T.C.
SELÇUK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DESTEK VEKTÖR MAKİNELERİNİN ETKİN EĞİTİMİ İÇİN YENİ YAKLAŞIMLAR

Emre ÇOMAK

DOKTORA TEZİ

ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ
ANABİLİM DALI

Konya, 2008

T.C.
SELÇUK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DESTEK VEKTÖR MAKİNELERİNİN ETKİN EĞİTİMİ İÇİN YENİ YAKLAŞIMLAR

Emre ÇOMAK

DOKTORA TEZİ

ELEKTRİK - ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

Bu tez 04.06.2008 tarihinde aşağıdaki jüri tarafından oybirliği / oyçokluğu ile kabul edilmiştir.

Prof.Dr. Ahmet ARSLAN
(Danışman)

Prof.Dr. Saadetdin HERDEM
(Üye)

Prof.Dr. Novruz ALLAHVERDİ
(Üye)

Doç.Dr. İbrahim TÜRKOĞLU
(Üye)

Yrd.Doç.Dr. Yüksel ÖZBAY
(Üye)

ÖZET

Doktora Tezi

DESTEK VEKTÖR MAKİNELERİNİN ETKİN EĞİTİMİ İÇİN YENİ YAKLAŞIMLAR

Emre ÇOMAK
Selçuk Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik-Elektronik Mühendisliği Bölümü

Danışman: Prof. Dr. Ahmet ARSLAN
2008, 125 sayfa

Jüri: Prof. Dr. Ahmet ARSLAN
Prof. Dr. Saadetdin HERDEM
Prof. Dr. Novruz ALLAHVERDİ
Doç. Dr. İbrahim TÜRKOĞLU
Yrd. Doç. Dr. Yüksel ÖZBAY

Sınıflandırma; yeni karşılaşılan veri örneklerinin önceden karşılaşılmış olan verilerden elde edilen bilgilerle fikir yürütülerek farklı sınıflara ayrıştırılması işlemidir. Sınıflandırma işleminin amacı, veri gruplarını oluşturan sınıf özellikleri ve bu sınıfların çıkış karakterleri arasındaki ilişkilerin keşfedilerek, bu ilişkiler ile yeni bir veri örneğinin sınıf etiketini tahmin etmektir.

Sınıflandırıcılar, eğiticili ve eğiticisiz olmak üzere ikiye ayrılır. Eğiticili öğrenmede, verilerle birlikte ait olduğu sınıflar da önceden sisteme bildirilir. Eğiticisiz öğrenmede ise, sınıflar önceden bilinmez ancak tahmin edilebilirler. Genellikle kümeleme metotları eğiticisiz olarak çalışırlar. Makine öğrenmesi yöntemleri, eğiticili (sınıflandırma ve regresyon) ve eğiticisiz (kümeleme) öğrenme problemlerinde kullanılmaktadır. Destek Vektör Makineleri ise, eğiticili veya yarı eğiticili olarak çalışabilen ve sınıflandırma ile regresyon uygulamalarında kullanılan bir makine öğrenmesi yöntemidir.

Destek Vektör Makineleri (DVM), güçlü bir teorik alt yapıya sahiptir. Çünkü istatistiksel öğrenme teorisinden yararlanan Vapnik-Chervonenkis boyutu ve Yapısal Risk Minimizasyonu prensipleri üzerine kurulmuş olan bir yöntemdir. Bu yaklaşımlar DVM'ne üstün bir genelleme yeteneği kazandırmaktadır. Bu yüzden birçok uygulama alanında başarıyla kullanılmaktadır. Fakat bazı zayıf yönleri de bulunmaktadır. Gerçekleştirilen tez çalışmasında, bu zayıf yönlerin kuvvetlendirilmesi hedeflenmiştir. Bu çerçevede; K-En Yakın Komşuluk (KEYK), Renyi entropi ve en küçük kareler regresyonu gibi metotlar yeni DVM eğitim algoritmaları geliştirilmesinde yardımcı metotlar olarak kullanılmıştır. Ayrıca, bu yardımcı metotlardan elde edilen ölçülerin değerlendirilmesinde Öklit uzaklıklarından da faydalanılmıştır. Bu yardımcı metotlar, DVM'ne yerel kontrol özelliği katmak için kullanılmış ve bu metotların yardımı ile zayıf yönleri kuvvetlendirilen yeni DVM eğitim algoritmaları geliştirilmiştir.

Geliştirilen metotlar (KKDVM, RKKDVM ve parametre düzenleyici yaklaşım) UCI (California Üniversitesi Bilimsel Veri Tabanı) ndan alınan İris, Şarap, Araç ve Tiroid veri kümeleri üzerinde ve Fırat Tıp Merkezi Kardiyoloji bölümünden elde edilen Dopler kalp kapakçıkları işaretlerine uygulanmıştır. DVM çeşitleri arasında, aynı amaç çerçevesinde geliştirilmiş olan En Küçük Kareler Destek Vektör Makineleri (EKKDVM) metodu ile de aynı veri kümeleri kullanılarak uygulamalar yapıldı. Tiroid veri kümesinde EKKDVM %95.62, KKDVM %96.62 ve RKKDVM %97.26, İris veri kümesinde EKKDVM %94.66, KKDVM %98.67 ve RKKDVM %100.00, Şarap veri kümesinde EKKDVM %96.59, KKDVM %97.73 ve RKKDVM %96.59 ve Araç veri kümesinde EKKDVM %68.72 ve RKKDVM %83.88 sınıflandırma doğruluğu test edilmiştir. Parametre düzenleyici yaklaşımın ise, en iyi parametre değerlerine başarıyla ulaştığı gözlenmiştir. Bu uygulama sonuçlarıyla, geliştirilen metodların üstünlüğü ortaya çıkarılmıştır.

Anahtar Kelimeler – Sınıflandırma, Destek Vektör Makineleri, Entropi, K-En Yakın Komşuluk, Regresyon, Dopler İşaretleri.

ABSTRACT

PhD Thesis

NEW APPROACHES FOR EFFECTİVE TRAINING OF SUPPORT VECTOR MACHINES

Emre ÇOMAK

Selçuk University

Graduate School of Natural and Applied Sciences

Department of Electrical-Electronics Engineering

Supervisor: Prof. Dr. Ahmet ARSLAN

2008, 125 pages

Jury: Prof. Dr. Ahmet ARSLAN

Prof. Dr. Saadetdin HERDEM

Prof. Dr. Novruz ALLAHVERDİ

Assoc. Prof. Dr. İbrahim TÜRKOĞLU

Assist. Prof. Dr. Yüksel ÖZBAY

Classification is trying to exactly separate the unknown data set to different groups according to knowledge about training data set. The aim of classification; discovering the relations between class characteristics and features of training data set and also estimating unknown class label of new datum via these relations.

Classifiers are divided into two groups such as supervised and unsupervised classifiers. In supervised learning, data set is given to classifier with output labels for classification. In unsupervised learning, output labels are not known during learning process. These labels can only be estimated from the data set. Clustering methods generally run as unsupervised learners. Machine Learning methods are used for supervised (classification and regression) and unsupervised (clustering) learning problems. SVM is a machine learning method used for classification and regression implementations and also run supervised or semi-supervised way.

SVM has been formed on powerful theoretical foundations. Since it was been formed by VC dimension and SRM principle which are very important concepts in statistical learning theory. These approaches provide reliable generalization ability to SVM. Therefore, SVM has been widely used in many application fields; however some deficiencies of SVM are available. In implemented studies, avoiding of these deficiencies is aimed. To achieve this goal, assisting methods to develop new SVM training algorithms such as KNN, Renyi Entropy and Logistic Regression have been used. These methods provide local control ability for learning process. Thus, new improved training algorithms have been introduced to provide local controlling property for SVM. In addition, for the evaluation of measurements obtained from these methods, Euclidean distance equation is used.

For analyzing performances of improved methods (CKSVM, RCKSVM and Parameter regularization approach), Iris, Wine, Vehicle and Thyroid data sets from UCI database and Doppler Heart Sound Signals from Firat Medical Centre are classified via improved methods. Among SVM classifiers, LS-SVM classifier has been selected as an alternative to our methods. Since, LS-SVM was been introduced to avoid the same deficiencies with our methods. Data sets classified by improved methods are also classified by LS-SVM classifier. Experimental results of our methods and LS-SVM classifier are compared. In Thyroid data set LS-SVM %95.62, CKSVM %96.62 and RCKSVM %97.26, in iris data set LS-SVM %94.66, CKSVM %98.67 and RCKSVM %100.00, in wine data set LS-SVM %96.59, CKSVM %97.73 and RCKSVM %96.59 and in vehicle data set LS-SVM %68.72 and RCKSVM %83.88 classification performance have been tested. Parameter regularization procedure has reached successfully to optimal parameter values. Experimental results approve efficiency of improved methods.

Key Words: Classification, Support Vector Machines, Entropy, K Nearest Neighbor, Logistic Regression, Doppler Signals.

ÖNSÖZ

Öncelikle, tez çalışmamda cesaretlendirici ve yol gösterici tavsiyelerinden dolayı danışmanım, sayın Prof. Dr. Ahmet Arslan'a ve tez süresince benimle paylaştıkları faydalı fikir alış verişlerinden ötürü tez komitemde yeralan sayın Prof. Dr. Saadetdin Herdem ve sayın Prof. Dr. Novruz Allahverdi hocalarıma teşekkürlerimi sunarım.

Ayrıca yıllardır bana olan desteklerini esirgemeyen tüm mesai arkadaşlarım ve hocalarıma teşekkür ederim.

Son olarak, doğduğum günden bu yana her konuda maddi ve manevi desteklerini sürekli yanımda hissettiren anneme ve babama saygılarımı sunar teşekkürü bir borç bilirim.

İÇİNDEKİLER

1.GİRİŞ

Öğrenme ve zekânın doğasının anlaşılması hakkında bazı çalışmalar yapılması ve insan doğasında bulunan bu kabiliyetlerin diğer alanlara uygulanabilmesi eğitim, idrak ile ilgili bilim dalları, bilgisayar bilimleri, sinir sistemi ile ilgili bilim dalları, mühendislik, sosyal bilimler ve fiziksel bilimler gibi bilim dallarını kapsayan geniş bir alanda artan bir ilgi görmektedir (Shen 2005). Çevreden edinilen bilgilerle öğrenme, insanların dünyayı ve doğru bilgiyi algılaması için temel yollardan birisidir. Günümüzde yoğun ve analiz edilmesi gereken bir veri akışı mevcuttur. Bunlar bilimsel, tıbbi, grafik tabanlı, finansal ve pazarlama alanlarında karşılaşılan türden veriler olabilir. Örneğin; tıbbi kurumlar teşhis kararlarını açıklamak için hastanın geçmiş verilerine odaklanırlar, mali şirketler pazarlama stratejilerine ve mali işlemlerine dair sağlıklı kararlar verebilmek için müşterilerden elde edilen büyük veri gruplarının bilgisayar yardımıyla otomatik olarak işlenmesine ihtiyaç duyarlar. Kamu ve özel üniversiteler ile şirketler, ulusal tehditleri ve yasal olmayan aktiviteleri belirlemek amacıyla güvenilir öğrenme tabanlı bir sistem geliştirmek isterler. Bu yüzden, öğrenmenin doğasını daha iyi anlamak ve bu tip problemleri çözmek için büyük ölçekli öğrenme algoritmaları ve onlarla ilişkili araçların geliştirilmesi gereklidir.

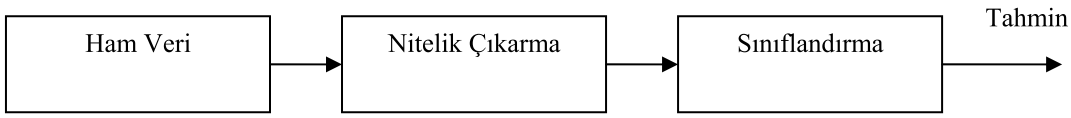
Bir örüntü, örneklem özelliklerinin birleşimiyle tanımlanan bir vektör olarak düşünülür. Örüntüyü tanımaya çalışıldığında, öğrenme boyunca tüm örneklerin görülemeyeceği problemiyle yüzleşilir. Bu yüzden mümkün olduğu kadar çok örüntünün tanınması istenir. Bu verilerin analiz edilmesi, sınıflandırılması ve eğiliminin belirlenmesi örüntü tanıma sistemlerinin ortaya çıkmasına sebep olmuştur. Bir istatistikçi için bu tanıma işlemi, çok sayıda örneğin toplanıp tümevarımsal metotlarla işlenerek tanınmaya çalışılması olabilir. Başka bir ifadeyle, görünebilir bir 3-boyutlu nesnenin fiziksel özelliklerinin 2-boyutlu görüntüler yardımıyla oluşturulduğu düşünülsün. Nesne mekanizmasının kanunlarını anlayarak örüntünün doğasını tanınmaya çalışılabilir. Aşağıdaki paragrafta, bu örüntü tanıma mekanizmalarının temelini oluşturan sınıflandırma, kümeleme ve regresyon işlemleri üzerinde durulmuştur.

Sınıflandırma; bilinmeyen bir veri örneğinin sonlu sayıda sınıftan hangisine ait olduğunun tahmin edilmesi işlemidir. Sınıflandırma işleminin amacı sınıf özellikleri ve çıkış özellikleri arasındaki bazı ilişkilerin keşfedilerek, bu ilişkiler ile sınıflandırılmamış yeni bir veri örneğinin sınıf etiketini tahmin etmek maksadıyla kullanılmasıdır. Sınıflandırıcılar genelde iki ana adımdan oluşurlar. İlk adımda, önceden belirlenmiş olan veri sınıfları veya kavramlarını tanımlayan bir model oluşturulur. Tipik olarak öğrenilen model, sınıflandırma kuralları, karar ağaçları veya matematiksel formüller gibi temsillerle ifade edilir. Bu model niteliklerle tanımlanmış veri gruplarının analiz edilmesiyle öğrenilir. Bu analiz edilen veri grubu eğitim kümesini oluşturur. Her bir eğitim örneğinin sınıf etiketi belirtildiği için bu adım eğiticili öğrenme olarak ta bilinir. Bir başka deyişle, modelin öğrenilmesi eğiticilidir ki, bu durum hangi sınıfın hangi eğitme örneğini içerdiğini açıklar. İkinci adımda model, sınıf etiketi henüz bilinmeyen örneklerin sınıf etiketlerini belirlemek maksadıyla kullanılır. Literatürde yarı eğiticili (verilerin bir kısmının çıkış değeri bilinen) öğrenme modelleri de geliştirilmiştir (Li ve ark.2003).

Kümeleme sınıflandırmanın tersine, önceden belirlenmiş bir sınıf etiketi olmaksızın veri grubundaki sınıfları belirleme amacı güder. Kümeleme eldeki verileri sınıflara veya kümelere ayırma işlemidir. Bu işlem gerçekleştirilirken aynı küme içindeki verilerin kendi aralarında yüksek benzerliğine ve tersine diğer kümelerdeki verilere yüksek oranda benzememezliğine önem verilir. Benzemezlik veri örneklerini tanımlayan nitelikler temel alınarak ölçülür.

Genellikle her bir eğitme örneğinin sınıf etiketinin bilinmediği bu öğrenme kalıbı eğiticisiz öğrenme olarak ta bilinir. Kümeleme, veri madenciliği, istatistik, makine öğrenmesi, görüntü işleme ve biyoloji gibi farklı birçok alanda ihtiyaç duyulan temel bir işlemdir.

Regresyon ise, aralarındaki ilişki gereği bağımsız olanının değerindeki değişim bağımlı olanının değerini değiştiren iki veya daha fazla değişken arasındaki ilişkinin, kalıpları önceden belirlenmiş olan matematiksel formüllerle modellenmesidir. Bu sayede sonlu bir gürültülü örnek kümesi üzerinde gerçek değerli bir fonksiyon tahmin edilebilir. Regresyonda sistemin çıktısı gerçek sayılar kümesine ait bir değişkendir (Hong ve Hwang, 2003).



Şekil 1.1 Temel Örüntü Tanıma Sistemi Yapısı

Bu süreçte nesnenin yapısını yansıtan farklı gözlemler veya fonksiyonlar elde edilebilir. Temel ham gözlemlerden yararlı nitelikler çıkarabiliriz. Bu tez çalışmasının konusu olan Destek Vektör Makinesi (DVM), istatistiksel öğrenme teorisi temelinde geliştirilen tümevarımsal bir algoritmadır ve öğrenme yapısını önceki bilgilerle birleştirerek tanıma işlemini zenginleştirir. Genel olarak örüntü tanıma sistemlerinin aşamaları şekil 1.1'de gösterilmiştir.

Makine öğrenmesi alanında, daha güçlü DVM sınıflandırıcılarının veya farklı durumları ele alabilen çekirdek makinelerinin geliştirilmesi üzerine çalışmalar devam etmektedir. Bir problemi çözmek için gerçek hayatta matematiksel modellere (dağılım ve olasılıkları temel alan istatistiksel modeller) ya da analitik modellere başvurulur. Hesaplama aşamasında, klasik veya esnek hesaplama teknikleri kullanılabilir. Klasik teknikler, sistemlerin doğal davranışlarını irdelemek amacı ile iyi kurgulanmış eşitlikler, diferansiyeller, dağılımlar ve olasılıklar kullanılarak oluşturulmuş sayısal gerçekleştirimler isterler. Gerçek hayatla ilgili veriler üzerinde çalışırken, bu koşulların tam olarak sağlanması çok zordur. Örneğin, klasik istatistiksel çıkarım metotlarının çoğu veri kümesinin normal dağılıma uygun olarak şekillendiği varsayımı üzerine kurulmuştur. Bazen hiç bir matematiksel model veya araç güncel problemin çözümünde yeterli olmayabilir. Ayrıca, uygun bir matematiksel model bulunsa bile, bu model problemin çözümü için oldukça karmaşık ve hesap yükü fazla olabilir (Shen 2005). Bu tip durumlarda analitik ve esnek hesaplama yöntemlerine yönelebilir. Bulanık sistemler, sinir ağları ve DVM sınıflandırıcılar mühendislik ve bilimsel problemlerin çözümünde en önemli bileşenlerdendir. Bulanık sistemler; bulanık kurallar, bulanık mantık ve diğer matematiksel yapıların kullanılması sonucunda, yapısal bilginin modellenmesine yardımcı olurlar. Sinir ağları ve DVM modelleri, veri ya da örneklerinden öğrenebilen metotlar grubuna girerler. Bu teknikler, eğitme verileri ile çıkışlar arasında doğrusal olmayan bir bağımlılık veya haritalama fonksiyonunu bulmaya çalışırlar (Shen 2005).

Destek Vektör Makinesi (DVM), Vapnik ve Chervonenkis tarafından geliştirilmiş olan istatistiksel öğrenme teorisinden türetilmiş yeni bir öğrenme metodudur (Vapnik ve Chervonenkis 1991). Sinir ağları, bulanık modeller ve sinir-bulanık ortak sistemleri gibi geleneksel öğrenme ve sistem modelleme metotlarıyla karşılaştırıldığında, DVM yüksek genelleme başarımı, en iyileme kapasitesi ve yüksek boyutlu küçük sayıda veri üzerinde dahi çalışabilme gibi özelliklere sahiptir. Genel olarak DVM, sınıflandırma ve regresyon işlemlerinde kullanılır. Günümüze kadar DVM, veri madenciliğinde (Burbidge ve Buxton 2001), finans sektöründe (Chen ve Shih 2006), çeşitli mühendislik problemlerinde (Chiang ve ark. 2004, Samanta 2004), görüntü işleme uygulamalarında (Sun ve ark. 2004), tıp (Übeyli 2007), (Çomak ve ark. 2007a, 2007b) ve gıda sektörü (Du ve Sun 2005) gibi çeşitli alanlarda başarıyla kullanılmıştır.

DVM sadece makine öğrenmesi alanında değil aynı zamanda esnek hesaplama alanında da önemli bir araştırma konusu olmuştur. Esnek hesaplama; bulanık mantık, sinir ağları ve olasılık tabanlı düşünmenin birleşimi olarak görülür. Geleneksel esnek hesaplama metotları; bulanık sistemleri, sinir ağlarını ve sinir-bulanık ortak sistemlerini içerir ki bu sistemler genelleme, esneklik ve iyi başarımlarından dolayı çok geniş uygulama alanlarında kabul görmüşlerdir. Esnek hesaplama alanında, DVM klasik esnek hesaplama modellerinin eğitimi için yeni bir öğrenme aracı olarak görülebilir (Shen 2005).

K-En Yakın Komşuluk (KEYK) sınıflandırıcıları, belirli bir benzerlik ölçüsünü temel alan ve parametrik olmayan bir örüntü sınıflandırma tekniğidir (Kim ve Shin 2000). KEYK, basitliği ve iyi performansından dolayı en çok kullanılan komşuluk tabanlı sınıflandırıcılardan birisidir. Etiketlenmemiş bir örnek verildiğinde, KEYK algoritması eğitme örnekleri arasından k tane en yakın (en benzer) komşusunu bulur ve bu komşularının ait olduğu baskın sınıf etiketini ilgili örneğin sınıf etiketi olarak belirler. Bu sınıflandırıcı yeni bir örnek ile karşılaşıncaya kadar hiçbir işlem yapmaz. Burada iki örnek arasındaki benzerlik basitçe onların nitelikleri arasındaki çakışma sayısı olarak tanımlanır. Eğer örnekler ikili nitelik vektörleriyle temsil ediliyorsa, benzerlik fonksiyonu iki nitelik vektörünün çarpım fonksiyonu olarak döndürülebilir. Tez çalışmasında, bu algoritma geliştirilen DVM öğrenme algoritmasına yardımcı bir yöntem olarak kullanılmıştır.

Renyi entropi kavramı, benzerlik ölçüsü olarak kullanılabilen ve kökleri bilgi teorisine kadar uzanan bir yöntemdir. İlk olarak Alfred Renyi tarafından Shannon bilgi teorisi tanımlamalarının entropi gibi değerlendirilmesiyle ve daha genel bir formda sunulmasıyla ortaya çıkmıştır (Renyi, 1976). Renyi entropi; kaynak ayırma, boyut azaltma ve nitelik çıkarma gibi problemlere uygulanmaktadır (Xu, 1999). Ayrıca Jenssen ve ark. (2003), bu metodu kümeleme işlemi için genişletmişlerdir. Bu tez çalışmasında da, ilgili metot kümeleme tabanlı yeni DVM öğrenme algoritması geliştirilmesinde ve DVM öğrenme parametrelerinin optimum değerlerinin bulunmasında kullanılmış ve gerçekleştirilen uygulamalarda iyi bir sınıflandırma başarımı elde edilmiştir.

1.1.Tezin Amacı ve Literatüre Katkıları

Her ne kadar DVM birçok uygulama alanında başarıyla kullanılan yeni bir makine öğrenmesi yöntemi olsa da bir takım kuvvetlendirilmesi gereken yönleri vardır. Bu zayıf yönler şu şekilde sıralanabilmektedir;

- Hesap yoğunluğu fazla olan (veri sayısının çok fazla veya verilerin çok yüksek boyutlu olduğu problemler) veri kümeleri üzerinde DVM metodunun eğitilmesinin çok zaman alması,
- DVM yönteminin eğitilmesi aşamasında kullanılması gerekli çekirdek fonksiyonu ve ayırıcı fonksiyon parametrelerinin en iyi değerlerinin en yüksek sınıflandırma doğruluğunu sağlayacak şekilde tespit edilmesi,
- DVM sınıflandırıcısının yanlış sınıflandırılması muhtemel verilere karşı hassaslığının giderilmesi,
- Ayırıcı fonksiyon hesaplanırken, bu düzlemin en az karmaşıklıkla aynı zamanda mümkün olduğunca en yüksek doğrulukta aranması,
- Son olarak da, DVM ile çok etiketli sınıflandırma işleminde karşılaşılan problemlerin çözülmesi.

Tez çalışmasında, bu problemlerden bir veya daha fazlasına çözüm getirecek yeni DVM öğrenme algoritmaları geliştirilmeye çalışılmıştır. Geliştirilen metotlar üç ana başlıkta sunulmuştur: Kümeleme Tabanlı K En Yakın Komşuluk Destek Vektör Makinesi (KKDVM), Renyi Kümeleme Tabanlı K En Yakın Komşuluk Destek Vektör Makinesi (RKKDVM) ve DVM eğitim parametrelerinin optimum değerlerine ulaşılmasını sağlayan yaklaşım.

Bu tezde, özellikle hesap yoğunluğu fazla olan veri kümelerinde sınıflandırma doğruluğu ve etkinlik bakımından başarılı olabilecek bir öğrenme algoritması geliştirilmesi üzerinde çalışılmıştır. Bu hedef doğrultusunda KKDVM ve RKKDVM yöntemleri geliştirilmiştir. Ayrıca DVM öğrenme işlemini tam otomatik hale getirecek parametre değerlerinin en iyi sınıflandırma başarımını verecek şekilde ayarlanmasını sağlayan bir prosedür geliştirilmiştir. Mevcut öğrenme algoritmaları bu tip verilerin analizinde yetersiz kalmaktadır. Hızlı ve sınıflandırma doğruluğu yüksek sınıflandırıcıların oluşturulması için bazı yardımcı metotların kullanılmasına da ihtiyaç vardır. Bu metotların uygun bir şekilde seçilmesi için de titiz çalışmalar yapılmıştır. Yapılan çalışmaların sonunda KEYK, Renyi entropi ve En Küçük Kareler Regresyonu gibi metotlar geliştirilen DVM öğrenme algoritmasında yardımcı metotlar olarak kullanıldı. Geliştirilen algoritmanın önemi aşağıdaki konular idrak edildiğinde daha iyi anlaşılmaktadır. Çünkü aşağıdaki konular geleneksel makine öğrenmesi yöntemlerindeki eksiklikleri gidermek için DVM'nin kurulması aşamasında tasarlanmıştır. Tezde geliştirilen eğitim algoritmaları bu konuları temel almakta ve orijinal DVM'de kullanılan halinden daha öteye taşımaktadır.

Teorik konular:

1. Geliştirilen DVM eğitim algoritmasının hangi veri kümesi üzerinde çalıştırılırsa çalıştırılsın sorunsuz olarak çıkış değerini hesaplayabilmesi gerekmektedir. Başka bir deyişle veriye göre uyarlanabilen esnek bir sınıf yapısının geliştirilmesi önemlidir. Çoğu durumda, karşılaşılan problemin yapısını istatistiksel temellere uygun olarak yansıtan başlangıç bilgisi yetersiz miktardadır. Genellikle eldeki verileri bilinen bir fonksiyona uyarlamak ya da bu yetersiz bilgi kaynağına göre modeller tasarlamak yetersiz sonuçlar verecektir.

2. İlk bakışta, sınıflandırma problemleri için daha çok sayıda nitelik üzerinde çalışmanın daha iyi ayırma gücü ve sınıflandırma başarımı ile sonuçlanması beklenir. Fakat pratikte görünen, daha çok niteliğin eklenmesi genellikle sınıflandırma doğruluğunu azaltır. Bellman (1961)'a göre bu boyut sıkıntısıdır (curse of dimensionality) ve nitelik sayısındaki artış eğitme örneklerinin sayısında üstel bir artışla desteklenmelidir.

3. Öğrenme algoritmasının sonlu veri kümesi üzerindeki genelleme başarımını kontrol edebilmesi bu algoritmaya üstünlük katacaktır. İstatistiksel öğrenme teorisi Vapnik (1995, 1998) tarafından önerilmeden önce, en yakın komşuluk sınıflandırıcısı gibi metotlar sonlu örneklem kümelerinde başarısız oldular ve genelleme yeteneği üzerine büyük etkisi olan sınıflandırıcı karmaşıklığını hesaba katmadan sadece deneysel riski minimum tutmaya çalıştılar. DVM gibi yapısal riskin minimum tutulması üzerine kurulan sınıflandırıcılar ise bu iki unsur arasında tercih sunarlar.

Algoritmik Konular

1. Çalışılan öğrenme algoritmasının düşük hesap maliyetli olması da ona ilave bir üstünlük kazandıracaktır. Örneğin, birçok öğrenme algoritması bir takım parametrelerle maliyet fonksiyonunu belirli koşullar altında minimum veya maksimum yapmaya çalışır. Bu algoritmaların, eğitme örneklerinin sayısına ve nitelik vektörü boyutuna doğrusal şekilde bağlı bir hız ve karmaşıklıkta yakınsama sağlaması beklenir.

2. Sınırlı sistem kaynakları ve hız gereksinimi arasında bir tercih yapılması gerekir. Her ne kadar günümüz bilgisayarlarının merkezi işlem birimleri oldukça hızlı da olsa, onların hafıza birimleri, bant genişlikleri ve ön bellekleri bir takım hesaplamalar için sınırlıdırlar. Bir öğrenme algoritması hafıza birimlerine tamamı ile yüklendiğinde oldukça fazla yer işgal eden büyük veri kümeleri üzerinde çalıştırıldığı zaman, disk ve hafıza birimleri arasında büyük bir veri transferi söz konusudur. Sonuçta, CPU biriminin kullanım oranı düşüktür. Geliştirilen algoritmanın çalışma süresi, bu kaynaklarla ve algoritmanın tasarlanma planlarıyla doğrudan ilişkilidir.

3. Paralel uygulamalar için iyi bir algoritma yapısının varlığı. Özellikle çok işlemcili ortamlarda paralel yapılı uygulamaların bu tip büyük veri kümeleri için geliştirilmesi gelecekte düşünülebilir.
4. Çevrim içi öğrenme; Sınıflandırıcının hata oranını azaltmak için başlangıçtaki eğitme örnekleri ile eğitilmiş sınıflandırıcıya yeni eğitme örnekleri eklenebilir. DVM konusunda bu hedefe yönelik başarılı bir çalışma Chen ve ark. (2003) tarafından literatüre katılmıştır. DVM sınıflandırıcısının eğitiminde diğer önemli hususlar da, sınıflandırma parametrelerinin eklenen verilere göre şekillendirilmesi mümkün müdür? Yoksa sınıflandırıcı yeniden mi eğitilmelidir? Sınıflandırma parametrelerinin optimum değerlerinin eğitme esnasında otomatik olarak bulunması için nasıl yaklaşımlar geliştirilebilir? Sorularına aranan cevaplarda gizlidir.

Veri analizi

Veri analizi konusunda istatistik tabanlı birçok değerli metot geliştirilmiştir. Fakat güncel hayat problemlerinin istatistiksel analizinde klasik problemlerden farklı olarak aşağıdaki problemler belirlenmiş ve ele alınmıştır.
1. Gürültülü veri: Çalışılan veri kümesi deneysel gürültüler veya aykırı veriler içeriyor olabilir. Verinin veya sınıf etiketlerinin elde edilişinde ölçüm veya karar aşamasında kolaylıkla tespit edilemeyen hataların meydana gelmesi de rastlanan bir durumdur. Bu tip veriler için bazı ön işlem metotlarının diğer aşamalara geçilmeden önce kullanılması gerekir.
2. Büyük veri kümeleri: Bu tip veri kümeleri nasıl ele alınabilir? Hesap yükünü ve hafıza depolama uzayını azaltmanın etkili bir yolu var mıdır? Sorularına cevap olabilecek yaklaşımların gerçekleştirilmesi gerekir.
3. Boyut sıkıntısı: Çoğu biyoloji probleminde, nitelik sayısı veri ile ilgili deneysel gözlem sayısından fazladır. Sınıflandırıcıyı eğitmek için ihtiyaç duyulan veri örneği sayısı, boyut sayısının üstel fonksiyonu şeklinde artar. Boyut sıkıntısının nasıl üstesinden gelinebileceği probleminin çözümü için de bazı metotlar geliştirilmiştir.
4. Doğrusal olmama: Güvenilir bir doğrusal veri analiz metodunu, doğrusal olmayan verileri ele alabilen bir metoda nasıl dönüştürülebilir? Klasik veri analiz metotlarının çoğunda yerel minimum sorunu mevcuttur. Bu problemin çözümü, DVM öğrenme metodunda çekirdek fonksiyonlarının yardımıyla sağlanmıştır.

Yukarıdaki teorik konular arasında, genelleme başarımının sonlu kümeler üzerinde nasıl kontrol edilebileceği istatistiksel öğrenme teorisinde (VC teorisi yapısında) formülleştirilmiştir.

1.2. Literatür Araştırması

Literatürde, bu tezin konusu ile doğrudan veya dolaylı olarak ilişkili olan bazı çalışmalar yapılmıştır. Özellikle son yıllarda yeni DVM eğitme algoritmalarının geliştirilmesi ve DVM öğrenme parametrelerinin en yüksek sınıflandırma başarımını sağlayan değerlerinin hesaplanması hususlarında çalışmalar yapılmaktadır. Bu çalışmaların ortak amacı ise sınıflandırma başarımı daha iyi, daha güvenilir ve etkin olarak çalışabilen DVM sınıflandırıcıların geliştirilmesidir.

Widodo ve ark. (2006), nitelik çıkarımı işlemi için Bağımsız Bileşen Analizi (BBA - Independent Component Analysis) ve Esas Bileşen Analizi (EBA - Principal Component Analysis) yöntemlerini kullanmışlar ve bu yöntemlerle kullanılan sınıflandırıcıların başarımlarını kıyaslamışlardır. Ayrıca DVM sınıflandırıcısı eğitilirken farklı olarak Sıralı Minimum Optimizasyon (SMO - Sequential Minimal Optimization) algoritması kullanılmıştır. Bu oluşturulan sistem mekanik motorun çalışması esnasında meydana gelebilecek hataların tespit edilmesi için kullanılmıştır.

Peddabachigari ve ark. (2007), bir bilgisayar sisteminde veya ağında meydana gelebilecek yetkisiz kullanımların tespiti amacıyla bileşik bir akıllı sistem kurmuşlardır. Sistem, karar ağaçları (decision trees) ve DVM (Support Vector Machines) sınıflandırma metotlarının birleştirilmesi ile oluşturulmuştur. Bu birleştirme işlemi esnasında düğüm elemanlarını DVM ile işleyerek yeni bir birleştirme metodu kullanılmıştır. Sistemin karar doğruluk oranının yüksek ve hesaplama maliyetinin düşük olduğunu gösteren deneysel sonuçlar elde etmişlerdir.

Shih ve Liu (2006), Ayrıştırıcı Nitelik Analizi (Discriminating Feature Analysis) ve DVM metotlarını kullanarak yeni bir yüz tanıma metodu geliştirmişlerdir. Metodun yeniliği Ayrıştırıcı Nitelik Analizi ile DVM metotlarının kullanılma şeklinden kaynaklanıyor. İlk aşamada Ayrıştırıcı Nitelik Analizi, giriş uzayından 1-boyutlu Haar Wavelet temsillerini ve yansıma büyüklük değerlerini çıkarır. Bu çıkarımlardan birincisi etkili bir temsil için yöntem sağlar. İkincisi ise, insan yüzünün yatay ve dikey özelliklerini içerir. Sonra modelleme ile yüzeyin yüz olup olmadığının olasılık yoğunluk fonksiyonu çıkarılır. Bu safhada 3 sınıf oluşur. Bu sınıflar; yüz sınıfı, yüz olmayan sınıf ve karar verilemeyen sınıf. DVM kullanılarak bu karar verilemeyen sınıfların yüz sınıfına mı, yoksa yüz olmayan sınıfa mı ait olduğu tespit edilir.

Acır ve ark. (2006), işitsel beyin kökü cevaplarının eşik değerini otomatik olarak bulan bir sistem tasarladılar. Ham veriler büyüklük değerleri olarak, Ayrık Kosinüs Dönüşümü (Discrete Cosine Transform) ve Ayrık Dalgacık Dönüşümü (Discrete Wavelet Transform) ile işlenerek üç farklı şekilde nitelik değerleri çıkartılmıştır. Daha sonra bu niteliklerin arasından en iyileri seçilerek nitelik azaltma işlemi gerçekleştirilmiştir. Son aşamada da DVM ile sınıflandırılmıştır.

Zhan ve Shen (2006), standart DVM den daha etkin ve daha iyi genelleme yeteneğine sahip bir ayırıcı düzlem (hyperplane) hesaplanması için yeni bir eğitme metodu sunmuşlardır. DVM metodunun etkinliği, kullanılan destek vektörlerinin sayısıyla doğrudan ilişkilidir. Gürültülü veriler DVM etkinliği ve genelleme yeteneği için sorundurlar. Bu çalışmada, DVM için yeni bir hedef fonksiyonu düşünülmüş ve gürültülü veriler için de farklı bir ceza terimi geliştirilmiştir. Yazarların geliştirdiği sistem UCI Adult veri kümesi üzerinde başarılı sonuçlar ortaya koymuştur.

Lerski (2006), bulanık mantık tabanlı bir sistemle DVM regresyon metodunu birlikte kullanmıştır. Bulanık mantık ile gerçekleştirdiği ön işlem sayesinde, DVM çekirdek (kernel) fonksiyonlarına alternatif bir haritalama işlemi gerçekleştirmiştir. Bu sayede çekirdek matrisinin veri kümesine göre otomatik olarak değişmesi sağlanmıştır.

Huang ve Wang (2006), eş zamanlı olarak hem DVM eğitme işlemini hem de DVM sınıflandırma başarımını en üst düzeyde tutacak alt nitelik kümesinin ve uygun çekirdek fonksiyonu parametre değerinin seçimini gerçekleştirecek yeni bir sistem tasarladılar. Bu işlemi gerçekleştirirken genetik algoritmalardan yararlanmışlardır. Çünkü DVM uygulamalarında nitelik kümesinin seçimi ve çekirdek fonksiyonlarının parametre değerlerinin seçimi sınıflandırma başarımını doğrudan etkileyen unsurlardır. Bu sistemin başarımı çeşitli veri kümeleri üzerinde değerlendirilmiştir.

Eitrich ve Lang (2006), türevsiz nümerik optimizasyon tekniğini kullanarak DVM parametre değerlerini otomatik olarak ayarladılar. DVM optimum bir şekilde eğitildiği zaman sınıflandırma ve regresyon işlemlerinde çok iyi sonuçlar veren çekirdek fonksiyonlarına sahiptir. Fakat dengesiz veri kümeleri için öğrenme parametrelerinin ayarlanması oldukça zordur. Bu çalışmada yeni bir uygunluk ölçüsü geliştirilmiştir. Geliştirilen sistem sayesinde bu parametre değerleri dengesiz veri kümelerinde bile daha kolay hesaplanmaktadır.

İflas tahmini, bankacılık sektöründe alınan kararlar üzerinde önemli etkiye sahip bir alandır. Son zamanlarda yapılan çalışmalar incelendiğinde, DVM tabanlı tahmin sistemlerinin Yapay Sinir Ağları (YSA) ve geleneksel regresyon sistemlerinden daha iyi sonuçlar verdiği görülmüştür. Min ve ark. (2006) na göre, genetik algoritmalarla (GA) birçok yapay zekâ metodunun birlikte kullanıldığı, fakat DVM ile birlikte kullanılan çalışmaların, çalışılabilecek alan daha geniş olmasına rağmen az sayıda olduğu anlatılıyor. Bu çalışmada GA, DVM ne ön işlem metodu (nitelik seçme/azaltma amaçlı) olarak ve en uygun parametre değerlerinin bulunmasında kullanılmıştır.

Chen ve Wang (2007), sezonluk zaman serileri tahmini için Sezonluk Otoregresyon Tümleşik Hareket Ortalaması (SORTHO - Seasonal Autoregressive Integrated Moving Average) modeli ile DVM metodunun üstün özellikleri birleştirilerek yeni bir regresyon sistemi tasarlamıştır. Bu üç sistemin (SORTHO, DVM ve birleşik yeni model) doğruluklarını karşılaştırmak için Tayvan endüstrisinden alınan veriler üzerinde bu üç sistem çalıştırılıp Normalleştirilmiş Ortalama Karesel Hata (NOKH - Normalized Mean Square Error) ve Ortalama Mutlak Yüzde Hatası (OMYH - Mean Absolute Percantage Error) hata değerleri hesaplanmıştır. Birleşik modelin hatası diğer ikisinden daha düşük çıkmıştır.

Lu ve ark. (2006), hem etiketli hem de etiketsiz verilerin bulunduğu bir görüntü verisi üzerinde yarı eğiticili DVM ile sınıflandırma yapmıştır. Etiketsiz verilerden de bir takım bilgiler çıkararak mümkün olduğunca birbirlerine yakın verilerin aynı sınıf etiketine atanması sağlanmıştır. DVM maksimum aşırı düzlem sağlanırken aynı zamanda lokal özelliklerin korunmasına da dikkat edilmiştir.

Pozdnoukhov ve Bengio (2006) çalışmalarında, öncelik bilgisini DVM gibi çekirdek metotlarını kullanan tekniklerle birleştiren genel bir metot sunmuştur. Bu metotta, öncelik bilgisi eğitme verilerinin tümü için tanımlanabilecek bir şekilde formülleştirilebildiği zaman tüm noktaların aynı sınıfa ait olduğu varsayılır. Bu uygulamada keskin ve esnek geometrik hedef nesneleri tanjant vektörleri ile oluşturulmuştur. Oluşturulan sistem yüz tanıma uygulamaları üzerinde denenmiş ve kullanılabilir kalitede olduğu gözlemlenmiştir.

Portföy optimizasyon problemi son zamanlarda çalışılan bir konudur. İnce ve Trafalis (2006) in yaptıkları çalışmada, portföy ve stok market verileri sınıflandırma problemine dönüştürülüp Minimum-Maksimum Olasılık Makinesi (MOM - Minimax Probability Machine) ve DVM gibi makine öğrenmesi teknikleriyle sınıflandırılmıştır. MPM metodu, hatalı sınıflandırma olasılığının sınırını bulur. DVM ise iki sınıf arası mesafeyi maksimum yapan aşırı düzlemi bulur. Kısa dönemli portföy yönetimi için iki metodun birbirine benzer sonuçlar verdiği gözlemlenmiştir.

Otomobil motorunun gücü ve torku etkili bir ayarlama ile belirlenir, bu da otomotiv mühendisinin tecrübesine bağlıdır. Vong ve ark. (2006) nın yapmış olduğu çalışmada, dinamometreden alınan eğitme verileri ile En Küçük Kareler Destek Vektör Makineleri (EKKDVM) eğitilmiştir. EKKDVM deki eğitim parametrelerin çıkarımı için Bayes karar kuralı kullanılmıştır. Kurulan sistemin eğitim süresi standart DVM ye göre daha kısadır.

Bio-informatik çalışmalarında sekans-yapı ilişkisinin anlaşılması önemlidir. Bunun için genellikle kümeleme metotları kullanılır. Geleneksel kümeleme metotları doğrusal olmayan düzlemde başarısızdır. DVM ise, bu tip düzlemlerde de iyi sonuçlar vermektedir. Fakat çok büyük veri kümeleri de DVM için uygun değildir. Zhong ve ark. (2007) nın çalışmalarında, bu sorunu aşabilmek için KDVM (Kümelemeli Destek Vektör Makinesi, Clustering Support Vector Machine) geliştirilmiştir. Geliştirilen tekniğin uygulamalar üzerinde yeterince genelleme yeteneği sağladığı da gösterilmiştir.

Hücredeki istenen bir proteinin tahmini proteinin fonksiyonunun anlaşılması açısından önemlidir. Bu konuda başarının artması çıkarılan niteliklerin uygunluğuna doğrudan bağlıdır. Kim ve ark. (2006) sunmuş oldukları çalışmada, uygun niteliklerin çıkarılması için Akıllı-çift Sıralı Düzenleme (ASD - Pairwise Sequence

Alignment) skorlarını kullanmışlardır. Sınıflandırıcı olarak ise, DVM kullanılmıştır. Deneysel çalışmalarda bu nitelik çıkarım tekniğinin yararlı olduğu görülmüştür.

DNA yapısını doğru olarak inceleyen metotlar bulunmasına rağmen, bunların hesap yükü oldukça fazla ve bu yüzden büyük veri yığınlarında yavaş çalışmaktalar. Zhang ve ark. (2006), yaptıkları çalışmada; ilk aşamada, Bayes karar kuralı ile giriş uzayından nitelik uzayına haritalama yapıldı. Daha sonra doğrusal DVM kullanılarak sınıflandırma işlemi gerçekleştirildi. Uygulama sonuçlarına göre sistemin diğer metotlardan önemli ölçüde hızlı ve hemen hemen aynı başarı oranını verdiği gösterilmiştir.

Gold ve ark. (2005), Otomatik İlişki Belirleme (OİB - Automatic Relevance Determination) ile ilişkili giriş niteliklerini seçip Bayes karar kuralı ile DVM sınıflandırıcısının parametre değerlerini en iyi şekilde ayarlayarak bir sistem geliştirmişlerdir. Yeni sistemin OİB kullanmayan DVM ye göre daha iyi sonuçlar verdiği gözlemlenmiştir.

Zhang ve Liu (2005), çalışmalarında sınıflandırmada ayrım işleminin etkinliğini azaltmadan daha küçük boyutlu alt uzaylar bulunması problemi üzerinde durmuşlardır. İlk olarak sınıflandırma için alt uzay kavramı regresyon ile ilişkili olarak formülleştirilmiştir. Sonra karar sınırı analizinin bu sayede gerçekleştirilmesi sağlanmıştır. Önceki karar sınırı analiz yaklaşımlarına oranla daha hızlı ve doğru sistemler olduğu gösterilmiştir.

Zhan ve Shen (2005), DVM ye daha fazla etkinlik kazandırmak amacıyla 4 aşamalı bir eğitim safhası tasarladılar. İlk aşamada, DVM tüm eğitme örnekleriyle eğitiliyor. Böylece belirli bir sayıda destek vektör üretiliyor. İkinci aşamada, ayırıcı yüzeyi daha eğimli yapan destek vektörler eğitme kümesinden çıkarılıyor. Üçüncü aşamada, DVM kalan eğitme örnekleriyle yeniden eğitiliyor. Sonuçta DVM sınıflandırıcısının karmaşıklığı yaklaşık bir ayırıcı düzlemle azaltılmış oluyor. Böylece sistemin performansı önemli bir ölçüde düşürülmeden etkinliği artırılmıştır.

Kulkarni ve ark. (2004) çalışmalarında, önemli mühendislik problemleri için üstün sınıflandırma başarımı sağlayan sağlam bir DVM sınıflandırıcısı üzerinde durmuşlardır. Sınıf genişliğinin minimum olması düşünülerek etkili bir ayarlama prosedürü geliştirilmiştir. Ajan tabanlı Genetik-Quasi_Newton optimizasyon algoritması ile optimum DVM parametreleri bulunmuştur. Algoritma çeşitli ikili ve çoklu sınıf etiketli uygulamalar üzerinde çalıştırılmıştır.

Cherkassky ve Ma (2005) bu çalışmalarında, çoklu model regresyon formülleştirmesi için yapısal bir öğrenme yaklaşımı tanımlanmıştır. Mevcut eğitme örnekleri farklı ve önceden bilinmeyen regresyon modelleri ile oluşturulur. Öğrenmede amaç iki aşamalıdır; mevcut verilerin çeşitli alt kümelere ayrılması ve her bir alt küme için regresyon modellerinin tahmin edilmesi. Ana kabul, verinin farklı kısımlarının farklı modellerle tanımlanabilmesidir. Böylece bu çalışma diğerlerinden farklı bir formülleştirme biçimini eğitme aşamasında kullanmaktadır.

Genel eğiticili bir uzaktan kontrollü görüntü sınıflandırma probleminde öncelik bilgisinin veri kümesi içinde mevcut olduğu varsayılır. Fakat bu öncelik bilgisi çoğu zaman nesneyi gerçekten tanımlamayabilir. Bu da sınıflandırıcının yanlış sınıflandırma kararları vermesine sebep olur. Bu yüzden Mantero ve ark. (2006) yarı eğiticili bir sınıflandırma sistemi tasarladılar. Test örnekleri arasında eğitme örneklerinin hiç birine uymayan örnekler bulunabilir. Bu tip örnekleri de ayırt edebilmek için DVM olasılık yoğunluk tahmini tabanlı bir Bayes çıkarım metodu geliştirilmiştir. Ayrıca bu metot yapay ve gerçek bazı veri kümeleri üzerinde çalıştırılmıştır.

Hong ve ark. (2004) çalışmalarında, kaba küme teorisi ile DVM'ni birleştirerek ağ güvenliği üzerindeki yetkisiz kullanımları bulmaya çalışan bir sistem geliştirilmiştir. DVM sınıflandırıcısının bir eksikliği olan çok büyük veri kümeleri üzerindeki yavaşlığını azaltmak için önce kaba küme teorisi ile verinin boyutu azaltılmış daha sonra da DVM ile daha kısa sürede sınıflandırma yapılmıştır.

Lin ve Wang (2004), önceki çalışmalarında Bulanık Destek Vektör Makinesi (Fuzzy Support Vector Machine) sınıflandırıcısının gürültülü veriler üzerinde etkili olduğunu göstermişlerdir. Bu çalışmada ise güvenlilik faktörü, değersizlik faktörü ve bir haritalama fonksiyonu yardımıyla bulanık üyelik fonksiyonlarını otomatik olarak oluşturmuşlardır. Ayrıca, uygulama sonuçlarının da olumlu olduğunu gösterdiler.

Chua (2003), standart EKKDVM sınıflandırıcısında kullanılan matris yerine daha küçük boyutlu bir matris kullanarak milyonlarca veri içeren bir veri kümesi üzerinde bile daha kısa sürede sınıflandırma yapılabilmesini mümkün kılınmıştır. 10 nitelik içeren 1 milyon veri üzerinde 45 saniyede eğitme yaptığını belirtmiştir.

Ayrık karar fonksiyonlarına sahip ve bire karşı kalanlar yaklaşımını kullanan DVM çok sınıflı sınıflandırıcısı, sınıflandırılamayan bölgeler içerir. Bu bölgeleri sınıflandırabilmek için sürekli karar fonksiyonları veya bulanık mantık yöntemi kullanılabilir. Hata Doğrulama Çıkış Kodları (ECOC) yaklaşımında sadece ayrık karar fonksiyonları yerine sürekli karar fonksiyonlarının kullanılması bu problemi çözer. Kikuchi ve Abe (2005) nin çalışmalarında, yeni bir sürekli karar fonksiyonu ile bulanık DVM sınıflandırıcı sistemi kurulmuş ve ECOC yaklaşımını kullanan DVM ile karşılaştırılmıştır. Önerilen bulanık DVM sınıflandırıcısının daha üstün olduğu gösterilmiştir.

Geleneksel DVM veri noktalarına iki ayrık değerden birisini atar. Jayadeva (2004) çalışmasında, uzaklık ölçütleri yerine bulanık üyelik fonksiyonlarını kullanarak bu iki ayrık değerden birisine atama yapılması düşüncesiyle bulanık bir sistem kurmuştur. Algoritma basit ve hızlıdır.

Camps-Valls ve ark. (2004) bu çalışmalarında, bulanık sigmoid fonksiyonunun DVM sınıflandırıcısında çekirdek (kernel) fonksiyonu olarak kullanılması üzerinde durmuştur. Bulanık sigmoid çekirdek, düşük hesaplama yüküne ve yüksek oranda çekirdek matrisinin Eigen değerlerine izin vermesiyle standart sigmoid çekirdeğinin bazı sınırlarını aşmıştır.

Transdüktif Destek Vektör Makinesi (TDVM - Transductive support vector machine), DVM sınıflandırıcılarının uyumlu bir şekilde kullanılmasını öngören bir çıkarımdır. TDVM, sınıflandırma işlemlerinde etiketsiz verilerden de yararlanarak standart DVM sınıflandırıcısından daha iyi başarım sağlar. TDVM yönteminin de hala bir eksikliği vardır, Eğitme işlemine geçmeden önce sisteme pozitif örnek sayısı verilmelidir ve bu sayı eğitim süresince değiştirilmemelidir. Bu eksiklik TDVM de kullanılan karşılıklı değişim koşullarından kaynaklanır. Wang ve Huang (2005) ın çalışmalarında, yeni bir bireysel değerlendirme ve değişim koşulu sunulmaktadır. Yeni metodun TDVM sınıflandırıcısının uyarlanabilme özelliğini geliştirdiği ve pozitif örnek sayısını bulmada daha iyi olduğu gösterilmiştir.

Yanıt modelleme işlemi, pazarlamada anahtar faktör haline gelmiştir. Kim ve ark. (2008) na göre, cevap modelleme genellikle iki aşamadan oluşur. İlk aşama cevaplayan kişilerin müşteri veri tabanından tanınması ve ikinci aşama ise müşterilerin satış miktarının tahmin edilmesidir. Bu çalışmada ikinci aşamaya odaklanılmıştır. Bu yüzden konu regresyon işlemi olarak ele alınmıştır. Son zamanlarda DVM gibi bazı doğrusal olmayan makine öğrenmesi teknikleri bu alana uygulanmıştır. Fakat veri kümeleri çok büyük olduğu için cevaplama çok uzun zaman alabilmekte veya kötü modeller üretilebilmektedir. Bu yüzden pratikte örneklem metotları da kullanılmıştır. Bu uygulama da kötü sınıflandırma sonuçlarına sebep olmuştur. Bu çalışmada, yazarlar örneklem işlemini yerine getirmek için Destek Vektör Regresyon metodunu kullanmışlardır. Diğer örneklem metotlarına göre daha iyi sonuçlar elde etmişlerdir.

Dong ve ark. (2008), DVM eğitim aşamasında kullanılan parametrelerden bir tanesi olan C maliyet parametresinin uygun değerinin bulunmasını sağlayan iki aşamalı bir yaklaşım önerdiler. Denge Kısıtlarıyla Matematiksel Programlama (DKMP - Mathematical Program with Equilibrium Constraints) formunda DVM ni modellediler. Sonra başlangıç C_0 en düşük bant genişliği değeri verilerek en uygun C değerine kadar artırma yapılarak değer bulunur. DVM nin kullanıldığı bazı sayısal problemlerde ilgili yöntemin makul sonuçlar verdiği görülmüştür.

Hong ve ark. (2008) nın sunmuş oldukları çalışmada, parmak izi sınıflandırma sayesinde parmak izlerinin önceden tanımlanmış sınıflarda gruplanarak otomatik parmak izi tanıma sistemlerindeki olası eşleşme sayısının azaltıldığını belirttiler. DVM nin parmak izi tanıma uygulamalarında başarılı sonuçlar verdiği için tercih etmişlerdir. Bire karşı diğerleri çoklu sınıf sınıflandırma yaklaşımını Bayes sınıflandırıcılarıyla düzenleyerek yeni bir metot önerdiler. Bu metot çok sınıflı sınıflandırma uygulamalarında görülen düğüm alanı probleminin çözümü için geliştirilmiştir. Daha detaylı olarak, yaklaşım bire karşı diğerleri DVM ve Bayes sınıflandırıcılarını eğitmek için parmak kodu olarak parmak izi niteliklerini, benzersizlikleri ve sırtları kullanır. Önerilen metot NIST-4 veri tabanı üzerinde doğrulanarak %90.8 oranında sınıflandırma doğruluğu gösterildi.

Coussement ve Poel (2008), grid arama ve çapraz geçerlilik testi temelinde çalışan iki parametre optimizasyon prosedürü geliştirdi. Bu metotlar DVM öğrenme parametrelerinin en uygun değerlerinin bulunması amacıyla kullanıldı. Bu yapıyı Müşteri İlişkileri Yönetimi (MİY - Customer Relationship Management) alanına uyguladılar ve performanslarını karşılaştırdılar.

Shih-Wei Lin ve ark. (2007), Parçalı Sürü Optimizasyonu (Particle Swarm Optimization) ile nitelik seçimi ve DVM için öğrenme parametrelerinin optimizasyonunu sağladılar. 12 UCI veri kümesi üzerinde uyguladılar. Genetik Algoritma temelli yapıyla karşılaştırılabilir sonuçlar elde ettiler.

Literatürdeki bu incelemelerimiz ışığında, son yıllarda DVM metoduyla ilgili olarak iki ana başlıkta yer alan çalışmaların yayınlandığını görmekteyiz. Bunlardan ilki, var olan DVM tekniklerinin yeni uygulama alanlarında da çalıştırılarak diğer tekniklere karşı güçlü bir alternatif teknik olduğunun ve avantajlarının gösterilmesidir. Diğer çalışma grubu ise, bu tezin de amacıyla yakından ilgili olan DVM öğrenme metodunun çeşitli yaklaşımlarla geliştirilmesidir. DVM yeni ve sağlam bir öğrenme metodu olmasına rağmen zayıf yönleri bulunmaktadır ve bu yönlerin yardımcı metotlarla kuvvetlendirilmesi gerekmektedir. Bu zayıf yönler özetle, büyük veri kümelerindeki hesaplama maliyetinin yüksekliği ve bunun yol açtığı eğitim zamanının uzaması, ayırıcı düzlemin sadece az sayıda eğitme örneği (destek vektör) referans alınarak oluşturulması sonucunda aykırı veri noktalarına karşı oluşan hassasiyet, öğrenme parametrelerinin sınıflandırma başarımını doğrudan etkilemesine karşın bunların optimum değerinin bulunmasını sağlayacak bir yaklaşımın orijinal DVM metodu içinde yer almaması, dengeli olmayan veri kümelerinde sınıflandırma başarısının azalması ve orijinal DVM sınıflandırıcısının sadece *+1, -1* çıkışlarını üretebilmesi şeklinde sıralanabilir.

1.3. Tezin Organizasyonu

Tez çalışmasının organizasyonu aşağıdaki gibi şekillendirilmiştir.

Bölüm 2'de DVM metodunun teorik alt yapısını teşkil eden istatistiksel öğrenme teorisinden, orijinal DVM metodundan, bazı problemler sonucu ortaya çıkmış olan DVM ne yardımcı yaklaşımlardan ve daha sonra geliştirilen EKK-DVM yapısından bahsedilmektedir.

Bölüm 3'te tez çalışmasının esas amacı olan yeni bir DVM eğitme algoritması geliştirilmesi hususunda DVM ana prensiplerine yardımcı olarak kullanılan Renyi

Entropi, K En Yakın Komşuluk ve En Küçük Kareler Regresyonu metotları üzerinde durulmaktadır.

Bölüm 4'te tez çalışmasında geliştirilen yeni DVM öğrenme algoritmaları sunulmaktadır.

Bölüm 5'te uygulamalarda kullanılan veri kümeleri ve uygulama sonuçları anlatılmaktadır.

Bölüm 6'da ise, tezin sonuç ve öneriler kısmı yer almaktadır.

2. İSTATİSTİKSEL ÖĞRENME TEORİSİ, ÇEKİRDEK FONKSİYONLARI VE DESTEK VEKTÖR MAKİNELERİ

2.1. İstatistiksel Öğrenme Teorisi

Genel anlamda veriden öğrenme, sonlu sayıda eğitme verisi kullanan tahmin fonksiyonunun öğrenilmesini sağlayan öğrenme makinesi veya algoritmanın oluşturulması olarak tanımlanır. Öğrenme işlemi genel olarak, sınıflandırma, regresyon, kümeleme ve nitelik çıkarımı gibi işlemleri kapsar. Makine öğrenmesi teknikleri istatistiksel açıdan dağılım temeline göre işleyen teknikler ve dağılımdan bağımsız olarak çalışabilen teknikler olmak üzere ikiye ayrılırlar. Bu teknikler de ayrıca kendi içinde eğiticili ve eğiticisiz olarak öğrenen teknikler olmak üzere ikiye ayrılırlar. DVM metodu, dağılımdan bağımsız olarak çalışabilir ve istatistiksel öğrenme teorisine göre eğiticili veya yarı eğiticili olarak sınıflandırma ve regresyon işlemlerini gerçekleştirebilir. Klasik istatistiksel tekniklere göre oluşturulan regresyon ve sınıflandırma uygulamaları, olasılık dağılım modelleri veya olasılık dağılım fonksiyonları olarak bilinen kesin varsayımlara göre çalışırlar. Bu varsayımların pratikte sağlanması ise zordur. Bu yüzden, olasılık dağılımının bilinmediği durumlarda dağılımdan bağımsız tekniklerin kullanılmasına ihtiyaç duyulur. Güncel uygulamaların çoğundaki yegâne bilgi sonlu veri kümesidir ki, o bilgi de çok boyutlu ve sınırlı sayıdadır. Bu yüzden öğrenme makinesinin bu koşullarda dahi işlem yapabilmesi istenir. Maalesef, sinir ağları ve bulanık sistemler gibi geleneksel metotlar bu yeteneğe sahip değillerdir ve bu tip durumlarda güvenilir olmayan sonuçlar üretirler. Ayrıca, yüksek boyutlu uzay ise boyut sıkıntısı (curse of dimensionality) problemine sebep olur.

Sezgisel olarak, daha iyi bir eğitim için daha çok eğitim verisinin kullanılması uygundur. Eğitim veri kümesi yeterince büyük olduğu zaman (sonsuza yakın), eğitme hatası çok küçük (sıfıra yakın) olabilir. Fakat eğer eğitme kümesi çok küçük ise, eğitme hatası oldukça büyük olacak ve öğrenme işleminin sonucunun güvenilir olması sağlanamayacaktır. Eğer eğitim veri kümesi gürültü içeriyorsa, eğitim hatasını azaltmak için yine daha çok eğitim verisine ihtiyaç duyulur. Kecman (2001), örneklem boyutunun belirlenmesi için eğitim verisi sayısının VC (Vapnik-Chervonenkis) boyutuna oranını ölçü almıştır. Bu oran değeri 20 den küçük çıkan veri kümeleri küçük, 20 den büyük çıkanlar ise orta ölçekli veri kümeleri olarak ele alınmıştır.

Buraya kadar dikkat edilen hususlar sadece eğitim hatası düşünülerek planlanmıştır. Genelleme hatası (bilinmeyen test verisi üzerindeki hata) veya öğrenme metodunun genelleme kapasitesi hiç hesaba katılmamıştır. Eğer giriş veri kümesi yüksek boyutlu veya verinin karakteristiğini ifade eden temel fonksiyon çok karmaşık ise, öğrenme işleminde daha çok veriye ihtiyaç duyulur. Ayrıca, boyut sıkıntısı (sinir ağları için çok sayıda gizli nöron ve bulanık modeller için çok sayıda bulanık kural anlamına gelen) giriş uzayının boyutunu artıracaktır. Bu iki problemin ele alınmasıyla birlikte, yeni öğrenme makinelerinin geliştirilmesi ihtiyacı ortaya çıkmıştır. DVM; yüksek boyutlu ve küçük sayıda eğitim verisinden öğrenebilen istatistiksel öğrenme teorisi çatısı altında yönlendirilmiş yeni nesil bir öğrenme metodudur (Shen 2005).

Genel olarak öğrenme, olasılık temeline göre işleyen bir süreçtir. Veriden öğrenme işlemi üç temel bileşenden oluşur: üretici, danışman ve öğrenme makinesi (Vapnik 1995,1998). Üretici bileşeni, giriş vektörlerini (bilinmeyen bir dağılıma uygun x vektörleri) üretir. Danışman, her bir giriş vektörüne göre y eğitme cevabı değerini döndürür. X girişleri ve y cevapları öğrenme makinesinin eğitilmesi amacıyla kullanılır. Öğrenme makinesi bu giriş ve çıkış kümeleri arasındaki bağlantıyı $f(x,\alpha)$ fonksiyonlar kümesini kullanarak öğrenir. Bu fonksiyonlar kümesi istatistik literatüründe hipotez uzayı olarak adlandırılır. L rasgele ve birbirlerinden bağımsız eğitim verisi incelensin. Veriden öğrenme problemi, danışmanın cevaplarını mümkün olan en iyi şekilde tahmin eden $f(x,\alpha)$ fonksiyonunun seçilmesidir. Sınıflandırma ve regresyon gibi işlemlerde, sonuç fonksiyonu üzerinde çalışılan veri kümesinin doğal yapısını en iyi şekilde yansıtmalıdır. Bu fonksiyon hipotez uzayından bir hipotezdir. Eğer çıkış uzayı sonlu sayıda elementten oluşuyorsa, bu öğrenme işlemi sınıflandırma adını alır.

Tahmin fonksiyonu $f(x,\alpha)$'nın kalitesini ölçmek için $(L(y,\ f(x,\alpha)))$ ile gösterilen kayıp fonksiyonu ölçümü kullanılmalıdır. Bu kayıp fonksiyonu, öğrenme makinesi tarafından üretilen yakınsama değeri ile danışmanın cevap değeri arasında bir kayıp değeri tanımlar. İstatistiksel araştırmalarda farklı uygulama alanları için birçok kayıp fonksiyonu kullanılmıştır. Geleneksel regresyon problemlerinde de kullanılan en yaygın iki kayıp fonksiyonu, ortalama mutlak hata (OMH) ve ortalama karesel hatadır (OKS). Aslında bu hata oranları, öğrenme makinesinin gerçek risk veya beklenen risk oranlarını yansıtmamaktadır. Beklenen risk veya kayıp fonksiyonunun beklenen değeri aşağıdaki formülle ifade edilebilir (Vapnik 1995,1998). Aşağıda *R(α)* beklenen riski, *dP(x, y)* olasılık dağılımını gösterir.

$$R(\alpha) = \int L(y, f(x,\alpha))dP(x,y) \tag{2.1}$$

Burada *L(y, f(x,α))* özel bir kayıp fonksiyonudur ve eğitme kümesinden hesaplanır.

Özel veriden öğrenme uygulamaları sınıflandırma ve regresyondur. Farklı tipte kayıp fonksiyonları tanımlayarak farklı uygulamalar için özel öğrenme makineleri oluşturulabilir. Popüler iki veriden öğrenme problemi; örüntü tanıma (sınıflandırma) ve regresyondur.

Sınıflandırma tipi problemlerde, danışmanın cevapları öğrenme makinesinin çıkış etiketleridir. Örneğin ikili sınıflandırma problemi için $y = \{0,1\}$ olur ve verilen *f(x,α)* fonksiyonlar kümesi 0 veya 1 değerlerinden birisini alabilir. İkili sınıflandırma problemi için kayıp fonksiyonu aşağıdaki gibi tanımlanır.

$$L(y, f(x,\alpha)) = \begin{cases} 0, y = f(x,\alpha) \\ 1, y \neq f(x,\alpha) \end{cases} \tag{2.2}$$

Örüntü tanıma probleminin amacı, eşitlik 2.1'deki beklenen riskin yani bilinmeyen bir olasılık dağılımına göre alınan eğitme verisi kümesi üzerindeki sınıflandırma hatası olasılığının minimum tutulmasıdır. Denklem 2.2'den de şu çıkarılabilir ki, danışmanın cevabı sistemin bulduğu fonksiyon çıkışı ile aynı olmadığı zaman bu risk değeri artar.

Regresyon problemleri için öğrenme makinesinin çıkış değerleri gerçel sayılar kümesine dâhildir. Verilen *f(x,α)* fonksiyonlar kümesi; gerçel fonksiyonlardan oluşur ve optimum regresyon fonksiyonunu içerir. Regresyon problemlerinde kullanılan iki yaygın kayıp fonksiyonu bulunmaktadır. Bunlardan birincisi karesel hata yapısını kullanan L2 norm (denklem 2.3), diğeri ise mutlak hata yapısını kullanan L1 norm (denklem 2.4) kayıp fonksiyonlarıdır.

$$L_2(y, f(x,\alpha)) = (y - f(x,\alpha))^2 \quad (2.3)$$

$$L_1(y, f(x,\alpha)) = |y - f(x,\alpha)| \quad (2.4)$$

Epsilon, ξ duyarsız kayıp fonksiyonu olarak adlandırılan yeni bir kayıp fonksiyonu Vapnik (1995,1998) tarafından tanıtılmıştır. Bu fonksiyonun uygulamalarda, DVM yapısına daha uygun olduğu görülür. Regresyon problemindeki amaçta, yine eşitlik 2.1'deki beklenen riskin dağılım fonksiyonu bilinmeyen eğitim veri kümesi üzerindeki değerinin bu kez 2.3 ve 2.4 eşitlikleri veya epsilon duyarsız kayıp fonksiyonuna göre minimum tutulmasıdır.

Vapnik ve Chervonenkis (1991) tarafından geliştirilen istatistiksel öğrenme teorisi küçük veri kümeleriyle de öğrenme işlemini başarmış ilk etraflı öğrenme teorisidir. Veriden öğrenme problemi analizine göre beklenen riskin tanımı şu şekilde özetlenebilir: $D(x_i,y_i)$, $i=1,...,\ell$ eğitim örneklem kümesi verilsin, denklem 2.1'deki eşitlik beklenen risk fonksiyonu olarak adlandırılır. f elemanıdır $R^n * R$ kümesidir ve n giriş uzayının boyutudur (nitelik sayısı). Beklenen risk değerini minimum tutmaya çalışan tümevarımsal öğrenme makinesi metotları, yapısal risk minimizasyonu prensibindeki öğrenme metotları olarak bilinirler. Fakat verilerin ait olduğu olasılık tabanlı dağılım bilinmediği için bu beklenen risk değeri tam olarak hesaplanamaz. Sonuçta, beklenen risk fonksiyonu doğrudan minimum yapılamaz. $D(x_i,y_i)$, $i=1,...,\ell$ eğitim örneklem kümesi verilsin, eşitlik 2.5 deki deneysel risk fonksiyonu (R_{emp}) olarak adlandırılır. f elemanıdır $R^n * R$ kümesidir ve n giriş uzayının boyutudur.

$$R_{emp}(\alpha) = \frac{1}{\ell}\sum_{i=1}^{\ell} L(y_i, f(x_i,\alpha)) \quad (2.5)$$

Eğer kayıp fonksiyonu mutlak hataya dayalı ($L1$ norm) ise, deneysel risk fonksiyonu (R_{emp}) eşitlik 2.6'daki şekli alır.

$$R_{emp}(\alpha) = \frac{1}{ÖS}\sum_{i=1}^{ÖS}\left|y_i - f(x_i, \alpha)\right| \tag{2.6}$$

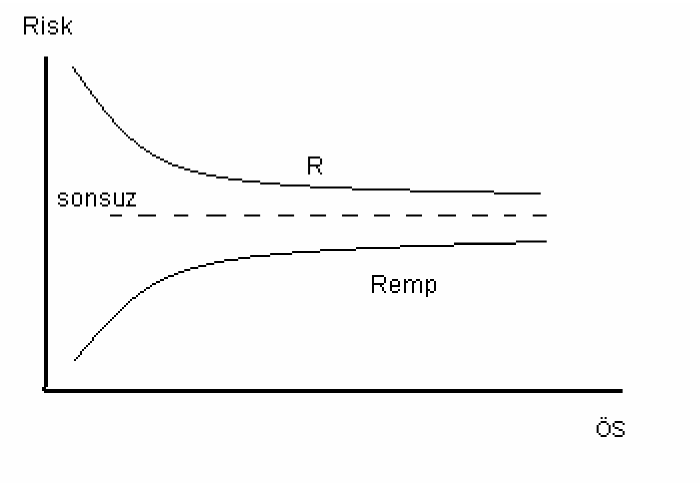
Burada *ÖS*, eğitme verilerinin sayısını gösterir. Genel deneysel risk fonksiyonu $R_{emp}(\alpha)$, geleneksel regresyon metotları ve öğrenme algoritmaları tarafından bulunur. Öğrenme makinesinin deneysel riskini minimum tutmayı amaç edinen tümevarımsal metotlar Deneysel Risk Minimizasyonu (DRM) prensibiyle ifade edilirler. Deneysel risk, eğitim örnekleri için yanlış sınıflandırılma olasılığını temsil ederken, beklenen risk bilinmeyen dağılımdaki farklı örneklerin yanlış sınıflandırılma olasılığını temsil eder. DRM genelleme prensibini temel alan metotlar deneysel risk fonksiyonu yapısında sadece verilen eğitim kümesi ile ilgili bilgiler barındırdığı için iyi bir performansı garanti edemezler. Bu hata oranı, gerçek hatadan veya beklenen riskten oldukça farklı olabilmektedir.

2.2. Deneysel Risk Minimizasyonunun Kararlılık ve Yakınsama Prensibi

En küçük kareler hesabı gibi bazı geleneksel yaklaşımlar 2.7 eşitliğinde verilen prensibi temel alırlar. Bu prensip; Yapay Sinir Ağları, Bulanık Mantık gibi geleneksel metotlar tarafından referans alınmaktadır.

$$\lim_{ÖS \to \infty} (R(\alpha) - R_{emp}(\alpha)) = 0 \tag{2.7}$$

Bu prensibe göre *ÖS* değeri sonsuza doğru artarken, deneysel risk değeri beklenen risk değerine yaklaşmaktadır. Her ne kadar bu prensip deneysel riski minimum yapan fonksiyonu garanti altına almasa da, beklenen riski minimum yapan fonksiyona veya gerçek riske yakınsar (Kecman, 2001). Bu önemli kavram şekil 2.1'de açıklanmıştır.



Şekil 2.1 Deneysel ve Gerçek Riskler

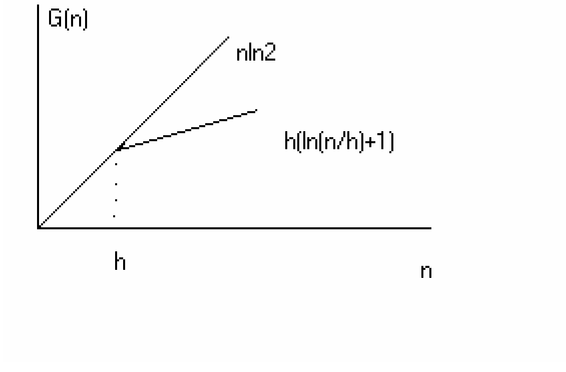
Bu tek düzenli yakınsama prensibine göre, *ÖS* değeri artarken deneysel risk fonksiyonu beklenen risk fonksiyonuna daha çok yaklaşır. Fakat bahsedilen DRM'nin uygunluğu ve yakınsama prensibi yapısal prosedürlerin nasıl oluşturulacağı hakkında bir yol göstermez. Gerekli modellerin oluşturulabilmesi ancak yapısal prosedürlerin tanıtılması, Yapısal Risk Minimizasyonu (YRM) tümevarımsal prensibi ve VC teorisinin geliştirilmesi sayesinde başarılmıştır. Bu kavramlar istatistiksel öğrenme teorisinin ve aynı zamanda DVM öğrenme metodunun da temelini oluşturmaktadır.

2.3 Yapısal Risk Minimizasyonu Prensibi ve Vapnik-Chervonenkis Teorisi

YRM tümevarımsal prensibi ve VC teorisi istatistiksel öğrenme teorisinde önemli bir yere sahiptir. YRM, küçük örneklem kümelerinden öğrenme ile ilgili uygulamalarda yararlı olan yeni bir tümevarımsal prensiptir (Kecman, 2001). Olasılıkta tek düzenli yakınsama, öğrenme makineleri için beklenen risk ile deneysel risk arasındaki sapma değerine bir sınır getirir (Trafalis ve Ince, 2000). İstatistiksel öğrenme teorisinde ancak, yakınsama fonksiyonları kümesini daha küçük bir hipotez uzayında kısıtlayarak ve aynı zamanda bu yakınsama fonksiyonlarının esneklik ve karmaşıklıklarını da kontrol ederek bu sınırlar minimum tutulabilir. Bu yüzden YRM prensibi ve VC teorisindeki temel fikir; çok sayıda aday model arasından beklenen risk veya genelleme hatasını minimum yapacak doğru oranda karmaşıklığa sahip (deneysel riski ve yakınsama fonksiyonlarının karmaşıklığını aynı anda minimum yapan) modelin seçilmesidir. Böylece, seçilen optimum öğrenme makinesi verilen eğitim kümesi için en uygun kapasiteye sahip olacaktır.

Bu fikir, sınıflandırma veya regresyon işlemleri gerçekleştirilirken meydana gelen fazla uyum ve eksik uyum problemlerinin analizinde de gözlenebilir. Eğer amacımıza uygun ve küçük bir veri kümesi ile çalışıyorsak, fazla uyum durumunun oluşması beklenir. Eğer basit bir yakınsama fonksiyonu yetersiz bir şekilde kullanılıyorsa, veri ile uygun yakınsama fonksiyonu arasındaki sapma çok büyük olabilir. Bu takdirde de eksik uyum durumu oluşabilir. Bu yüzden, deneysel riskin veya eğitim hatasının minimum tutulması, beklenen riskin veya genelleme hatasının küçük olmasını garanti edemez. Eksik uyum ve fazla uyum problemleri arasında bir denge her zaman mevcuttur. Bu denge noktası, düzenleyici bir parametre yardımıyla sağlanabilir. İstatistiksel öğrenme teorisinde, uygun fonksiyonun seçildiği fonksiyonlar kümesinin karmaşıklığı (kapasitesi) VC boyutu kullanılarak tanımlanır.

Öğrenme işleminin daha uygun bir şekilde yerine getirilmesi ve hipotez sınıfları üzerindeki PAC teorisinin genişletilmesi için VC teorisi ilk kez Vapnik ve Chervonenkis (1971) tarafından geliştirilmiştir. PAC teorisi; sabit fakat bilinmeyen bir dağılıma sahip giriş veri kümesi ve bu kümenin ikili çıkış değerlerini temel alarak, bu giriş uzayını çıkış uzayına haritalayan sınıflandırma fonksiyonunu yanlış sınıflandırma oranına bir sınır getirerek gerçekleştirilebilmesidir. VC teorisini açıklamaya başlamadan önce ayrıştırma, yükselme fonksiyonu ve VC boyutu kavramları açıklanmalıdır.



Şekil 2.2 Yükselme Fonksiyonunun Davranış Grafiği

VC boyutunu hesaplamada faydalı olan bir tanım ayrıştırma temsilidir. Eğer n tane örneklem temsilci fonksiyon kümesi tarafından mümkün 2^n durumun tümünde ayrılabiliyorsa, o zaman bu örneklem kümesinin bu fonksiyon kümesi ile ayrıştırılabildiği söylenir.

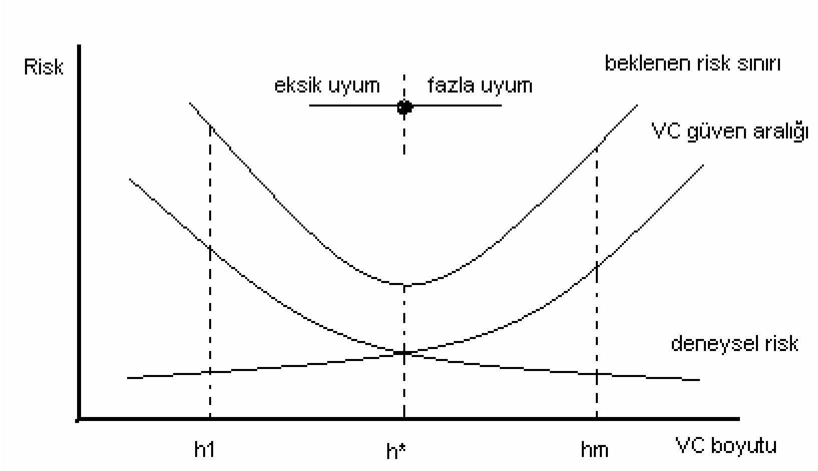
Eğer h tane örneklem bir fonksiyon kümesi tarafından ayrıştırılabiliyor fakat $h+1$ tane örneklem bu fonksiyon kümesi tarafından ayrıştırılamıyorsa, o zaman bu fonksiyon kümesinin VC boyutu h olur. Başka bir deyişle, bir fonksiyon kümesi tarafından hatasız bir şekilde mümkün tüm durumları etiketlenebilen örneklem sayısı VC boyutudur (Yang, 2002). VC boyutu DVM nin teorik bakımdan temelidir.

Hipotez uzayının karmaşıklığı veya H hipotez uzayının L sayıda örneklem üzerindeki açıklayıcı gücü bir yükselme fonksiyonu kullanılarak hesaplanabilir. Vapnik ve Chervonenkis (1971), n sayıda örnek için logaritmik bir yükselme fonksiyonu önerdiler. Şekil 2.2'deki gibi $n = h$ noktasında yükselme fonksiyonu eğrisi düşüşe geçer. Bu noktaya VC boyutu denir. Eğer onun değeri sonsuz değilse yükselme fonksiyonu yeterli örnek için doğrusal olarak artmaz (Cherkassky, 1998).

YRM prensibi genelleme hatası sınırlarını minimum tutmaya çalışır. Bu işlemi, yalnızca sonlu sayıda eğitim örneklerini kullanarak ve büyük sayıdaki aday modeller arasından en iyisini seçmek suretiyle gerçekleştirir. Bu şekilde, doğrudan gerçek hata oranını yansıtan beklenen risk fonksiyonunu minimum tutmaya çalışmak yerine YRM ve VC teorisi kullanılarak hipotez uzayındaki en uygun hipotez seçilir. Fakat DRM, sonlu sayıda ve genellikle birbirlerinden ayrık halde bulunan eğitim verilerine göre minimum tutulur. YRM prensibi, az sayıda örneklem içeren eğitme kümeleri için oldukça yararlıdır. Tersine, genellikle DRM mevcut eğitim verisi kümesinin küçük olduğu durumlara pek uygun değildir. Çünkü küçük deneysel risk beklenen riskin küçük olmasını garantileyemez. YRM prensibini temel alan istatistiksel öğrenme teorisi aynı anda hem deneysel riski hem de VC boyutunu minimum tutarak öğrenme makinesinin genelleme yeteneğini kontrol edebilir. Böylece, doğal olarak kaliteli bir yakınsama ile yakınsama fonksiyonunun karmaşıklığı arasında bir seçim yapabilme olanağı sunar. Başka bir deyişle, herhangi bir dağılıma sahip fonksiyon için YRM prensibi en iyi çözüme yakınsama garantisi verir (Vapnik 1995,1998). Ayrıca genelleme hatasının üst sınırının analiz edilmesi, hızlı bir yakınsama ve dağılımdan bağımsız DRM öğrenme şekli ile ortaya çıkan yakınsama fonksiyonlarının VC boyutu değerinin de sonlu bir değer olduğunu işaret eder. Bu, istatistiksel öğrenme açısından çok önemli bir sonuçtur.

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\frac{h(log(2ÖS/h)+1) - log(\eta/4)}{ÖS}} \tag{2.8}$$

Burada, $R(\alpha)$ gerçek riski, $R_{emp}(\alpha)$ deneysel riski, h VC boyutu değerini, $ÖS$ eğitim örneklerinin sayısını ve η, [0,1] aralığında bir değerdir. İstatistiksel öğrenme teorisi, öğrenme makinelerinin genelleme yeteneğini denklem 2.8'deki genelleme risk fonksiyonunu minimum tutarak kontrol eder. Bu işlem de fonksiyon kümesini mümkün olduğu kadar basit tutarken eğitim hatasının da minimuma ulaşması ile gerçekleştirilebilir (Cherkassky, 1998). Hiyerarşik fonksiyonlar sınıfı, $H_1 \in H_2 \in ... \in H^* \in ... \in H_m$ nin VC boyutu $h_1 \in h_2 \in ... \in h^* \in ... \in h_m$ şeklindedir ve azalan bir dizi değildir. 1. durumda, fonksiyonlar hipotezi en basit yapıda olanıdır ve büyük bir deneysel risk ile küçük bir VC terimine sahiptir. Pratikte, basit modeller yeterli temsil gücüne sahip değildirler, fakat sadece birkaç tane ayarlanabilir parametreye sahip olduklarından veriye karşı duyarlı değildirler ve kararlıdırlar. Son yani *m.* durumda, fonksiyonlar hipotezi en karmaşık yapıya sahiptir ve bu özelliğiyle de küçük eğitim hatası ile yüksek bir VC terimiyle sonuçlanır. Pratikte, güçlü yakınsama kapasitesine sahip daha karmaşık modeller daha fazla ayarlanabilir parametreyle kurulabilir, fakat bu durumda da fazla uyum problemi ortaya çıkar. Genelleme hatası sınırını minimum yapan optimum fonksiyon en basit ve en karmaşık olan fonksiyonlar sınıfının arasında bir karmaşıklığa sahiptir. Beklenen riskin (genelleme hatasının) denklem 2.8'de de gösterilen üst sınır fonksiyonu ve bu fonksiyonun düşük uyum ile yüksek uyum durumları arasındaki ilişkileri şekil 2.3'te sunulmaktadır.



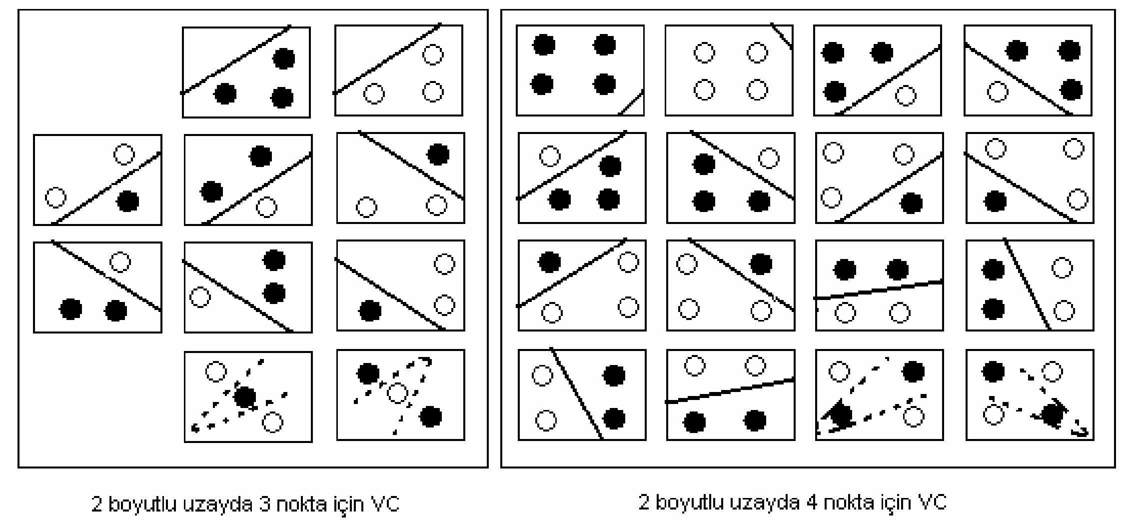
Şekil 2.3 Beklenen Risk Fonksiyonu Yapısı

YRM tümevarımsal prensibi ve VC teorisini genel olarak tüm makine öğrenmesi metotları üzerinde tartışmamız tez çalışması açısından yararlı olacaktır. Parametrelerin lineer olarak işlev gördüğü makine öğrenmesi metotlarında, VC boyutu bağımsız parametrelerin sayısına eşdeğerdir. Diğer bir ifadeyle öğrenme makinelerinin karmaşıklığı bağımsız parametrelerin sayısıyla ifade edilir. Azalmayan VC boyutlu hiyerarşik hipotez modeli yapısı, $H_1 \in H_2 \in ... \in H^{*} \in ... \in H_m$, azalmayan sayılı bağımsız parametrelere göre kurulmuş modelleri anımsatır. Her bir model kendisinden önceki daha az karmaşık yapıdaki modeli kapsar. Örneğin H_m, *m.* dereceden bir değişkenli çok dereceli fonksiyonlar kümesi, *m* farklı bulanık kuraldan oluşan bulanık bir model veya *m* tane gizli katman düğümüne sahip bir yapay sinir ağı olabilir (Kecman, 2001). Aslında hiyerarşik yapı tüm bu modellere uygulanabilecek genel bir tasarımdır. Bir taraftan hipotez uzayı H_1'den H_m'ye kadar daha yüksek dereceden bileşenler kullanarak artarken (bu durum örneğin bulanık modellerde kural sayısının artışına, çok dereceden fonksiyonlarda ise fonksiyon derecesinin artışına tekabül eder), öğrenme makinesinin kapasitesi de ilave ayarlanabilir parametrelerin kullanılmasıyla artar. Ancak diğer taraftan da, bu ilave parametrelerinin de aynı veri kümesi vasıtasıyla optimum değerlerinin bulunması gerektirir ki, bu da beklenen riskin minimuma ulaşmasını kötüye götürür. Deneysel riskin beklenen riske kararlı olarak yakınsamasını sağlamak için m, bağımsız parametre sayısındaki artışın veri kümesindeki artışla desteklenmesi gerekmektedir.

Bu yüzden VC boyutu kavramına göre kurulan bu tip modellerin hedefi, eğitme hatası ve model karmaşıklığını dengelemek amacıyla uygun sayıda bağımsız parametre içeren alt modellerden birisinin bulunmasıdır. Sinir ağları ve Bulanık modeller gibi çok dereceli fonksiyon alt yapısına sahip olarak çalışan yakınsayıcı veya sınıflandırıcılar ise; yüksek dereceli fonksiyonun optimum kuvvet değerini, yani gizli katmandaki optimum düğüm sayısını veya bulanık modeller için optimum bulanık kural sayısını bulmaya çalışırlar (Kecman, 2001). Bu tip bir optimum model karmaşıklığı seçimi beklenen risk sınırının minimuma ulaşmasını kesinleştirir. Bu durum da, zaten YRM tümevarımsal prensibinin ana fikrini yansıtır.

Bulanık modeller, sinir ağları, gauss ağları ve DVM gibi metotlar bazı koşulları sağlayan *m* tane sabit fonksiyonun birleşimi olarak temsil edilebilirler. Bu yüzden doğrusal parametrik ve m tane sabit fonksiyonun birleşiminden oluşan modeller için VC boyutu m+1 olarak düşünülür. Bu bakımdan, lineer modellerde optimum fonksiyonun bulunması lineer olmayan modellere göre daha kolaydır. Bahsedilen bu metotlar *m* tane değişken, şekli ayarlanabilen ve lineer olmayan fonksiyondan da oluşabilir. Bu tip metotlar için onların VC boyutunun hesaplanması yararlı, fakat karmaşık bir işlemdir (Kecman, 2001). Bunun sonucu olarak, bu tip metotlar öğrenme süreçlerinde yakınsama doğruluğu ile model karmaşıklığı arasında bir denge kurmak yolunu tercih ederler.

VC boyutu ile DVM tasarımında kullanılmakta olan tolerans aralıkları Destek Vektör Ssınıflandırma'da sınıfları ayıran bölge, Destek Vektör Regresyon'da ise fonksiyonun izin verilen sapma bölgesi) arasındaki bağlantı şunlara bağlıdır: VC boyutu, yakınsama fonksiyonlarının kapasitesini yansıttığından ve model karmaşıklığı için tümleşik bir ölçüt olduğundan istatistiksel öğrenme teorisinde önemli bir yer tutar. Boyutu *n* olan bir uzayda *N* tane örnek veri üzerinde çalışıldığı durumunu düşünelim. Bu örnekler 2^N mümkün şekilde etiketlenebilirler ve $f(x,\alpha)$ fonksiyonlar kümesinden bir fonksiyon tüm bu etiketlerin ayrılması için seçilebilir. Bu fonksiyonlar kümesinin VC boyutu, fonksiyonlar kümesinin üyeleri tarafından ayrılabilen eğitim örneklerinin maksimum sayısı olarak tanımlanır (Suykens ve ark., 2002). Genel olarak VC boyutunun değeri, hem $f(x,\alpha)$ yakınsama fonksiyonları kümesine hem de çözülecek olan öğrenme probleminin türüne (sınıflandırma, regresyon gibi) bağlıdır.



Şekil 2.4 2 Boyutlu Uzayda 3 Ve 4 Veri Noktası İçin VC Boyutu

Örneğin, basit bir ikili sınıflandırma durumunu düşünelim. Boyutu n olan uzay için doğrusal ayırıcı düzlemlerle sınıflandırma işlemi gerçekleştirilirken VC boyutu değeri, $h = n+1$ olur. Şekil 2.4'te N=3 ve N=4 veri örneği sayıları ve n=2 boyut değeri için mümkün olan sınıflandırma durumları gösterilmiştir. Bu şekilden de görülür ki, N=3 için mümkün 2^3=8 durumda da doğrusal ayırıcı düzlemlerle sınıflandırma yapılabilmektedir. Fakat N=4 için mümkün 2^4=16 durumun tamamında doğrusal olarak sınıflandırma yapılamaz. 16 durumdan 2 tanesinde XOR problemiyle karşılaşırız. Sonuç olarak, VC boyutu n+1 yani 3 tür. İki boyutlu uzaydaki doğrusal olmayan sınıflandırıcılar için VC boyutu daha büyük olabilir.

Aynı şekilde n boyutlu uzaydaki ayırıcı düzlemin VC boyutu da $h = n+1$ olur. Pratik uygulamaların çoğunda, örneğin sınıflandırma durumunda VC boyutunun değeri hesaplanabilir veya tahmin edilebilir. Fakat bazı uygulamalarda VC boyutunun hesaplanması hiç kolay değildir. Çünkü bilinmeyen bir değerde veya sonsuz olabilmektedir. Bu durumda iyi bir sınıflandırma için sonsuz sayıda eğitim örneğine ihtiyaç vardır.

Pratikte, VC boyutu her zaman öğrenme makinesinin bağımsız parametre sayısı ile ilişkili değildir. Bazen bağımsız parametrelerin sayısı artarken VC boyutu da artar. Bazen de, bu geçerli değildir. Örneğin sadece bir parametreye sahip bazı öğrenme makinelerinin VC boyutu bile sonsuz olabilir. Genelleme hatasının üst sınırı ise 2.8 denklemine göre VC boyutunun değeri de kullanılarak hesaplanabilir. Doğrusal fonksiyon yapısında verilen öğrenme makinelerinin VC boyutu kolaylıkla hesaplanabilirken doğrusal olmayan fonksiyonlar için hesaplanması oldukça zordur (Kecman, 2001).

2.4 Çekirdek Fonksiyonları

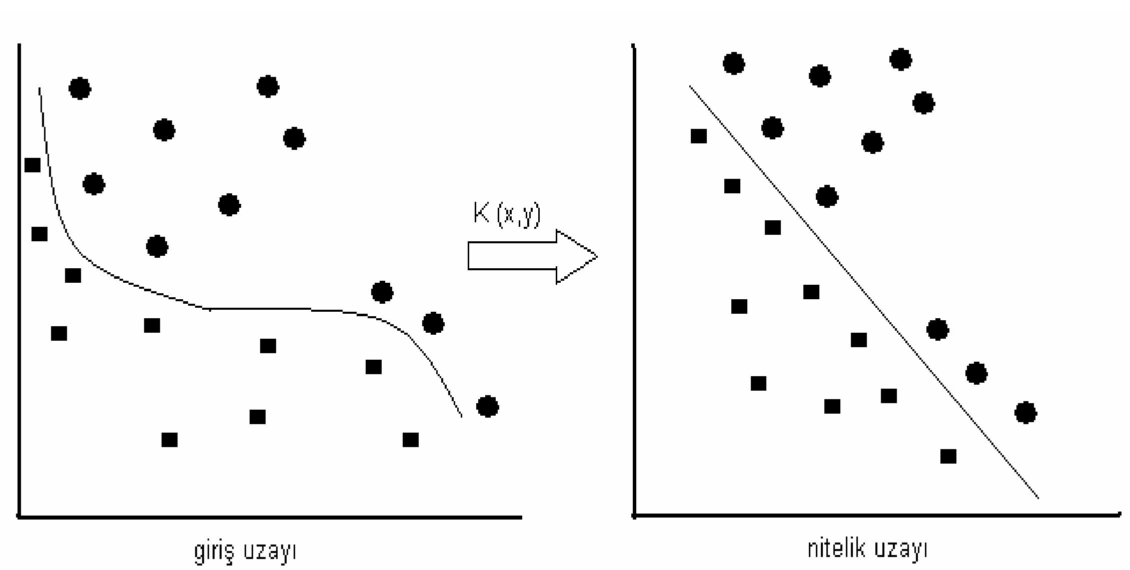
Çekirdek fonksiyonlarının (kernel functions) dayandığı temel fikir, bir takım bileşenlerin belirli bir kurala göre dönüştürülmesidir. Bu dönüşüm aslında giriş uzayında da gerçekleştirilebilecek bazı işlemlerin potansiyel çok boyutlu bir uzayda gerçekleştirilmesi üzerine kurulmuştur. Böylece çekirdek fonksiyonlarını kullanan metotlar (özellikle tezin de konusu olan DVM) giriş uzayında karmaşık temsillerde sunulabilen birçok uygulama üzerinde de çalışabilmektedirler. Doğrusal olmayan ayrılabilirlik durumu düşünüldüğü zaman, eğitim örnekleri orijinal giriş uzayında doğrusal olarak ayrılamazlar. Bu tip durumlarda DVM, doğrusal olmayan haritalama fonksiyonu yardımıyla orijinal giriş uzayından doğrusal olarak kolayca sınıflandırma yapabileceği yüksek boyutlu nitelik uzayına dönüşüm yapar. Bu yaklaşım şekil 2.5 te gösterilmektedir.

İlk bakışta çekirdek fonksiyonlarının kullanılması ek bir hesaplama maliyeti getiriyor gibi görülebilir. Fakat aslında DVM ayırıcı fonksiyonu doğrusal olarak nitelik uzayı üzerinde aramaya başladığından bu arama esnasında kayıp gibi görünen ek hesaplama maliyeti giderilir. Ayrıca, çekirdek fonksiyonlarını kullanmamızın basit nokta çarpımlarını kullanmaya göre önemli bir avantajı tam olarak neyi haritaladığımızı bilmemizdir. Çekirdek fonksiyonları kullanılırken nokta çarpımının değerlendirilmesine gerek duyulmaz. Ayrıca, boyut sıkıntısı probleminin çözüm yollarından da bir tanesidir. Doğrusal olmayan DVM sınıflandırıcı ve regresyon uygulamalarında iki temel problemle karşılaşabiliriz. Bunlardan birincisi, haritalama fonksiyonunun seçimi işlemidir. İkincisi ise, boyut sıkıntısı problemidir. Eğer giriş uzayının boyutu büyük ise, o zaman nokta çarpımı işleminde çarpım matrisinin sayısal değerlerinin oluşturulması da çok fazla zaman almaktadır. Boyuttaki bu genişleme çekirdek fonksiyonlarının kullanılmasıyla çözülebilir. Genel çekirdek fonksiyonu yapısı eşitlik 2.9'daki gibi gösterilir.

$$k(x_i, x_j) = z_i . z_j = \phi(x_i).\phi(x_j) \quad (2.9)$$

Çekirdek fonksiyonu kullanmanın temel avantajı, bütün değerlerin tekrar tekrar çarpım değerlerinin hesaplanarak bulunması yerine doğrudan çekirdek fonksiyonunda değerin yerine koyularak nitelik uzayındaki değerinin bulunmasıdır. Bu sayede, son derece yüksek boyutlu bir nitelik uzayı ile uğraşma olasılığı kalmaz. Seçilen çekirdek fonksiyonu ile DVM sonsuz boyutlu bir uzayda dahi oluşturulup çalıştırılabilir. Çekirdek fonksiyonlarının diğer bir avantajı da, eğitme aşamasında bir eğitim örneği için fonksiyon kurulup değerler bulunduktan sonra diğer örnekler için artık kalıp değerleri eğitim örneği dışında tamamen hazır olduğu için çok daha kolay hesaplanmasıdır (Kecman, 2001). Örneğin 2-boyutlu uzaydan 3-boyutlu uzaya dönüşüm yapan çekirdek fonksiyonu denklem 2.10'daki gibi olmaktadır.

$$\begin{aligned} \langle r,s \rangle &= r_1 s_1 + r_2 s_2 + r_3 s_3 \\ &= a_1^2 b_1^2 + 2a_1 a_2 b_1 b_2 + a_2^2 b_2^2 \\ &= \langle a,b \rangle^2 \end{aligned} \quad (2.10)$$



Şekil 2.5 Yüksek Dereceli Çekirdek Fonksiyonu İle Nitelik Uzayına Dönüşüm

Tablo 2.1 Yaygın Olarak Kullanılan Çekirdek Fonksiyonları

Çekirdek İsimleri	**Çekirdek Fonksiyonları**
Doğrusal	$K(x,y) = x.y$
d. dereceden yüksek dereceli fonksiyon	$K(x,y) = [(x.y)+1]^d$
Çok katmanlı algılayıcı (sadece bazı θ değerleri için tanımlıdır)	$K(x,y) = tanh(kx.y-\theta)$
Gauss Yarıçap Temelli Fonksiyon-Gauss YTF	$K(x,y) = K(x-y) = exp(-\frac{\lVert x-y \rVert^2}{2\sigma^2})$
Üssel YTF	$K(x,y) = K(x-y) = exp(-\frac{\lVert x-y \rVert}{2\sigma^2})$
Ters İkinci Dereceden	$K(x,y) = K(x-y) = (\lVert x-y \rVert^2 + c^2)^{-1/2}$
İnce Tabaka Spline	$K(x-y) = \lVert x-y \rVert^{2n+1}$ $K(x,y) = K(x-y) = \lVert x-y \rVert^{2n} ln(\lVert x-y \rVert)$
Oransal İkinci Dereceden	$K(x,y) = 1 - \frac{\lVert x-y \rVert^2}{\lVert x-y \rVert^2 + \theta}$
Dalga Yapılı	$K(x,y) = \frac{\theta}{\lVert x-y \rVert} sin(-\frac{\lVert x-y \rVert}{\theta})$
B-spline	$K(x,y) = B_{2n+1}(x-y)$
Fourier Serisi (trigonometrik çok dereceli)	$K(x,y) = \frac{sin(d+1/2)(x-y)}{sin\frac{(x-y)}{2}}$

Özellikle DVM alanında çalışan araştırmacılar için çekirdek fonksiyonları önemli bir yere sahiptir. Bir fonksiyonun çekirdek fonksiyonu olarak kabul edilebilmesi için, Mercer koşullarını yerine getirmesi gerekir (Shen, 2005). En sık kullanılan çekirdek fonksiyonları Tablo 2.1'de fonksiyon isimleri ve matematiksel ifadeleriyle birlikte gösterilmektedir.

Farklı çekirdek fonksiyonları farklı karakterlere sahiptir. Yüksek dereceden çekirdek fonksiyonları uygulamalarda tercih edilen çekirdek fonksiyonu çeşitlerinden birisidir. En sık kullanılanı ise, YTF (Yarıçap Temelli Fonksiyon, Radial Basis

Function RBF) çekirdek fonksiyonudur. Bu fonksiyon tipi yardımcı bazı tekniklerle de kullanılarak birden fazla merkez noktası üzerinde çalıştırılabilir. Tez çalışmasında gerçekleştirilen uygulamalarda YTF tercih edilmiştir. Çok katmanlı algılayıcı çekirdek fonksiyonu kullanıldığında DVM sınıflandırıcısının yapı taşları olan destek vektör noktaları, YSA çok katmanlı ağının ilk katmanındaki düğümlerinin görevini üstlenirler. Tanım aralığı olarak $[-2/\pi, 2/\pi]$ bölgesinde tanımlanan fourier çekirdek fonksiyonu nitelik uzayının $2N+1$ değerine genişletilmesi anlamına gelir.

2.4.1 Çekirdek Fonksiyonlarının Özellikleri

DVM nin önemli bir parçası olan çekirdek fonksiyonlarının birlikte kullanılması ile ilgili olarak aşağıda sıralanan özellikler mevcuttur (Schölkopf ve Smola, 2002).

- Eğer K_1 ve K_2 çekirdek fonksiyonları iseler ve $\alpha_1, \alpha_2 \geq 0$ ise, o zaman $\alpha_1 K_1 + \alpha_2 K_2$ de bir çekirdek fonksiyonudur.
- Eğer K_1 ve K_2 fonksiyonları bağımsız olarak çekirdek fonksiyonları iseler, o zaman $(K_1K_2)(x, \bar{x}) \equiv K_1(x, \bar{x})\, K_2(x, \bar{x})$ şeklinde tanımlanabilen K_1K_2 de bir çekirdek fonksiyonudur.
- Eğer K_1 ve K_2 çekirdek fonksiyonları sırasıyla X_1 ve X_2 üzerinde tanımlı iseler, o zaman $(K_1 * K_2)((x_1, x_2)(\bar{x}_1, \bar{x}_2)) = K_1(x_1, \bar{x}_1)\, K_2(x_2, \bar{x}_2)$ şeklinde tanımlanabilen bir şekilde kartezyen çarpım çekirdek fonksiyonu da $(x_1 * x_2)$ $X (x_1 * x_2)$ üzerinde tanımlıdırlar.
- Eğer K_1 ve K_2 çekirdek fonksiyonları sırasıyla X_1 ve X_2 üzerinde tanımlı iseler, o zaman $(K_1 + K_2)((x_1, x_2)(\bar{x}_1, \bar{x}_2)) = K_1(x_1, \bar{x}_1) + K_2(x_2, \bar{x}_2)$ şeklinde tanımlanabilen bir şekilde kartezyen toplam çekirdek fonksiyonu da $(x_1 * x_2) * (x_1 * x_2)$ üzerinde tanımlıdırlar.

2.5 Destek Vektör Makineleri (Support Vector Machines)

Birçok makine öğrenmesi, veri madenciliği ve istatistiksel veri analizi metodunun temel amacı, verilerden veya örneklerden öğrenmedir. Yapay sinir ağları gibi geleneksel öğrenme makineleri, örüntü tanıma, regresyon analizi ve sistemlerin kontrolü ve modellenmesi gibi çok çeşitli alanlarda kullanılmıştır. Fakat bu yöntemlerin bazı eksiklikleri mevcuttur. Bu eksikliklerin başında, yeterince eğitim verisine sahip olma ihtiyacı, düşük yakınsama oranı, yerel minimuma takılma problemi ve fazla uyum/eksik uyum (overfitting/underfitting) problemleri gelmektedir (Lu ve ark., 2002). Bu problemlerin çoğu geleneksel yöntemlerin deneysel risk minimizasyonuna bağlı olarak çalışmasından kaynaklanır. DVM, yapısal risk minimizasyonu prensibine bağlı olarak öğrenir ve bu problemlerin çoğunu böylece aşmıştır. DVM, YRM prensibinin ve VC teorisinin uygulandığı bir yakınsama metodudur. Bu özelliğiyle diğer makine öğrenmesi metotlarının çoğundan farklıdır. DVM, beklenen riskin minimuma ulaşması için hem deneysel riski hem de VC boyutunu minimum tutmaya çalışır. Fakat bu işlemi, bulanık modeller, sinir ağları ve yüksek dereceden fonksiyon kullanan yakınsayıcılar gibi VC boyutu değerini sabit tutarak gerçekleştirmez (Shen ve ark., 2004). Bu makine öğrenmesi metotlarında VC boyutunun sabitlenmesi, çok dereceli fonksiyonların derecelerinin, gizli katmandaki düğüm sayısının veya bulanık kuralların sayısının sabitlenmesiyle sağlanır. DVM, bu bahsedilen özelliğiyle genelleme yeteneğini geliştirir, küçük miktarda eğitim verisiyle öğrenebilir ve genellikle genel optimuma ulaşabilir. DVM, güçlü bir algoritmik alt yapıya sahiptir ve DVS (Destek Vektör Sınıflandırma) olarak bilinen sınıflandırma işlemleriyle DVR (Destek Vektör Regresyon) olarak bilinen tahmin işlemlerinde başarılıdır. DVM, yüksek boyutlu küçük sayıda veri içeren veri kümelerinde de başarılıdır (Shen ve ark., 2004).

DVM sınıflandırıcısı maksimum aralık sınıflandırıcılarından birisi olarak da tanınır. VC boyutu ve aralıklar arasındaki bağlantının kurulması pratik uygulamalar için gereklidir. Boyutu n olan bir uzayda eğitim örnekleri aşağıdaki formülde belirtilen yapıda bir ayırıcı fonksiyon ile birbirlerinden ayrılabilir.

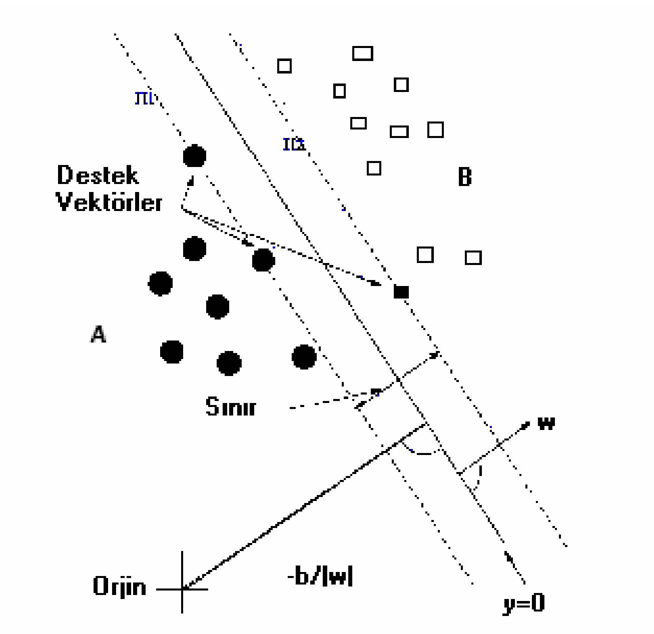
$$f(x) = w.x + b = 0 \qquad (2.11)$$

Bu fonksiyon w, ağırlık vektörünün yardımıyla hesaplanabilir. Eğitim verisinin ayrılabilir olduğu varsayılırsa ve N eğitim örneklerinin sayısını göstermek üzere, w ve b değerleri $|w.x_i + b| = 1$, i=1,...,N olacak şekilde ayırıcı düzleme en yakın veri noktaları değerlerini içeren bir formda ölçeklenebilir. Bu noktalardan bazıları eşitliği $w.x+b$=1 durumunda sağlarken bazıları da $w.x+b$=-1 durumunda sağlar. Bu ayırıcı fonksiyonun böldüğü sınırın toplam uzunluğu $2/\|w\|$ olur. Ayırıcı fonksiyonlar kümesinin VC boyutu, w (ağırlık vektörü) ile ilişkili sınırı ayarlayarak kontrol edilebilir. En yakın nokta veya noktaların ayırıcı fonksiyona uzaklığı $1/\|w\|$ dur. Farklı w vektör değerleriyle farklı ayırıcı fonksiyonlar oluşturulabilir. Ağırlık vektörü normunun küçük olması VC boyutu değerinin küçük olmasını sağlar. Bu yüzden $\|w\|$ değerinin minimum tutulması; sınır aralığının maksimum, VC boyutunun da minimum tutulması anlamına gelmektedir. DVM sınıflandırıcısına maksimum sınır sınıflandırıcısı da denmesinin sebebi budur (Cherkassky, 1998).

DVM, sınır aralıklarını maksimum yaparak YRM prensibini uygulamaya çalışır. Karmaşık fonksiyonların VC boyutu değerinin hesaplanması zor olduğundan, doğrusal fonksiyonlar için ise çok daha kolay olduğu bölüm 2.3'te açıklanmıştır. DVM yapısı da bu koşullar düşünülerek doğrusal formda inşa edilmiştir. Bu durumda karşılaşılabilen sorun eksik uyum sorunudur. DVM bu sorunu ise çekirdek fonksiyonlarını kullanarak aşmaktadır.

İstatistiksel öğrenme teorisi (YRM ve VC teorisi), Vapnik ve çalışma ekibi tarafından bulunan DVM metoduna sağlam bir teorik alt yapı sunar (Vapnik, 1995,1998). Eğitim verileri üzerinde hiç hata bulunmadığı durumlarda DVM, VC boyutu değerini minimuma ulaştırmaya çalışır. Böylece beklenen risk olasılığını azaltarak iyi bir genelleme yeteneği sağlayabilir. Eğitim verileri üzerinde bir miktar hatanın mevcut olduğu durumlarda ise DVM, tanıma doğruluğu ile öğrenme makinesinin karmaşıklığı arasında iyi bir denge sağlayan doğru VC boyutu değerinin bulunmasını amaç edinir. Bu teorik alt yapısı sayesinde, DVM pratikte karşılaşılan boyut sıkıntısı problemini de çözer.

DVM, sınıflandırma ve regresyon problemlerinde çalıştırılabilir. Sınıflandırma problemlerinin birkaç türü vardır. Bunlardan en temel olanı eğitim hatasının bulunmadığı doğrusal ayrılabilirlik durumudur. Bu durumda DVM giriş uzayı üzerinde doğrusal bir ayırıcı fonksiyon oluşturur. Bir başka durum, eğitim hatasının bulunmadığı doğrusal olmayan ayrılabilirlik durumudur. Bu durumda DVM sınıflandırıcısı giriş uzayı üzerinde doğrusal sınıflandırma işlemini gerçekleştiremez. Bu nedenle, ilk önce çekirdek fonksiyonlarının yardımıyla giriş uzayından nitelik uzayına dönüşüm yapılır. Daha sonra DVM sınıflandırıcısı nitelik uzayı üzerinde doğrusal ayırıcı fonksiyonunu oluşturur. Eğitim hatasının bulunduğu durumlarda da aynı işlemler yapılır. Fakat sisteme pozitif değerli bir esneklik parametresi eklenir. Hem sınıflandırma hem de regresyon işlemlerinde, öğrenme problemi ikinci dereceden amaç fonksiyonuna sahip bir optimizasyon problemi formunda temsil edilir. DVM regresyon metodundaki temel fikir, eldeki eğitim verilerinin karakterini mümkün olduğunca gerçeğe yakın bir şekilde yansıtan ve istatistiksel öğrenme teorisine uyan doğrusal ayırıcı fonksiyonun bulunmasıdır. Sınıflandırmaya benzer bir şekilde regresyonda da doğrusal olmayan durumların işlenebilmesi için çekirdek fonksiyonları kullanılır.



Şekil 2.6 DVM Sınıflandırıcı

Geleneksel modellerden farklı olarak, teorik yönden analiz edilmesi ve karmaşık modeller üzerinde çalıştırılması zordur. DVM nin iki belirgin özelliği; güçlü teorik alt yapısı ve pratikteki yüksek başarımıdır. DVM eğitim işlemi, ikinci dereceden bir optimizasyon probleminin çözümü gibi ele alınabilir. Çekirdek fonksiyonlarının kullanımı ile birlikte DVM yüksek boyutlu nitelik uzayında doğrusal bir modele dönüşür. Giriş uzayından daha yüksek boyutlu nitelik uzayına haritalama yapılması işlemi tüm gerekli hesaplamalar giriş uzayında yapıldığı için DVM eğitme işlemine ek bir hesap yükü getirmez. Bu yüzden DVM en karmaşık doğrusal olmayan öğrenme işlemlerinde bile basit bir öğrenme algoritmasıyla görevini yerine getirebilir.

Sınıflandırma durumunda şekil 2.6 da gösterildiği gibi, DVS nitelik uzayında mevcut eğitme verilerine göre optimum ayırma düzlemini öğrenmeye çalışır. Regresyon durumunda ise DVR, mevcut eğitme verilerine göre giriş ve çıkış uzayları arasındaki haritalama fonksiyonunu öğrenmeye çalışır. Hem DVS hem de DVR istatistiksel öğrenme teorisi çatısı altında geliştirilir ve örüntü tanıma, fonksiyon tahmini, zaman serileri tahmini, mühendislik ile biyoloji gibi birçok alana başarıyla uygulanmaktadır. DVM başlıca üç bileşenden oluşmaktadır; istatistiksel öğrenme teorisi, optimizasyon algoritması ve çekirdek fonksiyonları. Böylece, DVM çekirdek fonksiyonlarıyla dönüştürülmüş olan nitelik uzayında iyi bir genelleme teorisinin ışığında ve optimizasyon teorisinin yardımıyla doğrusal olarak eğitilen bir öğrenme makinesi olarak tanımlanabilir. DVM, yüksek boyutlu veri kümeleri üzerindeki sınıflandırma ve regresyon işlemlerinde iyi bir genelleme performansı sergilemiştir.

2.5.1 Doğrusal Olarak Ayrılabilen Veri Kümeleri İçin DVM

DVM sınıflandırma işleminde oluşturulmaya çalışılan ayırıcı düzlem, sınıflandırma probleminde eğitme verisini hatasız bir şekilde ayırma yeteneğine sahip doğrusal bir fonksiyondur (Cherkassky, 1998). İçeriğinde *n* tane örneklem barındıran bir eğitme veri seti üzerinde durulsun.

$$(x_1,y_1), \ldots ,(x_n,y_n), \;\; x \in R^d, \;\; y \in \{+1,-1\} \qquad (2.12)$$

Bu eğitme verileri uygun w ve w_0 katsayılarıyla şu ayırıcı düzlem karar fonksiyonu yardımıyla ayrılabilir;

$$D(x) = (w \times x) + w_0 \qquad (2.13)$$

Ayırma düzlemi n sayıda veri örneğinin ayrılmasını tanımlayan şu koşulları yerine getirir;

$$\begin{aligned} &(w \times x_i) + w_0 \geq +1, && \text{eğer } y_i = +1 \\ &(w \times x_i) + w_0 \leq -1, && \text{eğer } y_i = -1, \quad i = 1, \ldots ,n \end{aligned} \qquad (2.14)$$

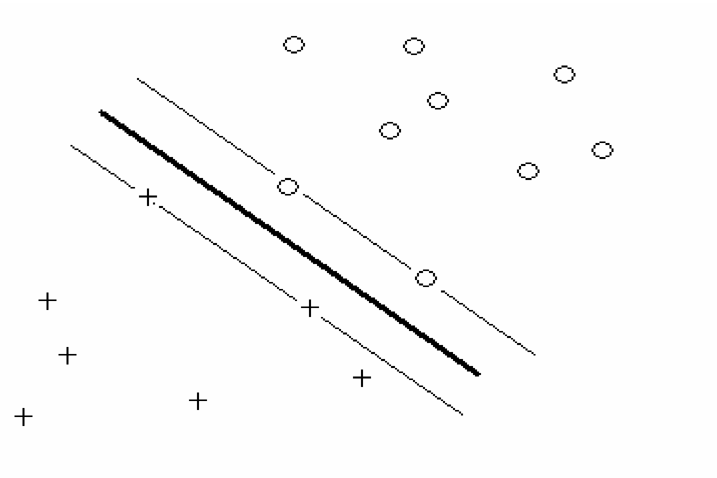
2.14 formundaki formülleri tek formüle indirilirse;

$$y_i\,[(w \times x_i) + w_0] \geq +1, \quad i = 1,\ldots,n \qquad (2.15)$$

Elde edilen 2.15 formülü 2.14 formülünün tek duruma indirilmiş halidir.

Verilen eğitme verisi kümesi için mümkün olan tüm ayırıcı düzlemler bu formda temsil edilebilir. Bu gösterim doğrudan eğitme verilerini kullandığı için önemli bir gözlemdir. Ayırıcı düzlemin formülleştirilmesi sınıflandırma problemini doğrudan çözmemize olanak sağlar (Cherkassky, 1998).

Ayırıcı düzlemden en yakın veri noktasına olan minimum uzaklığa sınır (margin) denir ve Γ ile gösterilir. Eğer sınır maksimum uzunlukta ise ayırıcı düzlem optimum ayırıcı düzlemdir. Çünkü daha büyük bir sınır daha iyi genelleme yapar (Cherkassky, 1998). Ayırıcı düzlem ve bir x' örnek verisi arasındaki mesafe $|D(x')| / \|w\|$ dur.



Şekil 2.7 Doğrusal Hatasız Ayrılabilen Veri Kümesi İçin DVM Sınıflandırıcı

Doğrusal olarak ayrılabilen verileri içeren bir sınıflandırma problemi için,

$$\frac{y_k \times D(x_k)}{\|w\|} \geq \Gamma, \qquad k = 1,\ldots,n \text{ ve } y_k \in \{-1,+1\}, \tag{2.16}$$

koşulu tüm eğitme örnekleri tarafından sağlanır.

Optimum ayırıcı düzlemin bulunması problemi Γ sınırını maksimum yapan w değerinin bulunması işlemine dönüşür. Ağırlık (w) değerinin alabileceği sonsuz sayıda farklı çözüm değeri bulunmaktadır. Bu çözümleri kısıtlamak için Γ ve $\|w\|$ değerleri $\Gamma\ \|w\| = 1$ olacak şekilde sabitlenir. Bu yüzden w değerinin minimumlaştırılması Γ değerinin maksimumlaştırılmasına eşdeğerdir. Optimum ayırıcı düzlem 2.15 formülündeki koşulları sağlar ve aşağıdaki eşitlik değerini minimum yapar.

$$\eta(w) = \|w\|^2 \tag{2.17}$$

Sınır, ayırma düzleminin genelleme yeteneğiyle doğrudan ilişkilidir. Daha büyük sınır sınıflar arasında daha fazla ayırma yapar. Sınır üzerinde oluşan veri noktaları (bu noktalar aynı zamanda 2.15 formülünü eşitlik halinde sağlar) destek vektörler adını alır. Destek vektörler karar yüzeyine en yakın veri noktaları olduğu için en zor sınıflandırılan ve karar yüzeyinin yapısını belirleyen veri noktalarıdır.

Örnek: XOR probleminin DVM de çözümü.

Çözüm:

$$\Phi(w) = \tfrac{1}{2} w^T w$$

$$L(w,b,\alpha) = \tfrac{1}{2} w^T w - \sum_{i=1}^{N} \alpha_i [d_i (w^T x_i + b) - 1]$$

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i \; - \tfrac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j x_i^T x_j$$

$$k(x, x_i) = (1 + x^T x_i)^2$$

$$= 1 + x_1^2 x_{i1}^2 + 2x_1 x_{i1} x_2 x_{i2} + x_2^2 x_{i2}^2 + 2x_1 x_{i1} + 2x_2 x_{i2}$$

$$\varphi(x) = [1, x_1^2, \sqrt{2} x_1 x_2, x_2^2, \sqrt{2} x_1, \sqrt{2} x_2]^T$$

$$Q(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \tfrac{1}{2}(9\alpha_1^2 \; - 2\alpha_1\alpha_2 - 2\alpha_1\alpha_3 + 2\alpha_1 + 9\alpha_2^2 + 2\alpha_2\alpha_3 - 2\alpha_2\alpha_4 + 9\alpha_3^2$$
$$- 2\alpha_3\alpha_4 + 9\alpha_2^4)$$

$$9\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 = 1$$
$$-\alpha_1 + 9\alpha_2 + \alpha_3 - \alpha_4 = 1$$
$$-\alpha_1 + \alpha_2 + 9\alpha_3 - \alpha_4 = 1$$
$$\alpha_1 - \alpha_2 - \alpha_3 + 9\alpha_4 = 1$$

$$\alpha_{o,1} = \alpha_{o,2} = \alpha_{o,3} = \alpha_{o,4} = \tfrac{1}{8}$$
$$Q_o(\alpha) = \tfrac{1}{4}$$

Qo(α) = 1/4 = 1/2||wo||2

w0 = 1/8[- Ψ(x1) + Ψ(x2) + Ψ(x3) - Ψ(x4)] = [0 0 -1/ 0 0 0]T

b biası, w_0'ın ilk elemanıdır ve 0' a eşittir.

Optimum ayırıcı düzlemin genelleme kabiliyeti, destek vektörlerinin sayısıyla doğrudan ilişkilendirilebilir. Vapnik'e göre, destek vektör sayısı test örnekleri için beklenen hata oranı sınırını şu eşitliğe göre etkiler (Cherkassky, 1998).

$$E_n[\textit{hata oranı}] \leq E_n[\textit{Destek vektör sayısı}] / n \qquad (2.18)$$

Burada, E_n tüm eğitme verisi üzerindeki beklentiyi gösterir. Paydada bulunan *n* ise eğitim örneği sayısıdır. Bu sınır uzayın boyutundan bağımsızdır. Küçük sayıda destek vektör ile yapılandırılmış bir optimum ayırıcı düzlem olduğunu varsayılırsa; o yüksek boyutlu uzayda bile iyi bir genelleme kabiliyetine sahip olacaktır (Cherkassky, 1998).

DVM metodunun geliştirilmesinde ayırıcı düzlem çalıştırılacağı için onun VC boyutu da yakınsama fonksiyonlarının katmanlı bir yapısını oluşturmak için belirlenmelidir. Bir örneklem kümesinin herhangi bir alt kümesi daha küçük VC boyutuna sahip olabilir. Vapnik (1995), böyle alt kümeler için bir sınır getirmiştir. $\|w\|^2 \leq c$, eşitliğini doğru olarak sağlayan ayırıcı düzlem fonksiyonları için VC boyutu denklem (2.19)'da verilmiştir.

$$h \leq min(r^2 \times c, d) + 1 \tag{2.19}$$

Burada bulunan *r*, eğitme girişini içeren en küçük dairenin yarıçapıdır ve *c* değerinin ölçeklendirilmesi için kullanılır. Ayırıcı düzlemin karmaşıklığının örneklem uzayının boyutundan bağımsız bir şekilde doğrudan kontrol edilmesi mümkündür. $\|w\|^2$ normunu kontrol ederek karmaşıklığın artışına göre ayırıcı düzlem yapısı kurulması mümkündür.

$$S_k = \{ \frac{(w \times x) + w_0}{\|w\|^2} \leq c_k \}, \qquad c_1 < c_2 < c_3 \ldots < c_k \tag{2.20}$$

YRM prensibine göre, iyi bir genelleme kabiliyeti sağlamak için garanti riski minimumlaştıran fonksiyon seçilmelidir. Sınıflandırma problemleri için garanti risk, deneyimsel risk ve güvenlik aralığının toplamıdır.

$$R(w) \leq R_{emp}(w) + \Phi \tag{2.21}$$

Tanımdaki varsayıma göre, ayırıcı düzlem her zaman sıfır deneyimsel riske (verinin doğrusal olarak ve hatasız ayrılabildiği durumlar için söz konusu) sahiptir. Bu yüzden garanti risk güvenlik aralığının minimum tutulmasıyla minimumlaştırılır. Güvenlik aralığı VC boyutu değerinin *(h)* minimum tutulmasıyla minimum hale gelir. Bu da 2.19 formülündeki $\|w\|^2$ değerinin minimum tutulmasıyla ilgilidir. Bu değerlendirmelere göre, minimum karmaşıklığa dolayısıyla maksimum genelleme kabiliyetine sahip ayırıcı düzlemin maksimum sınır değerine sahip olan ayırıcı düzlem olduğunu görülmektedir. Doğrusal olarak ayrılabilir durumlar için optimum ayırıcı düzlemin bulunması doğrusal sabitli ikinci dereceli optimizasyon problemidir.

$$\eta(w) = ½\,\|w\|^2\,, \tag{2.22}$$

eşitliğini

$$y_i\,[(w \times x_i) + w_0] \geq +1, \quad i = 1, \ldots ,n \tag{2.23}$$

formülüne uygun olarak minimumlaştıran w ve w_0 hesaplanır.

Boyutu d olan bir giriş uzayında bu problemin çözümü d+1 bağımsız parametreden oluşur. Bu optimizasyon problemi, ikinci dereceden programlama tekniğinin kullanılmasıyla çözülebilir. Çok yüksek boyutlu uzaylar için problemin şu anki formda çözülmesi pratik değildir, fakat bu problem ikili forma dönüştürülüp çözülebilir. Eğer maliyet fonksiyonu ve sabit fonksiyon tamamen dışbükey ise optimizasyon teorisi ikili formda bulunan bir optimizasyon problemine konumlandırılır. Bu durumda ikili problemin çözülmesi orijinal çözüme eşdeğerdir. Denklem 2.22 ve 2.23'te belirtilen optimizasyon problemi bu kıstasları gerçekleştirir ve bu yüzden ikili forma sahiptir. Bu durumda Kuhn-Tucker teoremi problemin ikili forma dönüştürülmesi amacıyla kullanılır (Cherkassky, 1998). Optimum ayırıcı düzlem probleminde bu teorem n örneklem sayılı olan ve d boyutlu olmayan ölçeklemelerle ikili optimizasyon problemi üretir. Bu yüzden çok yüksek boyutlu (milyon veya 10 milyon gibi) ve örnek boyutlarının değişebileceği problemlerde çözüm elde edilmesi için standart ikinci dereceden optimizasyon teknikleri kullanılabilir. 2.22 ve 2.23 denklemlerini iki adımda ikili forma dönüştürülür.

İlk adımda Lagrange çarpanlarını kullanılmaktadır;

$$Q(w, w_0, \alpha) = 1/2(w \times w) - \sum_{i=1}^{n} \alpha_i \{y_i\,[(w \times x_i) + w_0] - 1\} \tag{2.24}$$

Buradaki α_i değerleri Lagrange çarpanlarıdır. Bu fonksiyonun sırt noktası optimizasyon problemi için çözüm sağlamaktadır. Fonksiyon w ve w_0 açısından düşünüldüğünde minimumlaştırılmalı, $\alpha_i \geq 0$ açısından düşünüldüğünde ise maksimumlaştırılmalıdır.

İkinci adımda ise; 2.24 eşitliğindeki w ve w_0 parametrelerinin sadece α_i parametresi cinsinden ifade edilmesini sağlayacak olan Kuhn-Tucker koşulları kullanılır. O zaman, 2.24 formülü sadece α_i Lagrange çoğullayıcılarına göre maksimumlaştırılması istenen ikili problem olmaktadır. Kuhn-Tucker teoremine göre w^*, w_0^* ve α^* aşağıda belirtilen koşulları sağlamalıdır;

$$\frac{\partial Q(w^*, w_0^*, \alpha^*)}{\partial w_0} = 0 \qquad (2.25)$$

$$\frac{\partial Q(w^*, w_0^*, \alpha^*)}{\partial w} = 0 \qquad (2.26)$$

Bu türevlerin çözülmesi optimum ayırıcı düzlem ile ilgili aşağıdaki özellikleri vermektedir;

$$\sum_{i=1}^{n} \alpha_i^* y_i = 0, \qquad \alpha_i^* \geq 0, \qquad \text{i= 1, ... ,n ,} \qquad (2.27)$$

$$w^* = \sum_{i=1}^{n} \alpha_i^* y_i x_i, \qquad \alpha_i^* \geq 0, \qquad \text{i= 1, ... ,n ,} \qquad (2.28)$$

Kuhn-Tucker teoreminin yanı sısra, α_i^* parametreleri bir koşulu daha sağlamalıdır. Eğer veri örneği (x_i, y_i), 2.23 denklemini eşitlik halinde sağlıyorsa, başka bir ifadeyle destek vektör noktalarından birisi ise α_i sıfırdan farklı olmalıdır. Bunun formülsel olarak tanımı aşağıdaki gibidir.

$$\alpha_i^* [y_i(w^* \times x_i + w_0^*) - 1] = 0, \qquad \text{i=1, ... ,n ,} \qquad (2.29)$$

2.23 denklemini eşitlik halinde sağlayan veri örnekleri destek vektörlerdir. Destek Vektörler için α_i*, sıfırdan farklı bir değer alır. İkili problemin yapılandırılması için 2.27 ve 2.28 formüllerini kullanarak w ve w_0 parametrelerinin Lagrange ile yerleri değiştirilebilir. 2.24 eşitliği yeniden yazılırsa;

$$Q(w,w_0,\alpha) = 1/2(w \times w) - \sum_{i=1}^{n} \alpha_i y_i x_i \times w - w_0 \sum_{i=1}^{n} \alpha_i y_i + \sum_{i=1}^{n} \alpha_i \qquad (2.30)$$

2.27 koşul fonksiyonuna göre eşitliğin üçüncü terimi sıfırdır. 2.28 koşul fonksiyonunu Lagrange'da da kullanılırsa;

$$Q(\alpha(= -1/2 \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \times x_j) + \sum_{i=1}^{n} \alpha_i \qquad (2.31)$$

2.31 ifadesi ikili optimizasyon problemi fonksiyonudur. Bu fonksiyon $\alpha_1, \ldots ,$ α_n parametrelerine göre maksimumlaştırılmalıdır. 2.28 formülünü kullanarak 2.13 karar fonksiyonunu yeniden yazılırsa aşağıdaki formülü elde edilmektedir.

$$D(x) = \sum_{i=1}^{n} \alpha_i^{*} y_i (x \times x_i) + w_0^{*} \qquad (2.32)$$

Sonra destek vektör üzerindeki koşullar yardımıyla w_0* hesaplanır. (x_s, y_s) ile verilen bir destek vektör aşağıdaki eşitliği sağlar.

$$y_s[(w^* \times x_s) + w_0^*] = 1 \qquad (2.33)$$

2.28 denklemini de kullanarak;

$$w_0^{*} = y_s - \sum_{i=1}^{n} \alpha_i^{*} y_i (x_i \times x_s) \qquad (2.34)$$

Böylece Lagrange ve Kuhn-Tucker koşullarını kullanan ikili problemin yapılandırılması tamamlanmış olur. Özet olarak;

$$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - 1/2 \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \times x_j) \tag{2.35}$$

eşitliği

$$\sum_{i=1}^{n} \alpha_i y_i = 0, \qquad \alpha_i \geq 0, \quad \text{i= 1, ... ,n} \tag{2.36}$$

eşitliğine bağlı olarak verilen (x_i,y_i), i=1, ... ,n eğitme verisi kümesi maksimumlaştırılmalıdır. Optimum ayırıcı düzlem 2.32 fonksiyonu ile verilir. Buradaki α_i* ikili problemin çözümü ve w_0* 2.34 formülünün çözümü ile bulunur. Buradan görülür ki, 2.35 formülünün optimizasyonu ve 2.32 formülünün değerlendirilmesi giriş veri vektörleri arasındaki $(x * x')$ dâhili çarpım hesabına ihtiyaç duymaktadır. Gerçek dünya verilerinde destek vektör sayısı eğitme verisinin çok küçük bir yüzdesidir. Pratikte, bu optimizasyon problemi standart ikinci dereceden programlama metotlarıyla çözülebilir.

2.5.2 Lineer Olarak Belirli Oranda Hata ile Ayrılabilen Veri Kümeleri için DVM

Eğitme verisinin hatasız olarak ayrılamadığı durumlarda verilerin minimum sayıda hata ile ayrılması tercih edilir. Aynı zamanda veri noktalarını doğrusal yapıda ayıran optimum ayırıcı düzlemi de kullanılmalıdır. Ayırıcı düzlem yapısı düşünüldüğünde ise, veri noktası 2.15 formülünü sağlamadığı zaman ayırıcı fonksiyon tarafından ayrılamamaktadır. Bu durum, veri noktasının ayrım bölgesi içine düştüğünü veya yanlış sınıf etiketli bölgeye atandığını gösterir. Dikkat edilirse, ayrılamazlığın tanımı yanlış sınıflandırmanın tanımından farklıdır (Cherkassky, 1998). Yanlış sınıflandırma ile yüzleştiğimizde, veri noktasının karar düzleminin yanlış tarafına düştüğü görülmektedir. Bu durumda, ayırıcı düzleminin koşulları tanımlarken ξ_i, i=1, ... ,n pozitif gevşek değişkeni de koşul fonksiyonlarına eklenir.

$$y_i[(w \times x_i) + w_0] \geq 1 - \xi_i \tag{2.37}$$

Bir x_i eğitim örneği için, ξ_i esnek değişkeni uygun olan sınıfın sınırdan olan izin verilen sapma miktarıdır. Bizim tanımımıza göre, gevşek değişkenlerin değeri ayrılamazlık durumuna uygun olarak 0'dan büyük 1'den küçüktür ve gevşek değişkenlerin değeri yanlış sınıflandırma durumuna uygun olarak ta 1'den büyüktür.

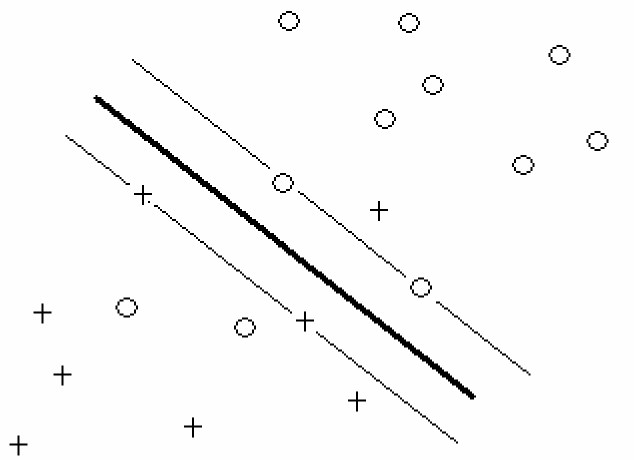
Ayrılamaz örnek sayısını minimumlaştırmak;

$$Q(\xi) = \sum_{i=1}^{n} \xi_i^{\,p} \tag{2.38}$$

eşitliğini minimumlaştırmaya eşdeğerdir. Her ne kadar (2.38) denklemi o zaman sadece ayrılamaz örnek sayısına yakınsa da, bu problemin kolaylıkla çözülmesi için p=1'e ayarlanacaktır. Bu değeri 1 alarak (2.38) eşitliğini (2.37) eşitliğine bağlı olarak minimum yapan ve 2.39 yapısını kullanan ayırıcı düzleme esnek sınır ayırıcı düzlemi denir (Cherkassky, 1998).

$$S_k = \{(w \times x) + w_0 : \|w\|^2 \leq c_k\} \tag{2.39}$$

Esnek sınırın (2.38) eşitliğiyle karşılaşınca ikinci dereceden optimizasyon teriminde bir problem ortaya çıkması muhtemel olsa da, esnek sınır ayırıcı düzlemini bulan optimizasyon problemi dışbükeydir. Bu ayırıcı düzlem aşağıdaki fonksiyonu minimumlaştıran w ve w_0 ile ayrıca 2.37 formülünü de sağlamakta ve yeterli büyüklükteki sabit C değeri ile tanımlanmaktadır (Cherkassky, 1998). Bu formdaki C sabiti, fonksiyon karmaşıklığı ile sınıflandırılamaz örnek sayısının azaltılması arasındaki tercih oranını etkiler. Eğer optimizasyon problemi yüksek boyutlu uzayda çözülüyorsa ikili formuna da dönüştürülmelidir. Bu safhada yapılması gereken işlemler bölüm 2.5.1'de verilen ayrılabilir durumdaki veriler üzerinde gerçekleştirilen işlemlerle aynıdır. Optimum esnek sınır;



Şekil 2.8 Doğrusal Hata İle Ayrılabilen Veri Kümesi İçin DVM Sınıflandırıcı

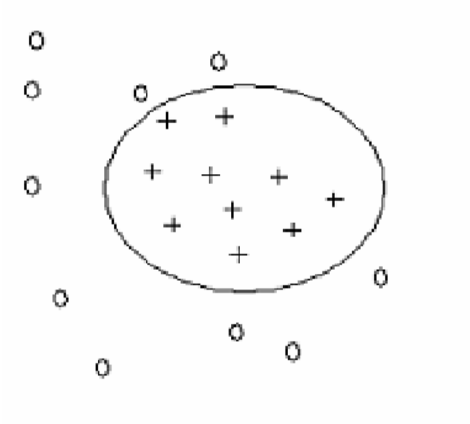
$$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - 1/2 \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \times x_j) \tag{2.40}$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0, \qquad 0 \le \alpha_i \le \frac{C}{n}, \qquad \text{i= 1, ... ,n} \tag{2.41}$$

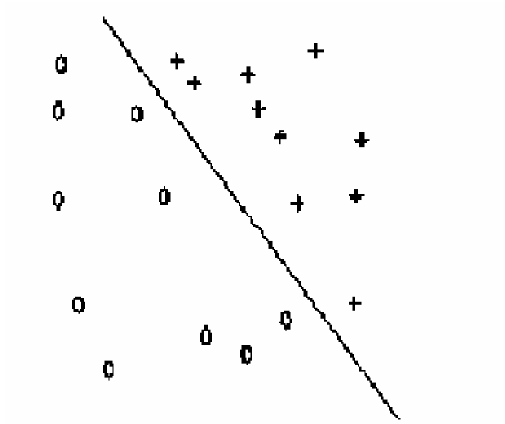
denklemine bağlı olarak bulunur. Ayırıcı düzlem karar fonksiyonu da tam olarak ayrılabilir durumdakiyle aynıdır;

$$D(x) = \sum_{i=1}^{n} \alpha_i^{*} y_i (x \times x_i) + w_0^{*} \tag{2.42}$$

(2.41) denkleminde öncekilerden farklı olarak sadece *C/n* ifadesi eklenmiştir. Bu ifade α_i çarpanları için üst sınır değerini temsil etmektedir.



Şekil 2.9 Lineer Olmayan Giriş Uzayı



Şekil 2.10 Lineer Hale Dönüştürülmüş Nitelik Uzayı

2.5.3 Lineer Olmayan Veri Kümeleri için DVM

Doğrusal olarak ayırt edilebilen veri kümelerini ayırabilen fonksiyon hipotez uzayları pratik uygulamalar için oldukça kısıtlıdır. Çünkü pratik uygulamaların çoğunu doğrusal DVM kullanılarak gerçekleştirilememektedir. Bu durumda uzaylar arası dönüşüm işlemini gerçekleştiren bazı dönüşüm fonksiyonları kullanılabilir. Bu nedenle önceki sonuçların düzenlenmesi gerekmektedir. Bu örnek kümelerini doğrusal yöntemlerle ayırmak mümkün olmadığı için, uzaylar arası dönüşümü doğrusal olmayan yöntemlerin kullanılarak gerçekleştirilmesi gerekmektedir. Bunlardan örneğe uygun olan çekirdek fonksiyonunu iç çarpım kurallarını yerine getirmek için seçilmelidir (Sanchez, 2003).

Çekirdek fonksiyonu ile giriş veri uzayı *N*'i yüksek boyutlu nitelik uzayı *H*'e haritalama işleminden sonra, 2 sınıflı sınıflandırma işlemi için DVM teknikleri çalıştırılır. Örneğin YTF çekirdek fonksiyonunun işlevi, şekil 2.9 ile gösterilen giriş uzayından şekil 2.10 ile gösterilen nitelik uzayına dönüşüm yapmaktır.

Lineer olmayan DVM için ana fikir, uygun bir çekirdek fonksiyonu belirleyerek *H* nitelik uzayına veri haritalama algoritmasının işletilmesi ve sonra da orijinden maksimum sınırla haritalanmış vektörleri ayıran düzlemin bulunmasıdır. Bunu gerçekleştirmek için şu yol izlenebilir;

- Eğitme probleminde dönüşüm için görünen tek temsil yolu (X_i, X_j) nokta çarpımları biçimidir.
- Daha yüksek boyutlu uzayda (doğrusal olmayan öğrenme işlemi için belirleyici olan nitelik uzayı) ayırıcı fonksiyon, doğrusal ayırıcı yapısına benzer şekildedir yani yüksek boyutlu nitelik uzayında doğrusal yöntemler kullanılabilir.
- R^n giriş uzayında bulunan veri noktalarını daha yüksek boyutlu nitelik uzayına haritalanmaktadır.
- Bu durumda, eğitme algoritması sadece *H*'deki çarpım noktalarına bağlı olacaktır. Bu çarpım noktaları da $\Phi(X_i)* \Phi(X_j)$ fonksiyon biçiminde gösterilir. Başka bir deyişle, $F(x) = w \times x = \sum_{i=1}^{n} \alpha_i \times y_i \times x_i \times x + b$ DVM eğitim

formülündeki x_i ve x yerine $\Phi(x_i, x)$ şeklinde bir çekirdek fonksiyonu kullanılır.

- Fakat geçiş operatörü Φ' nın hesaplanması zordur.
- Eğer $K(X_i, X_j) = \Phi(X_i) * \Phi(X_j)$ gibi *K* çekirdek fonksiyonu olsaydı eğitme algoritmasında sadece *K* kullanmaya ihtiyacımız olmaktadır.
- Örneğin; $K(X_i, X_j) = e^{-/Xi - Xj / 2} / 2\sigma^2$ çekirdek fonksiyonu gibi bir yapıda olabilir.
- Önceki doğrusal DVM çıkarımlarının tümünde çarpım noktaları yerine çekirdek fonksiyonu kullanılarak elde edilen nitelik uzayında yine doğrusal ayırma işlemi yardımıyla öğrenme işlemleri yapılabilir.
- Doğrusal olmayan uzaydaki öğrenme işlemlerinde en son adım ise, eğitim verisinin daha yüksek boyutlu nitelik uzayına Φ fonksiyonu ile haritalanması ve oradaki maksimum sınırlı ayırıcı düzlemin yapılandırılmasıdır.
- Bu işlemler sonucunda, giriş uzayındaki karakteristiği doğrusal olmayan bir veri kümesi için karar sınırının bulunmasını sağlayacak ortam oluşturulur. Çekirdek fonksiyonlarının kullanılmasıyla, nitelik uzayına tam olarak haritalama yapma dışında onun doğrusal ayırma düzlemini hesaplamamız da mümkün olur.

En yaygın kullanılan çekirdek fonksiyonları çeşitleri tez çalışmasının 2.4 bölümünde sunulan Tablo 2.1'de gösterilmektedir.

2.5.4 DVM de Çok Sınıflı Sınıflandırma

DVM sınıflandırıcılar orijinal olarak iki sınıf etiketli durumlar düşünülerek eğitilirler. Çünkü DVM iki sınıfı birbirinden doğrudan ayırabilen bir ayırıcı fonksiyonun araştırılması üzerine kurulmuş olan bir yöntemdir. DVM eğitimi için pozitif veya negatif sınıfa ait olan iki çeşit veriye eğitim aşamasında ihtiyaç duyulur. Çok sınıflı sınıflandırma durumunu içeren uygulamalarda çıkış uzayı, $y_i \in \{-1,1\}$, $i = \{1,\dots,k\}$ şeklinde sınıf etiketleriyle birlikte temsil edilir ve çözüm için birden çok sayıda ayırıcı fonksiyonun bulunması gerekmektedir. Bu tip sınıflandırma işlemlerinin gerçekleştirilebilmesi için literatürde iki temel yaklaşım bulunmaktadır (Kikuchi ve Abe, 2005). Bunlardan birincisi bütün sınıf etiketlerinin doğrudan ayrıştırılmasıdır. İkincisi ise standart ikili ayrıştırma planına göre çalışan dolaylı ayrıştırmalardır ki ayrıştırmada, belirli bir sayıda sınıflandırıcı tüm eğitme örnekleri üzerinde eğitilirler. Örneğin, bire karşı kalanlar (1-v-r, one versus remains) prensibinde sınıflandırıcı i. sınıfın örneklerine +1 etiketi ve diğer tüm sınıflara ait örneklere de -1 etiketi atamalıdır (Angulo ve ark., 2003). Başka bir genel ayrıştırma metodu k sınıf etiketli eğitme kümesinden mümkün tüm $L = k\ (k\text{-}1)\ /2$ tane ikili sınıflandırıcı DVM formunu yapılandırır daha sonra da, her bir destek vektör sınıflandırıcısı sadece iki çıkış üzerine eğitilir. Bu metotla eğitilmiş sınıflandırıcılar bire karşı bir (1-v-1, one versus one) veya akıllı çiftler (pairwise) isimleriyle adlandırılır. ECOC (Error Correcting Output Codes - Hata Düzeltme Çıkış Kodları) yaklaşımı da bu alana uygulanabilir (Angulo ve ark., 2003). Bu yaklaşımda ise, veriler arası benzerlik ilişkisini yansıtan bir uzaklık matrisi kullanılmaktadır. Aşağıdaki bölümde bu ikili sınıflandırıcıları çoklu sınıflı sınıflandırıcılara genişleten metotları ayrı başlıklar halinde daha detaylı bir şekilde incelenecektir.

2.5.4.1 Doğrudan Çok Sınıflı Sınıflandırma

Çok sınıflı etiketler üzerinde doğrudan sınıflandırma yapılabilmesi için karar fonksiyonun tüm etiketleri bağımsız olarak ayırt edebilmesi gerekmektedir. Bu tip bir planlama Crammer ve Singer (2001) ile Weston ve Watkins (1998) tarafından, r indisiyle doğrusal ayırıcı fonksiyonları tanımlayan (2.43) eşitliğiyle gerçekleştirilmiştir.

$$F_r(x) = \langle w_r, x \rangle \tag{2.43}$$

Karar fonksiyonu ise, şu şekilde ifade edilebilir.

$$f(x) = \underset{r=1,\ldots,k}{argmax} \; F_r(x) \tag{2.44}$$

Bu fonksiyon her bir $F_r(x)$ sınıfı için ayrı bir değer hesaplar ve en yüksek değere sahip sınıfı karar sınıfı atar. DVM sınıflandırıcısını çok sınıflı sınıflandırmaya uyarlayan herhangi bir eğitim örneğinin sınır değeri eşitlik 2.45'teki gibi tanımlanır.

$$\gamma_i = F_{yi}(x_i) - \max_{y \neq yi} F_y(x_i) \tag{2.45}$$

Burada sunulan formül herhangi bir eğitim örneği açısından onun ait olduğu doğru sınıf ile ait olmadığı yanlış sınıflar arasındaki en yüksek değere sahip olanının farkını yansıtmaktadır. Farklı k tane sınıf etiketi düşünüldüğünde ikinci dereceden optimizasyon formuna şu şekilde eklenebilir.

$$F_{yi}(x_i) - F_y(x) = \langle w_{yi} - w_y, x \rangle \geq 1, \qquad \forall y \neq y_i, \tag{2.46}$$

Optimizasyon problemi bu ilavelerle yapılandırılarak çözüldüğü zaman çok sınıflı sınıflandırma problemi çözülmüş olmaktadır.

2.5.4.2 Dolaylı Çok Sınıflı Sınıflandırma

Bu bölümde kısım 2.5.4 de kısaca bahsedilmiş olan dolaylı çok sınıflı sınıflandırma yaklaşımları ele alınmaktadır.

- *Çok sınıflı sınıflandırıcılar için bire karşı kalanlar (1-v-r) prensibi* **:** Vapnik tarafından 1998 de önerilen bir yöntemdir. Bu metottaki ana fikir, metodun her bir çalıştırılma zamanında bir sınıf etiketine ait olan bireylerin diğer tüm sınıf etiketlerine ait olan bireylerden ayrılması yaklaşımı üzerine kuruludur. Veri kümesinin büyük olduğu durumlarda her sınıfın ayrımında ayrıca eğitim yapıldığından fazla eğitim zamanı tüketebilir.
- *Çok sınıflı sınıflandırıcılar için bire karşı bir (1-v-1) veya akıllı çiftler (pairwise) prensibi* **:** Bu teknikte k farklı sınıf etiketi bulunan sınıflandırma uygulaması için $k*(k-1)/2$ sayıda destek vektör makinesinin eğitilmesi gerekir. Eğitme zamanı açısından bire karşı kalanlar metoduna tercih edilebilir. Çünkü her eğitim aşamasında sadece iki farklı sınıf verisinin alınması yeterlidir. Sınıf etiketlerinin sayısına bağlı olarak kullanılacak destek vektör makinesi sayısı değişir. Örneğin üç sınıflı sınıflandırma problemi için birinci sınıflandırıcı, bir ve iki etiketli sınıfları birbirinden ayırır. İkinci sınıflandırıcı, bir ve üç etiketli sınıfları birbirinden ayırır. Üçüncü sınıflandırıcı ise, iki ve üç etiketli sınıfları birbirinden ayırır. Bu metotta destek vektör makinelerinin bölüm 2.5.1'de belirtilen karar fonksiyonu, (2.13) denklem (2.47) biçimine dönüşür.

$$D_{ij}(x) = (w_{ij}.x) + w_{0ij} \tag{2.47}$$

K sayıda sınıftan her birinin karar fonksiyonu değeri 2.48 ve 2.49 eşitliklerine göre hesaplanır.

$$D_i(x) = \sum_{j \neq i, j=1}^{k} sign(D_{ij}(x)) \tag{2.48}$$

Böyle hesaplamalarda $D_{ij}(x) = - D_{ji}(x)$ kuralı da geçerlidir.

$$sign(x) = \begin{cases} 1, x > 0 \\ 0, x \leq 0 \end{cases} \tag{2.49}$$

$$\underset{i=1,\dots,k}{argmax} \; D_i(x) \tag{2.50}$$

D_i sonuç değerleri (2.50) eşitliğiyle hesaplanır. Bu sonuçların değerleri her bir sınıf için birbirine eşit değil ise, veri en yüksek değere sahip olan sınıf etiketine atanır. Eğer eşitse, o zaman ilgili verinin sınıflandırılamayan bölgeye düştüğü anlaşılmaktadır. Bu durum DVM ile dolaylı çok sınıflı sınıflandırma prensiplerinin tamamında karşılaşılan bir sorundur. Sorunun aşılması için dolaylı prensiplerin yanında ilave bir yaklaşım kullanılması ya da doğrudan çok sınıflı sınıflandırma prensibinin tercih edilmesi gerekmektedir. Bu problemin çözümü için ise, literatürde bulanık mantık metodunu kullanan bir çalışma yapılmıştır. Söz konusu çalışmada da, KEYK metodunu kullanan bir yöntem geliştirilmiştir (Çomak ve Arslan, 2006).

- *Çok sınıflı sınıflandırıcılar için Hata Düzeltme Çıkış Kodları (ECOC) prensibi* **:** Bu teknik, Dietterich ve Bakiri (1995) tarafından önerilmiştir. Bu teknikteki temel fikre göre, öncelikle sınıflandırılacak bir test örneği için destek vektör makineleri ayrı ayrı eğitilir. Daha sonra L sayıda $f_1,\dots,f_L$ çıkışı hesaplanır. C kod değerleri aşağıdaki fonksiyonun kullanılmasıyla hesaplanır.

$$Ci = \begin{cases} 1, \rightarrow fi \geq 0 \\ 0, \rightarrow f_i < 0 \end{cases} \tag{2.51}$$

Hesaplanan C kod değerleri, önceden atanmış değerlerle karşılaştırılır. Sonra test verisinden hesaplanan değerlerle, C kod değerleri arasındaki Hamming uzaklığı formunda en yakın olan sınıf etiketine atanır (Goh ve ark., 2001).

Literatürde, sınıflandırılamayan bölgeler probleminin çözülmesi için çalışmalar sürmektedir. Bu problemin çözümüyle ilgili olarak Abe ve Inoue (2002), bulanık mantık metodunu kullanarak yeni bir yaklaşım önermişlerdir. Çıkarılan bulanık kurallar yardımıyla sınıflandırılamayan sorunlu bölgeyi de, bilinen sınıf bölgelerine ait olacak şekilde bir paylaşım yapılmıştır.

Ayrıca, (Çomak, 2004), (Çomak ve Arslan, 2006) çalışmalarında geliştirilmiş olan yeni bir yaklaşım ile bu problemin çözümü gerçekleştirilmiştir. Bu yaklaşımda ise, sorunlu bölgeye dâhil olan her bir test örneği bir sorgu noktası olarak alındıktan sonra bu sorgu noktasının en yakın komşuları incelenmiştir. En son olarak da, test örneği bu komşular arasında en çok görülen sınıf hangisi ise o sınıfın etiketine atanmıştır. Algoritmanın tasarlanması ve gerçekleştirilmesi Abe ve Inoue (2002)'nin çalışmasında kullanılan algoritmaya göre çok daha basittir. Aynı zamanda, bazı veri kümelerinde eşdeğer bazılarında ise daha iyi sınıflandırma başarımı sağlanmıştır. Bu algoritmaların detayları (Çomak, 2004), (Çomak ve Arslan, 2006)'da bulunmaktadır.

Çok sınıflı sınıflandırma ile ilgili son olarak, bu tez çalışmasında da yeni bir yaklaşım önerilmiştir. Ayrık sınıflandırma değerleri yerine sürekli değerler kullanan bu yaklaşımın ayrıntıları da 4. bölümde sunulmuştur.

2.5.5 En Küçük Kareler Destek Vektör Makineleri (EKKDVM)

Tez çalışmasında, EKK-DVM (En Küçük Kareler Destek Vektör Makinesi) makine öğrenmesi metodundan bahsedilmesinin iki nedeni vardır: bunlardan birincisi, yardımcı bazı metotlar kullanarak gerçekleştirilen çalışmalarda EKK-DVM sınıflandırıcı çeşidinin kullanılmış olmasıdır. Diğer ve daha önemli olan nedeni ise, bu tez çalışmasında tasarlanan eğitim algoritması ile EKKDVM sınıflandırıcısı arasında amaç yönünden benzerlikler olması ve mukayese edilmeye uygun olmasıdır.

EKKDVM, Suykens ve Vandewalle (1999) tarafından önerildi. Bu DVM çeşidi de standart DVM sınıflandırıcılarında olduğu gibi iki sınıf etiketli sınıflandırma işlemi düşünülerek oluşturulmuştur. Standart DVM sınıflandırıcılardaki optimizasyon işleminin çözümü amacıyla kullanılan ikinci dereceden programlama metotları bu teknikte kullanılmaz. Onun yerine doğrusal eşitlik kümesi kullanılır. Yani;

$$½ \|w\|^2 + \frac{C}{2}\sum_{i=1}^{n}\xi_i^2 \quad (2.52)$$

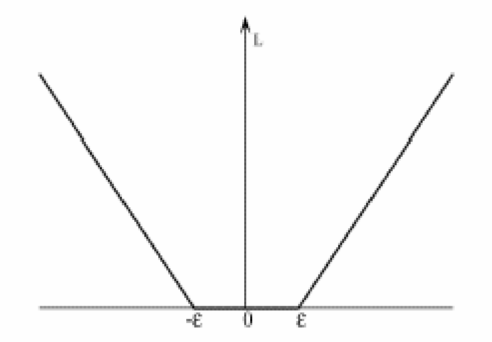
eşitliği,

$$y_i\,[(w.x_i) + w_0] = 1 - \xi_i, \qquad i=1,\ldots,n. \quad (2.53)$$

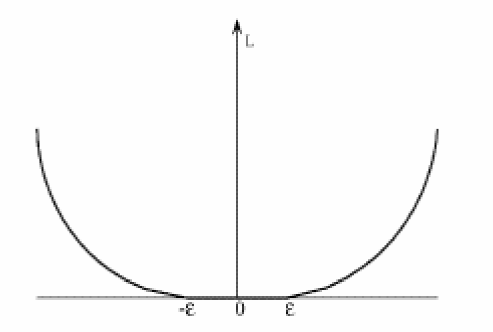
koşullarına bağlı olarak minimumlaştırılır. Standart DVM sınıflandırıcılarda ise (2.53) denklemi eşitlik durumunda değil eşitsizlik durumunda bulunur. Optimizasyon probleminin çözümünden sonra standart DVM sınıflandırıcılarda olduğu gibi optimizasyon eşitlikleri 2.54 eşitliğindeki gibi ikili problem yapısına dönüştürülür.

$$Q\,(w,b,\alpha\,,\xi\,) = ½\,\|w\|^2 + \frac{C}{2}\sum_{i=1}^{n}\xi_i^2 - \sum_{i=1}^{n}\alpha_i\{y_i[(w.x_i) + w_0] - 1 + \xi_i\} \quad (2.54)$$

Bu eşitliğin de standart DVM sınıflandırıcılardan bir farkı vardır. O da, α çarpanları standart DVM sınıflandırıcılarında pozitif olması gerekirken EKKDVM sınıflandırıcılarında pozitif veya negatif olabilmesidir (Tsujinishi ve Abe, 2003), (Suykens ve Vandewalle, 1999). Bu formülsel farklılıklar ve daha az eğitme zamanı gerektirmesiyle EKKDVM sınıflandırıcıları standart DVM sınıflandırıcılardan ayrılır. Sınıflandırma performansı ise, bazı uygulamalarda standart DVM öğrenme metoduna yakın da olsa düşük çıkabilmektedir. Bunun sebebi de eğitme aşamasında daha kuvvetli ama zaman alan ikinci dereceden programlama yerine daha basit ama kısa sürede uygulanabilen doğrusal programlamayı kullanmasındandır.



Şekil 2.11 Lineer ε Kayıp Fonksiyonu



Şekil 2.12 İkinci Dereceden ε Kayıp Fonksiyonu

2.5.6 Destek Vektör Makineleri Regresyon Uygulamaları

Regresyon, destek vektör makinelerinin kullanıldığı bir diğer öğrenme işlemi çeşididir. Tez çalışmasının konusu asıl sınıflandırma algoritmaları ile ilgili olduğu için regresyon konusu bu kısımda kısaca özetlenmiştir.

Regresyon, sonlu bir gürültülü örnek kümesi üzerinde gerçek değerli bir fonksiyonun tahmin edilmesidir. Regresyonun çıktısı gerçek değerli sayılar kümesine ait bir değişkendir. Gerçek değerli çıkışlar için öğrenme işlemi teorik olarak istatistiksel öğrenme teorisini kullanır. Önceki bölümlerdeki ayırıcı karar fonksiyonunun yerine, aynı yapıda fakat veriyi modelleyen $f(x) = w\,x + b$ fonksiyonu ve gerçek değerli y_i hedefleri arasındaki sapma değeri olan ε a her iki yönde de izin vermek için y_i - $w\,x_i - b \leq \varepsilon$ ve $w\,x_i + b - y_i \leq \varepsilon$ şekline getirilir (Hong ve Hwang, 2003). Bunu $f(x)$ hipotez fonksiyonu etrafında +/- $(\theta$ - $y)$ boyutlu bir tüp olarak gösterebiliriz ve bu tüpün dışında bulunan bir nokta eğitme hatasıdır. İki eğitme hatası tipi için ξ_i ve $\xi_i^\wedge$ gevşek değişkenleri tanımlansın. Bu gevşek değişkenler tüpün iç tarafında kayıp fonksiyonuna göre sıfırdır ve tüpün dışındaki noktalar için de doğrusal olarak artar. Şekil 2.11 de gösterilen bu genel yaklaşım ε-DVM regresyonu adını alır ve DVM regresyon'da yaygın şekilde kullanılır.

Regresyon'da sistemin çıktısı rasgele bir değişkendir (Cherkassky, 1998). Bu rasgele değişken, gerçek değer kümesinden değerler alır ve kararlı bir fonksiyon toplamı olarak yorumlanabilir.

$$y = g(x) + \varepsilon \tag{2.55}$$

Buradaki kararlı fonksiyon g(x), çıkış koşullu olasılığının ortalamasıdır.

$$g(x) = \int yp(y/x)\, dy \tag{2.56}$$

Regresyon'da genel bir kayıp fonksiyonu aşağıdaki gibi ifade edilir.

$$L(y,f(x,w)) = (y - f(x,w))^2 \tag{2.57}$$

Bu durumda öğrenme, risk fonksiyonunu minimumlaştıran $f(x,w_0)$ formülünün bulunmasıdır. Regresyon için risk, gürültü varyansıyla doğruluk yaklaşımının birleşimidir. En az riskli tahmini yani, en doğru bilinmeyen $g(x)$ fonksiyonunu bulmaya eşdeğerdir.

3. KULLANILAN YARDIMCI METOTLAR

3.1 Gevşek Öğrenme Metotları

Gevşek öğrenme (lazy learning) kavramı bir algoritma ailesi altında toplanır. Bu algoritmalar aşağıdaki koşullara göre çalışmaktadırlar:

- Verilen etiketlenmiş tüm örnekler kümesinin saklanması,
- Daha önce görülmemiş örneklerinden sınıflandırılma isteği alınana kadar sınıflandırma hakkında yapılacak tüm işlemlerin ertelenmesi.

Eğer bir istek meydana geldiği zaman tüm hesaplamaların yapılması zorunlu ise, elbette görülmemiş örneğin sınıflandırılması için gereken zaman istekli öğrenme (eager learning) tekniklerinden daha fazla olmaktadır. Bu tip algoritmaların avantajı, onların tek bir hipotez çıktısı vermek zorunda olmamasıdır. Fakat onlar yerel olarak iyi bir şekilde oluşturulan hedef fonksiyonu için farklı yaklaşımlar kullanabilirler. Bu yol, sınıflandırma işlemi boyunca daha fazla dikkat isteyen örneklere benzemektedir.

Bu tip metotların işletilebilmesi için bir benzerlik ölçüsü gereklidir. Bu benzerlik ölçüsü, gözlenmeyen bir örneğin sınıflandırılmasında diğer örneklerin öneminin değerlendirilmesini sağlar. Eğer örnekler gerçek değerli bir vektör ile temsil ediliyorsa o zaman benzerlik, olağan vektör uzaklığı (tüm niteliklerin aynı etkiye sahip olup olmadığı sorusunu çıkaran) ile ölçülebilir. İlişkisiz niteliklerin yararlı olmayan etkisinin azaltılması ve her bir niteliğe uygun etki derecesinin atanması için bu metotlar kullanılabilir.

Makine öğrenmesi algoritmalarının bir dalı olan gevşek öğrenme; girişlerin değerlendirilmesi işlemini erteler, depolanmış veriyi kullanarak bilgi istemlerine cevap verir ve sonuçta yapılandırılmış cevapları atar. Yapay zekâ; örnek tabanlı, hafıza tabanlı ve yerel öğrenme gibi durum tabanlı uygunluk ve KEYK sınıflandırıcılarında kesişen eş anlamlı konuları içermektedir. Bu konu, yapay zekâ araştırmacıları için bir makine öğrenmesi dalıdır. Çünkü bu tekniğin geniş bir katkısı vardır. Bu teknik ayrıca veri madenciliği, durum tabanlı uygunluk, istatistik ve örüntü tanıma ile de ilgilidir.

Klasik makine öğrenmesi literatüründe iki farklı öğrenme tipi vardır: bireylere göre yapının oluşturulup sonradan karşılaşılan örneklerin saklanmadığı istekli öğrenme ve örneklerin saklanarak sınıflandırma veya tahmin istendiği zaman işlemin yapıldığı gevşek öğrenme. Günümüz çalışmalarının amacı iki öğrenme tipinin insan doğasında tanımlanabilir olup olmamasının kurulmasıdır. İstekli öğrenme sinir ağlarındaki artışsal öğrenmeye benzemektedir, hâlbuki gevşek öğrenme kategorisel örnek modellerini hatırlatır. Bu konu daha fazla hafıza yoğunlu gevşek öğrenme ile aşılmaya çalışılmıştır, oysa hırslı öğrenme metotları hafıza yoğunlu değillerdir.

Tez çalışmasında önerilen yeni eğitim algoritması, en yaygın örnekleri KEYK en yakın komşuluk veya örnek tabanlı öğrenme olan gevşek öğrenme yapısından da yararlanmaktadır. Gevşek öğrenme sınıflandırma isteği ile karşılaşıncaya kadar hesaplamaları bekletme özelliğiyle ayırt edilir. Örneğin karar ağaçları veya kuralları gibi eğitme zamanında oluşturulan açık teorileri yoktur. Bir test örneğiyle yüz yüze gelindiği zaman, gevşek öğrenme algoritmaları tahmin yapmak için eğitme örneklerine başvurur.

Gevşek öğrenme; en iyi yerel model yapısını belirlemek için sorgu tabanlı bir yaklaşım kullanan hafıza tabanlı bir tekniktir (Bontempi ve ark., 2001). Gözlenen veriden uygun makine öğrenmesi modelinin kurulması, global veya yerel tekniklerle gerçekleştirilebilir. Gevşek öğrenme teknikleri yerel teknikler grubuna dâhildir. Sorgu tabanlı bir mantıkla yerel bir alanda uygulanan gevşek öğrenme metotları, bu yerel alanda elde ettikleri sonuçları birleştirerek global sonuçlara da ulaşabilirler. En önemli iki yerel yaklaşım tekniği, gevşek öğrenme algoritmaları ve bulanık çıkarım sistemleridir. Gevşek öğrenme teknikleri bulanık çıkarım sistemlerine alternatif bir yaklaşımdır. Modelleme ve kontrol alanlarında da verdiği sonuçlarla bulanık sistemlere karşı ciddi bir alternatif haline gelmişlerdir.

Gevşek öğrenme, sınırlı miktarda giriş çıkış verisi mevcut olduğu zaman lineer olmayan dinamik sistemlerin kontrolü ve modellenmesi için yerel bir metot olarak başarıyla kullanılabilir (Bontempi ve ark., 1998).

3.1.1 KEYK Sınıflandırma Metodu

KEYK, makine öğrenmesinde sıklıkla kullanılan ve geniş bir uygulama alanında çalıştırılmış bir metottur. Bu metodun C4.5 metodundan daha iyi başarım gösterebildiği Han ve ark. (2001) tarafından belirtilmiştir. KEYK sınıflandırıcısı, parametrik olmayan bir örüntü sınıflandırma tekniğidir (Kim ve Shin, 2000). En yakın komşu sınıflandırıcıları benzerlik tabanlı öğrenme metotlarını kullanırlar. KEYK metodu iki ana konu üzerine kurulmuştur. Bunlar; benzerlik ölçüsü ve çıkış belirleme yöntemidir. Eğitme örnekleri *n* boyutlu sayısal niteliklerle tanımlanırlar. Her bir örnek *n* boyutlu uzaydaki bir noktayı temsil eder. Bu şekilde, tüm eğitme örnekleri *n* boyutlu örüntü uzayında temsil edilir. Bilinmeyen bir örnek verildiği zaman, KEYK sınıflandırıcısı bilinmeyen örneğe en yakın *k* eğitme örneğini bu örüntü uzayı içinde belirler. Bu *k* eğitme örneği bilinmeyen örnek verisinin en yakın *k* sayıda komşusudur. Çıkış belirleme yöntemi olarak ise genellikle KEYK ta, en çok oy alan kazanır yaklaşımı kullanılmaktadır. KEYK yaklaşımının uygulama sonuçları diğer makine öğrenmesi sınıflandırıcı metotları ile karşılaştırılabilir bir ölçüdedir. Benzerlik ölçüsü, öklit uzaklık formülüyle belirlenir. $X = (x_1,x_2,...,x_n)$ ve $Q = (q_1,q_2,...,q_n)$ gibi iki nokta arasındaki öklit uzaklığı şöyle ifade edilir. Benzerlik ölçüsü ile öklit uzaklığı değerleri birbirlerine ters orantılı olarak değişirler. Yani, daha küçük öklit uzaklığı daha çok benzerlik anlamına gelmektedir.

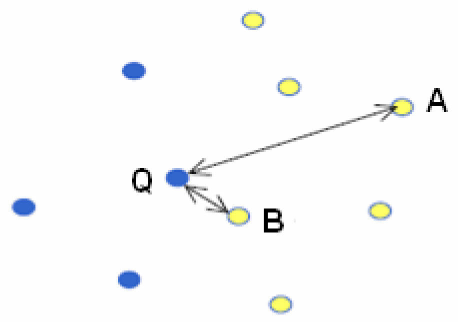
$$(X, Q) = \sqrt{\sum_{i=1}^{m} (x_i - q_i)^2} \tag{3.1}$$

Bahsedilen en yakın komşuluk işlemi, sınıfları, gerçek değerli verileri veya kümeleri temsil edecek şekilde genişletilebilir. Tez çalışmasında, KEYK metodunun bu özelliğinden yardımcı metotların desteğiyle de birlikte yararlanılarak kümeleme ve veriler üzerinde değerlendirme işlemleri gerçekleştirilmiştir. Kümeleme temsillerinde, küme içindeki verilerin birbirlerine benzerlik oranı diğer kümelerdeki verilere göre daha fazla olmalıdır. Bu safhada bazı yakınlık ölçülerinden yararlanılabilir. Tez çalışmalarında 3.1 eşitliğinde gösterilen öklit uzaklığı kullanılmıştır.

Bilinmeyen veri örneği en çok oy alan kazanır yaklaşımına göre, onun k sayıda en yakın komşusu arasında en yaygın olan sınıfa atanır. Bu *k* değeri 1 alındığı zaman, bilinmeyen örnek örüntü uzayındaki kendisine en yakın olan eğitme örneğinin sınıfına atanmış olur.

Örneğin şekil 3.1'de, bir sorgu noktasının iki ayrı nokta ile arasındaki uzaklıklar sunulmaktadır. Uzaklıklar uygulamalarda küçükten büyüğe doğru sıralanıp ilk *k* tanesi alınır. Bunların arasında fazla olan sınıf etiketine atama yapılır.

En yakın komşu sınıflandırıcıları; tüm eğitme örneklerinin tutulduğu, fakat yeni alınan test örneğinin sınıflandırılmasına ihtiyaç duyulana kadar sınıflandırıcı yapısının oluşturulmadığı örnek tabanlı (veya gevşek öğrenme) sınıflandırıcılarındandır. Bu düşünce, karar ağaçları ve geriye yayılma gibi sınıflandırma işlemleri için yeni bir örnek alınmadan önce bir genelleme modelinin oluşturulduğu istekli öğrenme metotlarından farklıdır. Esnek öğreniciler, karşılaştırılacak etiketsiz veri miktarı çok fazla ise sınıflandırma aşamasında yüklü bir hesaplama maliyetine sebep olurlar. Bu yüzden, onlar etkili indeksleme teknikleri isterler. Beklendiği gibi gevşek öğrenme metotları eğitme açısından istekli metotlardan daha hızlıdırlar, fakat tüm hesaplamalar o ana kadar ertelendiği için sınıflandırma açısından daha yavaştırlar. En yakın komşuluğun dezavantaj oluşturabilecek bir yönü de, sınıflandırmada her bir niteliğe eşit ağırlık değeri vermesidir. Veriler içinde çok sayıda ilişkisiz nitelik olduğu zaman, bu özelliği kötü durumlara yol açabilir.



Şekil 3.1 En Yakın Komşuluk Kavramı ve Uzaklık Ölçüsü

En yakın komşu sınıflandırıcıları tahmin için de kullanılabilir. Bu durumda, sınıflandırıcı bilinmeyen örneğin k en yakın komşusuyla gerçek değerli etiketlerin ortalama değerini döndürür. Bir başka deyişle, aslında sadece çıkış değerleri ve buna bağlı olarak ta çıkışın KEYK tarafından yansıtılması farklılık gösterir.

3.2 Entropi Kavramı

Entropi, Shannon (1948) tarafından bilgi teorisi ile bağlantısı üzerine önerilmiştir. Herhangi bir olasılık yoğunluğunun belirsizlik oranını ölçmek için önerilen bir yöntemdir. Örneğin, bir metin dokümanını bir kaynak noktasından hedef noktasına göndermek istediğimizi düşünelim. Dokümanda bulunan harfler $\Omega = \{\xi_1, \xi_2 ...\}$ gibi belirli indislerle ifade edilmiş olsun. Harflerin elektronik transferine olanak sağlamak için, ikili sistemde kodlama yapılması uygun olacaktır. Bu yüzden alfabede bulunan temsiller için ihtiyaç duyulan dijit sayısı da bilinmelidir. Bu indislerle ifade edilen temsiller rasgele bir değişken gibi düşünülüp, bu değişkenin alfabedeki temsiller için belirli olasılık değerleri alması sağlanabilir. Alfabenin 4 harf içerdiğini (A,B,C,D) varsayılırsa, rasgele değişkenin etki alanı belirlenmiş olur. Fakat her bir harf için yansıtacağı olasılık değeri de belirlenmelidir.

Eğer bütün harfler aynı olasılığa (tek düzenli dağılım) sahipseler, olasılık değerleri $p(\xi_i)$ = ¼ = 0.25, i=1,2,3,4 olur. Mümkün 4 çıkış durumunu temsil edebilmek için 2 bit kullanılmaktadır. Bu yüzden 2, bu olasılık dağılımının belirsizlik ölçüsünü temsil edebilir. Tek düzenli dağılım söz konusu olduğu zaman, belirsizlik ölçüsü mümkün çıkış sayısına bağlıdır. Daha çok çıkışın olması daha yüksek belirsizlik anlamına gelmektedir. Dağılım tek düzenli olmadığında ise, belirsizlik ölçüsü her bir temsil için hesaplanır ve rasgele değişkenin belirsizlik ölçüsünü belirsizliğin beklenen değeri cinsinden bulunur. Entropi, eşitlik (3.2) ile ifade edilir.

$$H(\xi) = -\sum_{\xi \in \Omega} p(\xi) \log_2 p(\xi) \tag{3.2}$$

Ayrık değerli olasılık dağılımının mevcut 4 temsil için sırasıyla 1/2, 1/4, 1/8 ve 1/8 olduğu kabul edilirse, (3.2) eşitliğine göre hesaplanan entropi değeri 1,75 olur. Tek düzenli dağılım için entropi değeri ise 2 bulunmuştu. Buradan ve daha birçok örnek uygulamadan görülebilir ki, genellikle tek düzenli dağılımda bulunan entropi değeri, tek düzenli olmayan dağılımlarda bulunan entropi değerlerine göre daha yüksek çıkmaktadır. Zaten teorik olarak bulunabilecek maksimum entropi değeri $\log_2$ tabanında temsil sayısıdır. Yani, bu örnek için $\log_2 4 = 2$ olur.

Bu konu ile ilgili olarak üzerinde durulması gereken iki özel durum vardır. Eğer olasılık değerleri 1 ise, o zaman entropi 0 olmaktadır. Diğer taraftan, olasılık değerleri 0 ise, o zaman entropi değeri yine 0 olmaktadır. Çünkü her iki durumda da temsil değeri tamamen belirlidir, her zaman vardır veya yoktur. Bu alanda daha sonra önemli çalışmalar yapan Renyi (1976) tarafından gerçekleştirilen araştırmalar Shannon (1948) tarafından önerilen çalışmaların aslında daha genel bir tanımlama kümesinin özel bir durumu olduğunu göstermiştir. Bu genel tanımlar Renyi Entropi veya Renyi ortak bilgisi olarak adlandırılmıştır. Daha eski ve önceden kabul görmüş bir metot olduğu için Shannon'ın yaklaşımı 1990'lı yıllara kadar farklı uygulama alanlarında daha çok rağbet görürken Renyi Entropi yaklaşımı pek kullanılmamıştır. Bu yıllardan itibaren örüntü tanıma ve şifreleme gibi alanlarda kullanılmaya başlanmıştır (Sahoo ve ark., 1997), (Cachin, 1997).

Renyi entropi formülü ikinci dereceden alındığında, bilgi teorisi ile ilişkisi eşitlik (3.3)'de ifade edildiği gibi daha net görülmektedir.

$$H(x) = -log \int_{-\infty}^{\infty} f_x^2(x)dx \qquad (3.3)$$

Erdoğmuş ve Principe (2002), Renyi tanımlamalarını ve Jensen eşitsizliklerini kullanarak verilen giriş sınıfına göre sınıflandırıcılarda oluşabilecek çıkış hatasını yansıtan üst ve alt sınır değerlerini bulmuşlardır. Bu tez çalışmasında, gerçekleştirilen bazı uygulamalarda da bahsedilen özelliği yansıtan uygulama sonuçları elde edilmiştir. Bu sonuçlar, tezdeki amaçlardan bir tanesi olan DVM parametre düzenleme prosedürü için oldukça faydalı olmuştur. Sürekli değerler alan rasgele bir değişkenin entropi diferansiyelini tahmin edebilmek için çeşitli metotlar geliştirilmiştir. Bu metotlar aşağıda incelenmektedir.

3.2.1 Plug-in Tahmin Metotları

Plug-in metotları, basitçe sürekli bir yoğunluk tahmini verisinin entropi ifadesindeki doğal yoğunluk tahmini fonksiyonuna eklenmesi ile elde edilir. Bu metotların oluşturduğu 4 çeşit yaklaşımdan birisi tercih edilebilir. Bunlardan ilki olan integral tahminleri, entropi tanımında var olan sonsuz integralleri doğrudan veya yaklaşık olarak değerlendirir. İkinci dereceden Renyi entropi tahmincisi, bu gruba dâhil bir metottur ve doğrudan değerlendirme yapar. İkinci yaklaşım olan yeniden yerine koyma tahminleri, daha çok entropi tanımındaki beklenti operatörünün örneklem ortalamasına yakınsamasını sağlar. Bu gruba çekirdek tabanlı olasılık yoğunluk fonksiyonu tahminleri girer. Özellikle çok boyutlu verilerle gerçekleştirilen uygulamalarda görülmüştür ki, verinin boyutundaki artışla beraber ihtiyaç duyulan örneklem sayısı da hızla artmaktadır. Shannon entropi, özellikle üzerine tahminlerde bulunan elektrik mühendisleri tarafından kullanılır. İlgilenilen uygulamaya bağlı olarak, bu tahminler yeniden şekillendirilebilir. Bu tahminlerin entropi değerlerinde de küçük değişimler olabilmektedir. Parametrik olmayan Renyi entropi de bu gruba dahildir. Bölünmüş veri tahminleri olarak adlandırılan üçüncü yaklaşım, bir farklılık dışında yeniden yerine koyma tahminlerine benzer. Bu fark ta, örneklem kümesinin ikiye bölünüp bir kısmının yoğunluk tahmini için ikinci kısmının da örneklem ortalamasının belirlenmesi için kullanılmasıdır. Son olarak, çapraz geçerlilik tahmini adı verilen dördüncü yaklaşım, yeniden yerine koyma tahmininde birer parça ayırma (leave-one-out) prensibini kullanır. Entropi tahmini birer parça ayırma tahmininin ortalamasıyla elde edilir.

3.2.2 Uyarlama ve Öğrenme için Entropi Tahmini

Literatür taramalarından yansıyan bulgulara göre, Shannon entropi üzerinde uzun zamandır çok sayıda çalışma yapılmış ve geliştirilmesi için büyük çabalar harcanmıştır. Shannon entropi hakkındaki bu fikirlerin çoğu Renyi entropi çalışmalarına da uyarlanabilir. Bu konuda özellikle, hesap yükü az olan, uygulanabilirliği açısından basit, sürekli değerli ve örnekler üzerinde ayrım yapabilen entropi tahminleriyle ilgilenilmiştir. Tez çalışmasındaki başlıca amaç, sadece geliştirilmiş olan bu entropi değerlerinin bulunması değildir. Başlıca amaç, bulunan değerler yardımıyla kümeleme işleminin gerçekleştirilmesi ve DVM eğitim algoritmasında kullanılması, ayrıca DVM sınıflandırıcısının öğrenme parametresi değerlerinin tahmin edilmesidir. Parametre tahmini işlemi, bulunan entropi değerleri ile kümeleme yapılması temeli üzerine kurulmuştur.

3.2.3 Renyi Entropi ile Kümeleme

Kümeleme, örüntü tanıma uygulamalarında kullanılan temel problem çeşitlerinden birisidir. Eğiticisiz öğrenme durumunda çalışarak örnek verilerinin doğal organizasyonunu, onları küme gruplarına ayırarak sağlar. Kümelemenin, veri madenciliği (Judd ve ark., 1998), görüntü ayrıştırma çalışmaları (Frigui ve Krishuapuram, 1999), sinyal sıkıştırma (Abbas ve Fahmy, 1994) ve makine öğrenmesi (Carpineto ve Romano, 1996) gibi alanlarda uygulamaları mevcuttur.

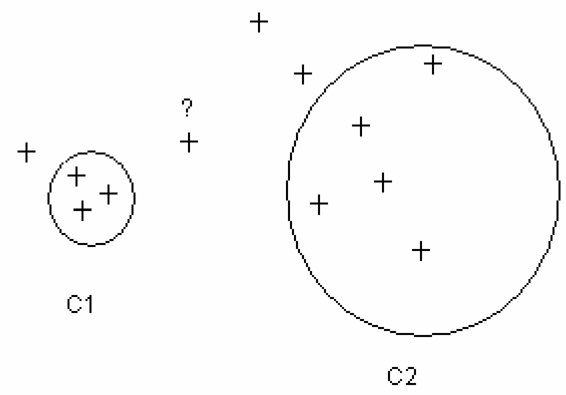
Genel anlamda kümeleme metotları ikiye ayrılabilir (Jain ve Dubes, 1988). Bunlar parçalı algoritmalar ve hiyerarşik algoritmalardır. Parçalı algoritmalar, veri ile küme merkezi arasındaki hataların kareleri toplamı gibi herhangi bir maliyet fonksiyonunu minimuma ulaştıracak şekilde bir ayrıştırma planını çalıştırır. Örneğin K-ortalamalar (K-means) metodu bu gruba dahildir. Bu metot küresel ve eliptik yapılı veri kümelerinde daha başarılıdır. Çünkü sadece ikinci dereceden bir istatistiksel özelliği kullanmaktadır. Ayrıca bu metot, verilerin gruplandırılacağı küme sayısını da öncelik bilgisi olarak kullanmaktadır ve sadece bu önceden verilen küme sayısı ile tolerans değerine göre kümeleme yapar. Farklı küme sayıları için bir fikir vermezler.

Diğer taraftan hiyerarşik algoritmalar, hiyerarşi içinde farklı sayıda küme ve farklı küme elemanları içeren bir yapı sunar. Bu durum da, kümeleme işlemine farklı alternatifler sunar ve veri grubunun doğasına daha uygun bir yapının seçilmesi olasılığını artırır.

Son yıllarda, Eltoft ve deFigueiredo (1998) tarafından YSA (Yapay Sinir Ağları) tabanlı ve Ben-Hur ve ark. (2001) tarafından da DVM (Destek Vektör Makineleri) tabanlı daha karmaşık yapılı kümeleme metotları geliştirilmiştir. Bu metotlar herhangi bir şekle sahip tüm yapıları kümeleyebilmekte ve öncelik bilgisine de gerek duymamaktadır. Fakat bu metotlar da karmaşık yapılı metotlardır ve uygulamalar üzerinde çalıştırılırlarken bazı parametrelerinin optimum değerlerinin bulunması gerekir. Bu işlem de ayrıca zaman alır.

Renyi entropi değerlerine göre kurulmuş olan kümeleme metodu ise, basit ve sezgisel bir kümeleme algoritmasıdır. Renyi entropi kendi doğası gereği parametrik olmayan yoldan ve geleneksel entropi yöntemlerindeki gibi ölçüm değerlerinin ayrıca değerlendirilmesine gerek duymayan bir kümele yöntemidir. Renyi entropi temelli kümeleme yöntemi ilk kez Jenssen ve ark. (2003) tarafından önerilmiştir. Bu yaklaşım sadece veriler arasındaki ikinci dereceden istatistiksel ilişkilere dayanmamaktadır. Verinin tabiatında dahil olduğu dağılıma daha iyi yakınsayan bir kümeleme yöntemidir. Ayrıca öncelik bilgisine de ihtiyaç duymamaktadır.

Renyi entropi kümeleme yönteminde; kümelerden bir tanesi çıkarılıp onun elemanları diğer kümelere aktarılırken, mümkün tüm kümeler arasında küme içi entropi değerinin en az artırdığı kümeye eklenmesi tercih edilir. Kümeler arası değerlendirmelerde ise, ilk kez Gokcay ve Principe (2002) tarafından önerilen kümeler arası entropi değeri adı verilen yaklaşım kullanılır. Bu metot da; herhangi bir şekle sahip olan tüm kümeleri bulabilme yeteneğine sahiptir, küme sayısı gibi bir öncelik bilgisine gerek duymaz. Ayrıca YSA ve DVM temeline göre kurulmuş olan kümeleme metotlarına göre çok daha basittir, uygulanması kolaydır ve hesap yükü azdır.



Şekil 3.2 Fark Temelli Entropi Yaklaşımı

Şekil 3.2'de gösterilen durum düşünülürse, herhangi bir veri kümesi şekil 3.2'de gösterildiği gibi dağılsın ve bir kısmı c_1 kümesine bir kısmı da c_2 kümesine ait olsun. Bu kümeler şekil 3.2'de çember içine alınarak gösterilmiştir. Kümeleme işlemindeki karar aşamasında "?" ile gösterilen yeni bir örneğin hangi kümeye (c_1 mi, yoksa c_2 mi) dahil edilmesinin daha uygun olacağı sorusuna cevap aranır.

Bu aşamada Jenssen ve ark. (2003)'ı bir çözüm geliştirmişlerdir. Eğer kümesi aranan yeni x örneği yanlışlıkla c_1 kümesine atanırsa, c_1 kümesinin entropi değerinde meydana gelecek artış hesaplanır. Aynı artış c_2 kümesi için de hesaplanır. Eğer c_1 kümesindeki artış değeri c_2 de bulunan değerden yüksek ise örnek c_2 kümesine, aksi durumda ise c_1 kümesine dâhil edilir şeklinde karar verilir. Bu çözüm yolu k sayıda küme için eşitlik (3.4) ile ifade edilir.

$$H(C_i + x) - H(C_i) < H(C_j + x) - H(C_j), \quad i \neq j, \qquad i,j=1, 2,\dots,k, \qquad (3.4)$$

Bu çözüm yolu, fark temelli entropi kümeleme olarak adlandırılır. Bu konuda, üç sorun ile yüz yüze karşılaşılabilir. Bunlardan birincisi, doğrudan veriyi kullanarak entropi değerinin nasıl hesap edileceğidir. İkincisi, eldeki veri kümesinden başlangıç durumu kümelerinin nasıl oluşturulacağıdır. Son olarak ta, mevcut kümelere eklenecek olan sonraki veri örneğinin nasıl seçileceği konusudur. Bu sorunların çözümleri de yine Jenssen ve ark. (2003) tarafından sunulmuştur. Bu çözümler aşağıda başlıklarda incelenmektedir.

- *Küme İçi Entropi:* Entropi, bazı varsayımlar olmadan değerlendirilmesi zor bir ölçüdür (Gokcay ve Principe, 2002). Fakat, Renyi entropi temelli yapının doğrudan veri kümesinden parametrik olmayan tahminler yapabildiği saptanmıştır (Principe ve ark., 2000). Bir f_x olasılık yoğunluk fonksiyonuna uyan x değişkeni için Renyi entropi eşitlik (3.5) ile ifade edilir.

$$H_R(x) = \frac{1}{1-\alpha} log \int f_x^{\alpha} dx \,, \qquad \alpha > 0, \qquad \alpha \neq 1, \tag{3.5}$$

Özel olarak α = 2 alınırsa, eşitlik 3.3'te gösterilen ikinci dereceden Renyi entropi elde edilir.

Bu ifade, çok boyutlu Gauss penceresi kullanılarak doğrudan veriden elde edilebilir. Örneğin N_k adet veri içeren C_k kümesi için olasılık yoğunluk fonksiyonu veriye göre eşitlik (3.6) ile ifade edilir.

$$\hat{f}_x = \frac{1}{N_k} \sum_{i=1}^{N_k} e^{-\frac{(x-x_i)^2}{2\sigma^2}} \tag{3.6}$$

Burada N_k, C_k kümesinde bulunan verilerin sayısıdır. Simetrik Gauss fonksiyonu ise, σ^2I kovaryans matrisi ile oluşturulmuştur. Denklem (3.6)'yı denklem (3.5)'te yerine yazarsak eşitlik (3.7) elde edilir.

$$H(C_k) = -logV(C_k) \tag{3.7}$$

$$V(C_k) = \frac{1}{N_k^2} \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} e^{-\frac{(x_i-x_j)^2}{2\sigma^2}} \tag{3.8}$$

Eşitliğe göre, entropi aynı kümeye ait veriler üzerinde hesaplandığından küme içi entropi adını almıştır.

Tezde, bu kümeleme işlemi uygulanırken N sayıda veri kümesi için $N*N$ ölçülerinde simetrik bir uzaklık matrisi oluşturulur. Bu matris ve önceden belirlenen σ değeri yardımıyla hesaplama yapılır. Buradaki bir diğer konu σ, çekirdek genişliğinin belirlenmesidir. Bu parametre değerinin seçimi, genel parametrik olmayan olasılık yoğunluk fonksiyonu seçimine bağlıdır. Ayrıca, geliştirilmiş olan yönteme benzer bir şekilde, gerçekleştirilen uygulamalarda kullanılan veri kümeleri de [-1,1] aralığında normalleştirilmiştir.

- *Kümelemeye Başlangıç Aşaması:* Veri kümesinde ilk aşamada, K_{init} sayıda başlangıç kümesi oluşturulur. Bunun için başlangıçta veri kümesinden K_{init} sayıda örnek her birisi ayrı bir kümeyi temsil edecek şekilde rasgele seçilir. Daha sonra bu örneklere en yakın diğer örnekler ilgili kümeye dahil edilerek küme genişletilir. Bu genişletme işlemi her kümede N_{init} sayıda örnek barındırılıncaya kadar devam eder. N_{init} değeri de önceden seçilmiş bir değerdir. Fakat bu kümeye yeni örnek seçme işlemi en yakın komşuluk sınıflandırıcılarında uygulanan planlamalardan farklıdır. Çünkü verilerin doğasına daha uygundur.

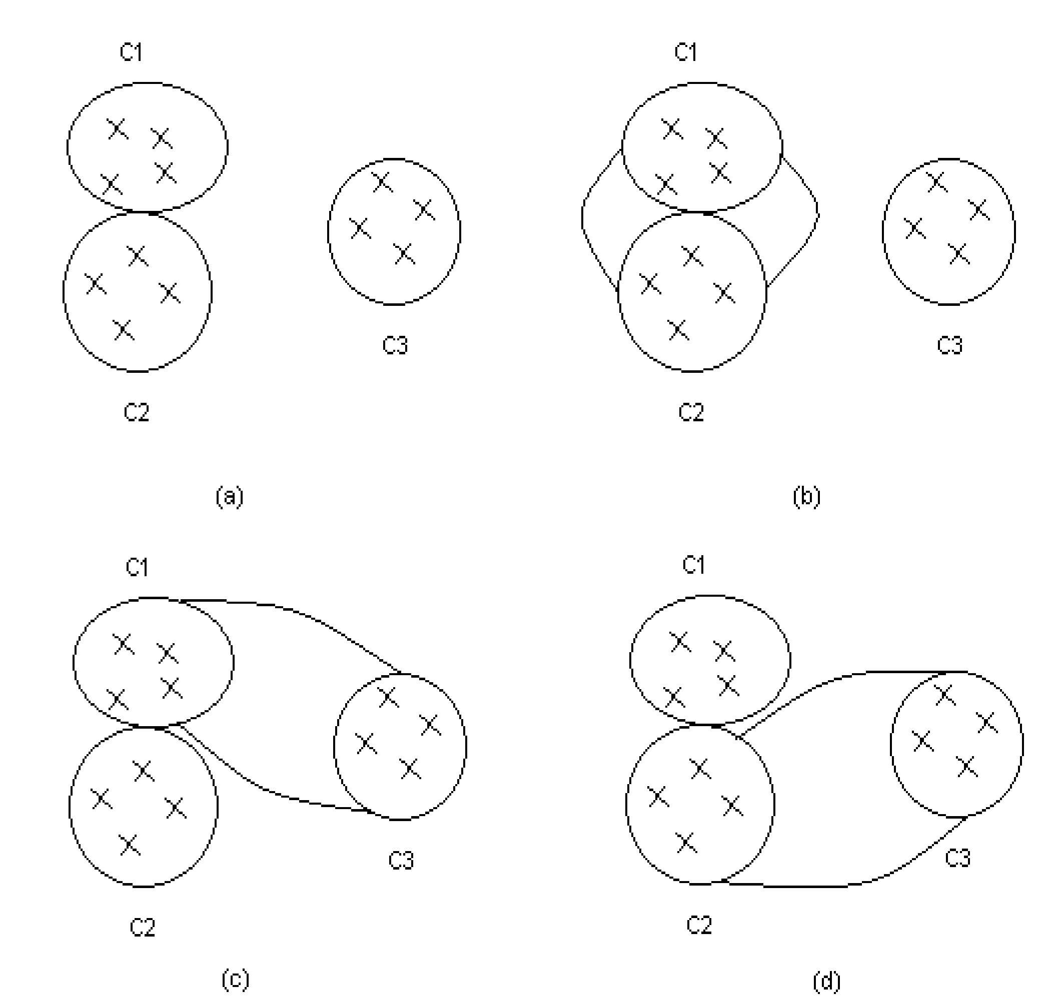
- *Kümeleme için Sonraki Verinin Seçilmesi:* Kümeye dâhil edilecek sonraki veri iki yolla seçilebilir. Birincisi rasgele seçimdir. Bu seçim şeklinde, rasgele alınan veri noktaları başlangıç verisinden çok uzakta çıkabilir. Bu durumda, güvenilir bir kümeleme işlemi gerçekleştirilemez. İkinci yol kümenin doğasına en uygun ya da yapısına en yakın verinin seçilmesidir. Bu yol birinci yola göre daha güvenilir kümeleme sonuçları vermektedir. Küme yapısını ölçmek için de küme elemanlarının ortalama değeri kullanılır.

  Tez çalışmasında gerçekleştirilen uygulamalarda, daha güvenilir olduğu için ikinci seçenek tercih edilmiştir. Kümelemenin başarılı olması için başlangıçta verilen K_{init} ve N_{init} sayıları büyük seçilmelidir. Böylece veri kümesi daha iyi temsil edilebilecektir. Çünkü sonraki adımlarda zaten en kötü küme bulunup o kümenin elemanları fark temelli entropi kümeleme ile diğer kümelere aktarılacak ve küme sayısı bir azaltılmış olacaktır. Bu azaltma işlemi küme sayısı ikiye indirilinceye kadar devam ettirilebilir. Ayrıca N_{init} sayısına da giderek küçülen değerler verilip bu değerler için bulunan entropi ölçüleri mukayese edilebilir.

- *Kümeler Arası Entropi:* Küme içi entropi hesaplanırken kullanılan denklem (3.8) yerine, tüm veri örnekleri üzerinde çalışan ve bir üyelik fonksiyonu kullanan yeni bir formül tasarlanmıştır (Jenssen ve ark., 2003). Eğer x_i ve x_j örnekleri farklı kümelere aitse, $M(x_{ij})$ üyelik fonksiyonu 1 değerini alır. Aynı kümeye aitse sıfır değerini alır. Buna göre kümeler arası entropi denklem (3.9) ile ifade edilir.

$$H(C_i) = -logV(C_i), \qquad i=1,\ldots,k. \tag{3.9}$$

K sayıda küme üzerinde çalışıldığında, küme değerlendirme fonksiyonu eşitlik (3.10)'da belirtilen biçime dönüşür (Gokcay ve Principe, 2002).



Şekil 3.3 En Kötü Kümenin Belirlenmesi

$$V(C_i) = \frac{1}{2\prod_{k=1}^{K} N_k} \sum_{i=1}^{N}\sum_{j=1}^{N} M(x_{ij}) e^{-\frac{(x_i - x_j)^2}{2\sigma^2}}, \qquad i=1,\ldots,k. \tag{3.10}$$

Eğer çalışılan veriler üzerinde iyi bir ayrıştırma yani iyi bir kümeleme gerçekleştirildiyse, eşitlik (3.10)'dan küçük bir değer ve dolayısıyla eşitlik (3.9)'dan büyük bir değer elde edilecektir. Bu kıstastan kümeleme sırasında faydalanılacaktır.

- *En Kötü Kümenin Belirlenmesi:* Şekil 3.3'te en kötü kümenin bulunması için kümeler arası entropinin nasıl kullanıldığı gösterilmektedir.

Şekil 3.3 kısım (a) da, veri grubu 3 kümeye ayrılmıştır. Her bir değerlendirme aşamasında, bir küme veri grubundan çıkarılır ve kalan veri grubu için kümeler arası entropi değeri hesaplanır. Örneğin, 3 küme için gösterilen şekil 3.3te, 3 ayrı kümeler arası entropi değeri hesaplanır. Çünkü sırayla 1., 2. ve 3. kümeler veri grubundan çıkarılmıştır. Veri grubundan çıkarıldığında, kalan veri grubu için kümeler arası entropi değeri en yüksek çıkan küme en kötü küme olarak belirlenir. Çünkü kalan kümeler en iyi ayrılan kümelerdir. Örneğin şekil 3.3'teki gösterimde en kötü küme c_1 veya c_2 çıkacaktır.

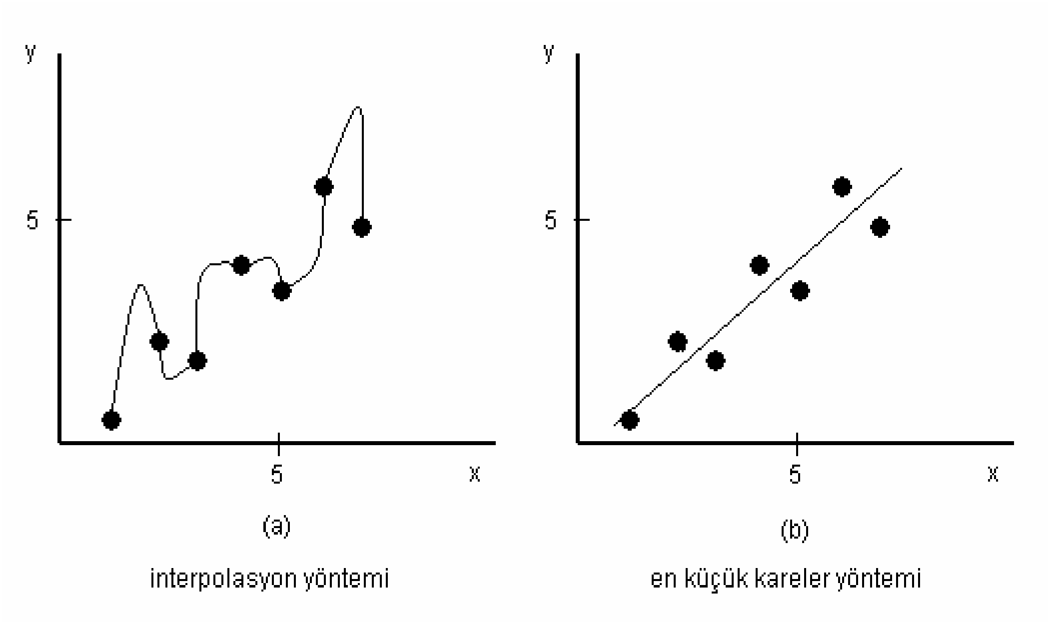
- *Sonuç Durumu Kümelerine Karar Verilmesi:* Her bir durumda veri grubundaki küme sayısı bir azalır ve çıkarılan kümenin elemanları diğer kümelere fark tabanlı entropi yaklaşımına göre aktarılır. Her bir küme sayısı için hesaplanan entropi değerleri tutulur ve bu liste bir kümeleme hiyerarşisi oluşturur. İki küme kalana kadar işlem devam ettirilir. Sonuç durumuna bu hiyerarşi değerleri arasından karar verilmelidir. Her bir adımda, kümeler arası entropi değeri tüm kümeler üzerinde denklem (3.9)'a göre hesaplanır. Küme sayısı azaltılmadan önceki ve sonraki durumlar için bulunan bu kümeler arası entropi değerleri karşılaştırılır. Deneysel uygulamalardan da elde ettiğimiz sonuçlara göre, küme sayısı azaltılırken entropi değeri doğru küme sayısına ulaşıncaya kadar artar ve bu noktada tekrar azalır. Böylece en iyi küme sayısı da tespit edilmiş olur. Başlangıç kümeleri rasgele belirlenen veriler üzerinde şekillendirildiği için aynı veri kümesinde birkaç kez kümeleme yapıldığında farklı entropi değerleri ve farklı küme sayıları elde edilebilir. Bu sonuçlar arasından en iyisi ise, yine kümeler arası entropi değerleri ile tahmin edilebilir. Kümeler arası entropi değeri diğer durumlara göre daha düşük olan durum en iyi durum olarak belirlenir.

3.3 Regresyon

Bir problemin çözümü ile ilgili olarak üzerinde çalışılan veriler çoğu zaman belirli bir karakteristiği yansıtacak şekilde ve sürekli bir yapıda dağılım gösterirler. Elde mevcut olmayan verilerin sahip olduğu bir bağımsız değişken değeri için bağımlı değişken değerinin ne olacağı gerekli olabilir. Bu durumda, genel olarak eğri uydurma ismi verilen yöntemler kullanılır.

Eğri uydurma yaklaşımlarında elde bulunan verilerin hata oranının az veya çok olmasına bağlı olarak iki genel yol izlenebilir (Chapra ve Canale, 2003). Bunlardan birincisi verilerin hata oranının çok olduğu bilgisine dayanır. Bu durumda verilerin tümünü sağlayacak yapıdaki bir eğrinin üretilmesi gereksizdir. Bu yüzden amaç veri grubunun genel eğilimini yansıtan eğrinin oluşturulmasıdır. Bu yapıya göre geliştirilen metot En Küçük Kareler Regresyonu olarak adlandırılır (Chapra ve Canale, 2003).

Verilerin hata oranının az olduğunun varsayıldığı ikinci yaklaşımda ise, veriler güvenilir olduğu için bu verilerin tümünden geçecek bir eğri üretilmeye çalışılır. Bu yapıyı temel alarak geliştirilen metot ise interpolasyon olarak adlandırılır. Bahsedilen iki yaklaşım şekil 3.4'te gösterilmiştir.



Şekil 3.4 Eğri Uydurma Yöntemleri

Tez çalışmasında, doğruluğu şüpheli veri kümeleri üzerinde çalışıldığı için ve DVM metodu yapısına da daha uygun olduğu için En Küçük Kareler Regresyon metodu tercih edilmiştir. Çünkü veriler, gerek hazır veritabanlarından elde edilen veriler olsun, gerekse araştırmacılar tarafından bizzat ham olarak alınan ve işlenen veriler olsun belirli bir miktar gürültü içermektedirler.

Geçmişte yapılan bazı çalışmalarda, verilerin dağılımına uyan doğrunun çizilmesi işlemi grafik tabanlı olarak göz kararı çizimlerle de gerçekleştirilmiştir. Ancak bu durumda farklı analizciler farklı doğrular çizebilmektedir. Bu yüzden bahsedilen doğrunun oluşturulması amacıyla En Küçük Kareler Regresyonu tasarlanmıştır. Bu metot standart bir matematiksel kalıba uygun olarak tasarlandığı için, işlemi gerçekleştiren analizciler arasındaki uygulama farklılıkları da giderilmiştir.

3.3.1 Doğrusal En Küçük Kareler Regresyonu

En Küçük Kareler Regresyonunun en temel hali doğrusal eğri uydurmadır. Elde bulunan (x_1,y_1), $(x_2,y_2),\ldots$, (x_n,y_n) şeklindeki veri kümesine göre bu veri kümesinin genel karakteristiğini en iyi ifade eden doğrunun bulunmasıdır. Bu doğru eşitlik (3.11) ile ifade edilir.

$$y = a_0 + a_1 x + e \tag{3.11}$$

Burada a_0 kesme noktasını, a_1 ise eğimi ifade eder. Doğru denklemine eklenen e ise, model ile gözlemler arasındaki hatadır. Uydurulabilecek en iyi doğru, bütün veriler için oluşacak bu hata değerlerinin toplamını minimum yapan doğrudur. Bu işlem matematiksel olarak eşitlik (3.12) ile yapılandırılır;

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i) \tag{3.12}$$

Fakat bu eşitliğin bir sakıncası vardır. Bu sakınca da; oluşturulan doğrunun alt ve üst tarafında aynı büyüklükte hatalar meydana geldiğinde, bu değerlerin birbirlerini toplamda sıfırlamaları ve hatanın toplam değerini minimuma ulaştırmalarıdır.

Denklem (3.12)'deki eksiklik düşünülerek daha sonra hataların mutlak değer cinsinden ölçüleri toplamının minimum yapılması düşünülmüştür. Böylece denklem (3.12), denklem (3.13)'e dönüşmüştür.

$$\sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |y_i - a_0 - a_1 x_i| \tag{3.13}$$

Şekil 3.5, bu kriterin de bazı durumlarda uygun olmadığını gösterir. Şekilde gösterilen 4 nokta için çizilen kesikli çizgiler arasındaki her bir doğru, hataların mutlak değerleri toplamını minimum yapacaktır. Ancak bu doğrulardan sadece bir tanesi ulaşılmak istenen en iyi doğrudur. Bu yüzden eşitlik (3.13)'e göre geliştirilen fikir de en iyi doğruyu vermez.

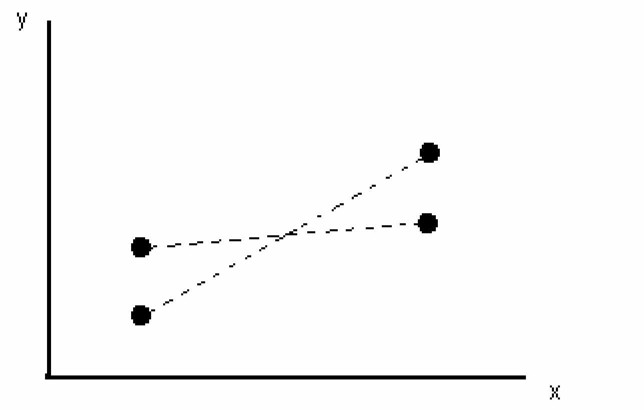
Yukarıda anlatılan yaklaşımların eksikliklerini gideren yaklaşım ise, ölçülen değerlerle modelin verdiği değerler arasındaki hataların karelerinin toplamını minimum yapan yaklaşımdır. Bu yaklaşım eşitlik (3.14) ile ifade edilir ve En Küçük Kareler Regresyonu olarak adlandırılır.

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_{i,ölçülen} - y_{i,model})^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \tag{3.14}$$

Eşitlik (3.14)'te çıkarılan hata oranına göre, a_0 ve a_1 değerlerinin belirlenmesi için, eşitliğin a_0 ve a_1 değerlerine göre türevleri alınıp sıfıra eşitlenmelidir. Bu işlemlerden sonra da (3.15)'teki eşitlikler elde edilir.

$$\begin{aligned} n.a_0 + (\sum_{i=1}^{n} x_i) a_1 &= \sum_{i=1}^{n} y_i \\ (\sum_{i=1}^{n} x_i) a_0 + (\sum_{i=1}^{n} x_i^2) a_1 &= \sum_{i=1}^{n} y_i . x_i \end{aligned} \tag{3.15}$$

Bu eşitliklerde iki bilinmeyen ve iki denklem bulunduğu için a_0 ve a_1 değerleri hesaplanabilir. Böylece, doğru denklemi de kurulmuş olur.



Şekil 3.5 Mutlak Hata Tabanlı Eğri Uydurma

Örnek: En Küçük Kareler Regresyonu, Bir akaryakıt şirketi ısınmak amacıyla sattığı gaz miktarı ve günlük sıcaklık arasındaki ilişkiyi belirlemek istemiştir. *x* günlük ortalama sıcaklığı, *y* gaz satış miktarını göstermek üzere satış ve sıcaklık ile ilgili veriler Tablo 3.1'de verilmiştir.

Tablo 3.1 En Küçük Kareler Regresyonu Örneği

Günlük Ortalama Sıcaklık (x)	Satışlar (bin litre) (y)
-17	44
-14	35
-11	33
-10	26
-7	26
-6	19
-4	17
-1	10
3	8
7	4

x üzerinde *y*'nin regresyon doğrusunun denklemini bulunuz.

Çözüm: Verilerden; $\Sigma x_i = -60$, $\Sigma y_i = 222$, $\Sigma x_i y_i = -2183$, $\Sigma x_i^2 = 866$

$$a_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{(10)(-2183) - (60)(222)}{(10)(866) - (60)^2} = \frac{-21830 + 13320}{8660 - 3600}$$

$$a_1 = \frac{-8510}{5060} \cong -1{,}68$$

$$a_0 = \frac{\Sigma y - a1\Sigma x}{n} = \frac{222 - (-1{,}68)(-60)}{10} = \frac{222 - 100{,}8}{10} = 12{,}12$$

Tahmin edilmiş regresyon doğrusu:

$\hat{y}_i = 12{,}12 - 1{,}68x_i$

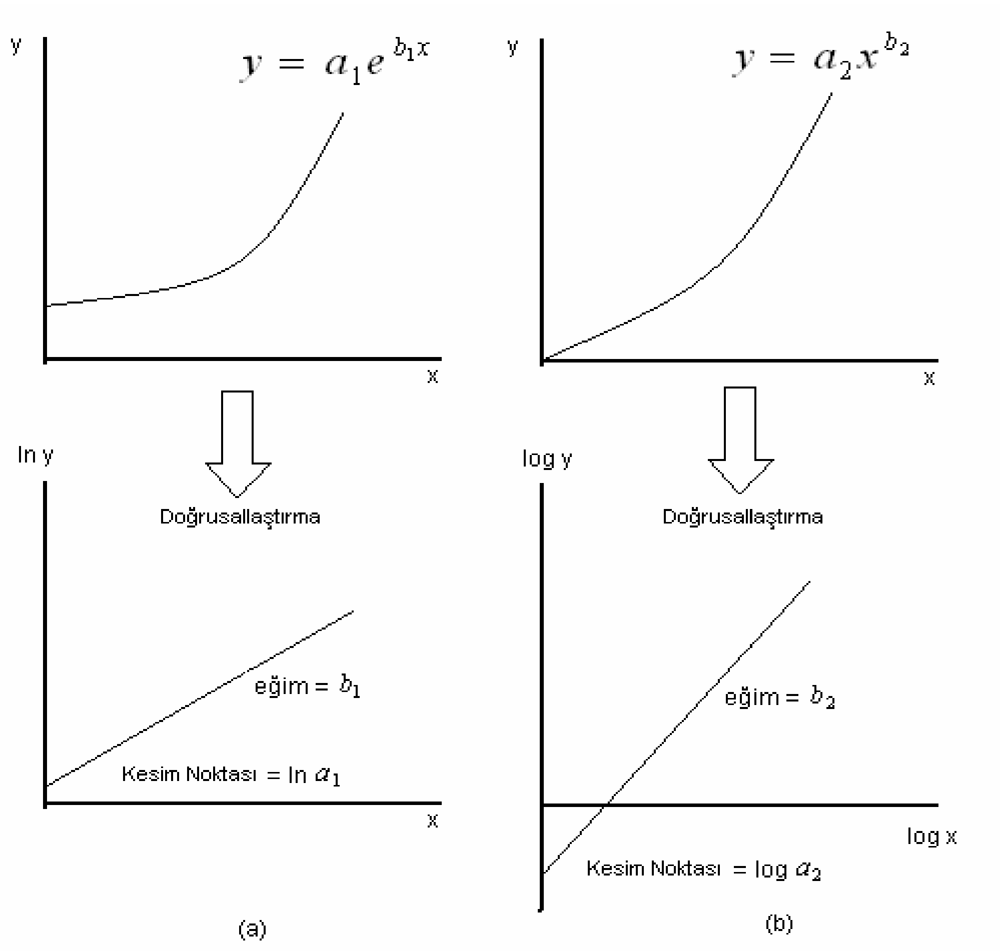
3.3.2 Doğrusal Olmayan En Küçük Kareler Regresyonu

En Küçük Kareler Regresyonu doğrusal veriler üzerinde düşünülerek kurulmuş bir eğri uydurma yöntemidir. Fakat çalışılan veriler her zaman doğrusal bir dağılım göstermezler. Bu tür durumlarda dönüşümler uygulanmalıdır. Yani, doğrusal olmayan veri dağılımı doğrusal hale getirilmelidir.

Tez çalışmasında, DVM ile ilgili olarak YTF temelli çekirdek fonksiyonu kullanılmıştır. Bu fonksiyona uygun bir fonksiyon olması nedeniyle, DVM parametre düzenleme işleminde gerçekleştirilen uygulamada da şekil 3.6 (a)'da gösterilen üstel tabanlı fonksiyon tercih edilmiştir. Bu fonksiyon mühendisliğin birçok alanında kullanılmaktadır. Dönüşüm yapılırken ise, ilk önce eşitliğin iki tarafının logaritmik değeri alınır.

Böylece, şekil 3.6 (a)'da gösterildiği gibi bulunan yeni doğrusal fonksiyonun kesme noktası ve eğimi bulunmuş olur.

Şekil 3.6 (b)'de ise, üstel yapıdaki logaritmik tabanlı bir fonksiyonun dönüşüm işlemi gösterilmektedir.



Şekil 3.6 Örnek Regresyon Dönüşümleri

4. GELİŞTİRİLEN DVM EĞİTİM ALGORİTMALARI

Tez çalışmasında, daha önceden belirlenen bir takım hedeflere yönelik eğitim algoritmaları geliştirilmiştir. Her ne kadar DVM son zamanlarda yaygın bir şekilde kullanılsa ve birçok uygulama alanında karşılaşılan problemlere tatmin edici çözümler getirse de kuvvetlendirilmesi gereken yönleri vardır. Bu yönler aşağıdaki gibi sıralanabilir;

- Çok büyük sayıda örneklem içeren veri kümeleri üzerinde DVM metodunun eğitilmesinin çok zaman alması,
- DVM metodunda kullanılan çekirdek fonksiyonu ve ayırıcı fonksiyon parametrelerinin en uygun değerlerinin bulunması,
- DVM sınıflandırıcısının gürültülü verilere karşı hassaslığı,
- Ayırıcı fonksiyon bulunurken, bu düzlemin en az karmaşıklıkla aynı zamanda mümkün olduğunca en yüksek doğrulukta olması,
- Son olarak da, DVM ile çok etiketli sınıflandırma işleminde karşılaşılan problemler.

Tez çalışmasında, bu problemlerden bir veya daha fazlasına çözüm getirecek yeni DVM öğrenme algoritmaları geliştirilmeye çalışılmıştır. Geliştirilen metotlar üç ana başlıkta sunulmuştur.

4.1 Kümeleme Tabanlı K En Yakın Komşuluk Destek Vektör Makinesi

DVM yeni ve popüler bir öğrenme makinesi olmasına rağmen yukarıda bahsedilen bir takım zayıf yönleri de bulunmaktadır. Bu zayıflıkların giderilmesi amacıyla geliştirilmiş olan ve birçok uygulama alanında yaygın olarak kullanılan metot bölüm 2.5.5'te bahsedilen En Küçük Kareler Destek Vektör Makineleri (EKKDVM) dir. EKKDVM, standart DVM ne göre daha kısa zamanda eğitim yapabilen ve bununla birlikte genelleme yeteneğinde ise küçük miktarda da olsa düşüş gözlenebilen bir DVM öğrenme metodu çeşididir. Bu özelliğinden dolayı geliştirilen KKDVM (Kümeleme Tabanlı K En Yakın Komşuluk Destek Vektör Makinesi) yöntemi yeni bir eğitim algoritması kullanan ve EKKDVM metoduna alternatif bir metot olarak görülebilir. Ayrıca KKDVM metodunun bazı ek özellikleri de mevcuttur. Eğitim algoritmasının bu özelliklerden ayrıntılı olarak bölüm 4.1.1'de bahsedilmiştir.

4.1.1 KKDVM Öğrenme Algoritması

Geliştirilen KKDVM öğrenme algoritması, özellikle büyük veri kümeleri üzerinde etkili biçimde eğitim yapabilecek bir algoritmadır. Eğitim verileri üzerinde yerel kontrol yeteneği de bulunan eğitim zamanı düşük bir metottur. Ayrıca, konu ile ilgili olarak taranmış olan çalışmaların hiçbirinde bulunmayan "Destek Vektörlerinin Olasılık Tabanlı Değerlerle Temsili" ilk kez bu çalışmada gerçekleştirilmiştir (E. Çomak ve A. Arslan 2007). Öğrenme algoritması dört temel aşamadan oluşmaktadır. Bu aşamalardan üçü eğitim, sonuncusu ise test ile ilgilidir.

1. Aşama: Bu aşamada, ilk önce YTF çekirdek fonksiyonu aracılığıyla giriş uzayından nitelik uzayına geçiş yapılmıştır. Daha sonra bu uzay üzerinde kümeleme işlemi gerçekleştirilmiştir. Burada var olan bir kümeleme metodunun kullanılmasından ziyade işlemin daha kolay ve daha çabuk olması için basit bir kümeleme planı işletilmiştir. Küme sayısı her iki sınıf için eşit olacak şekilde önceden belirlenen bir değerde alınmıştır.

Bu kümeleme planına göre, önce her iki sınıf için bağımsız olarak verilerin birbirlerine uzaklıklarından oluşan uzaklık matrisi oluşturulmuştur. Bu uzaklıklar eşitlik 3.1'de verilen öklit formülü temel alınarak hesaplanmıştır. Sonra her bir sınıf için uzaklık matrisindeki en büyük değer bulunarak birbirlerine en uzak iki veri noktası tespit edilmiştir. Bir başka deyişle, bir sınıf için iki nokta olmak üzere toplam dört veri noktası belirlenmiştir. Daha sonra yine her sınıf için, öncelikle birinci uç nokta referans alınarak, belirlenmiş olan küme sayısı ve iki uç nokta arası maksimum mesafe yardımıyla veriler kümelere ayrılmıştır. Aynı işlem bir kez de ikinci uç nokta referans alınarak tekrarlanmıştır. Her iki durumda da kümelere düşen veri sayılarına göre standart sapma değerleri bulunmuş ve standart sapması daha düşük olan uç nokta referans noktası olarak kabul edilmiştir. Bu işlemin gerçekleştirilmesinin sebebi ise, veri gruplarının kümelere mümkün olduğunca homojen olacak şekilde dağılmasının istenmesidir. Böylece kümeler daha iyi temsil edilebilecektir. Standart sapması düşük olan nokta referans alınarak gerçekleştirilen kümeleme de uygulamada kabul edilen kümeleme olarak belirlenmiştir.

2. Aşama: Küme sayıları ve her bir kümeye ait olan veriler birinci aşamada belirlenmiştir. İkinci aşamada öncelikle bu verilerin ortalamaları alınarak her kümenin merkezi belirlenir. İki sınıf için de küme sayısı eşit olarak alındığından aynı sıradaki 1. ve 2. sınıf kümeleri karşılıklı olarak ele alınır. Karşılıklı küme merkezleri arasındaki mesafe 1 olacak şeklide normalize edilerek bu karşılıklı kümelerde bulunan tüm verilerin 1. sınıf merkezine olan uzaklıkları (n+1)inci nitelik ve 2. sınıf merkezine olan uzaklıkları da (n+2)inci nitelik değeri olarak her bir veriye eklenir.

Bu aşamada, uygulamada genellikle [1,1.5] civarında değerler verilen bir eşik değeri belirlenir. (n+1). veya (n+2). niteliğinin değeri bu eşikten büyük olan veriler veri kümesinden çıkartılır. Bu işlem, hem eğitim için önemsiz olan verilerden kurtulmayı hem de sistemin işlem yükünün azalmasını sağlar.

Bundan sonra geride kalan veriler üzerinde KEYK metoduna göre en yakın komşuluk sayıları bulunur. Bu nedenle de, KEYK metodunda gerekli olan komşuluk sayısının *(k)* tespit edilmesi şarttır. Bu sayıyı standart bir biçimde belirlemek için de aşağıda ifade edilen denklem (4.1) kullanılır.

$$k_{ij} = \begin{cases} f(\sqrt{c_{ij}}), & \text{eğer } \mathrm{f}\left(\sqrt{c_{ij}}\right) mod\, 2 = 1 \\ f(\sqrt{c_{ij}}) - 1, & \text{diğer durumlarda} \end{cases} \qquad (4.1)$$

Burada j sınıfların, i ise kümelerin temsilidir. Bir diğer temsil olan c_{ij} ise, j sayılı sınıfın i sayılı kümesindeki eğitim verisi sayısıdır. Fonksiyon f $\left(\sqrt{c_{ij}}\right) mod 2 = 1$, aldığı değerin tamsayı kısmını döndürmektedir, ondalık kısmını ise ihmal eder. Bulunan k_{ij}, j sayılı sınıfın i sayılı kümesinin komşuluk değeridir. (4.1) eşitliğine göre belirlenen değer ilgili küme verileri için aranan maksimum komşuluk sayısıdır.

Her bir eğitim verisinin ait olduğu küme için hesaplanan k_{ij} sayısı ile temsil edilen komşusu belirlenir ve bu komşularının etiketine bakılır. Bu komşular arasında kendisiyle aynı etikete sahip komşularının sayısı veriye (n+3)üncü nitelik olarak eklenir.

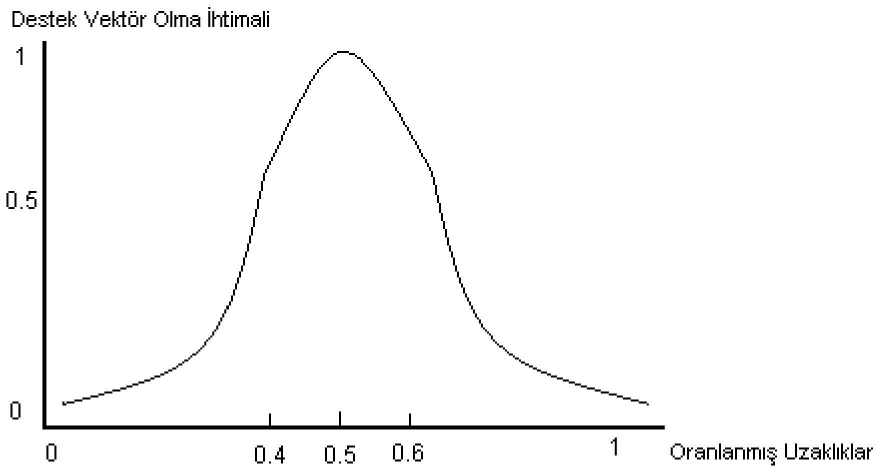
3. Aşama: Bu aşamada, her veri üzerinde öncelikle (n+3)üncü nitelik değerinin k_{ij} değerinin yarısından az olup olmadığının incelemesi yapılır. Az olduğu görülürse, bu tip veriler diğer verilerden ayrılarak "Destek Vektör Olma İhtimalleri" sıfırlanır. Geri kalan tüm verilerin az ya da çok "Destek Vektör Olma İhtimalleri" vardır. Bu ihtimaller şu plana göre bulunur;

1. İterasyon sayısı 1 e ayarlanır,
2. (n+3)üncü niteliğinin değeri, (k_{ij} + 1 – iterasyon sayısı) olan veriler listelenir,
3. Üzerinde çalışılan küme verileriyle ilgili olarak uzaklıklar farklarını (sınıf 1 de (n+2)inci nitelik değeri – (n+1)inci nitelik değeri, sınıf 2 de ise (n+1)inci nitelik değeri – (n+2)inci nitelik değeri) bul ve minimum olan değer kaydedilir,
4. İterasyon sayısına 1 ekle ve sınıf j, küme i için iterasyon sayısı round(k_{ij}/2) değerine eşit veya küçükse adım 2 ye geri dönülür,
5. Kaydedilen değerler azalıyorken o verilerin (n+3) üncü niteliğinin değeri de azalıyorsa, kaydedilen değerler arasında minimum uzaklıklar farkına sahip olan veri tam olasılıklı olarak "Destek Vektör" noktası seçilmiştir. Diğer durumlarda, (n+3) üncü nitelik değeri en yüksek olan veri 1 tam olasılıklı olarak "Destek Vektör" noktası seçilmiştir. Yapılan çalışmalarda, bu tam olasılıklı "Destek Vektör" noktalarına "Ana Destek Vektör" noktaları adı verilmiştir (Çomak ve Arslan, 2007).

Ayrıca tez çalışmasında, destek vektör olma ihtimalini belirli bir seviyede sınırlandırmak için bir de tolerans değeri belirlenmiştir. Bu değerin yüksek seçilmesi halinde, destek vektör olma ihtimali olan veri sayısı artacak ve hesap yükü de az da olsa fazlalaşacaktır. Tolerans değeri çalışmalarımızda [0.05,0.15] civarında ve verilerin küme merkezlerine olan oranlanmış uzaklıkları temel alınarak seçilmiştir. Ana destek vektörün (n+1) ve (n+2). nitelik farklarına oranlanmış tolerans değeri eklenir. Ana destek vektörlerle ilgili başlangıç şartlarını sağlayan ve aynı zamanda (n+1) ve (n+2) inci nitelik farkları da toleransı ekleyerek bulunan bu yeni değerden az olan veri noktaları da eşitlik (4.2) ile hesaplanan ihtimalde destek vektör olur.

$$p(x_i) = \frac{1}{exp\left(\left\|x_i(1,n+1) - x_i(1,n+2)\right\| - \left\|sv(1,n+1) - sv(1,n+2)\right\|\right)} \tag{4.2}$$

Burada *sv* ana destek vektörü, x_i ise belirli bir olasılıkla destek vektör olabilen yardımcı destek vektör noktasını temsil etmektedir. Eşitlik yardımcı destek vektör noktasının “Destek Vektör Olma İhtimali” ni vermektedir. Verilerin destek vektör olma ihtimali şekil 4.1 de gösterilen dağılıma uymaktadır. Şekilde gösterilen dağılım, ana destek vektörün mesafeler farkı 0.5 ve tolerans oranıyla bulunan oranlanmış uzaklığın ise 0.32 olduğu durum içindir.



Şekil 4.1 Oranlanmış Uzaklıkların Gauss Dağılımı

Bu yapının kullanılmasıyla birlikte, küme başına en az bir ana destek vektör belirlenecektir. Böylelikle, ana destek vektörler ve yardımcı destek vektörler belirlenmiş olur. Daha sonra, bu noktalara göre regresyon eğrisi uydurularak ayırıcı fonksiyon bulunmaktadır. Ayırıcı fonksiyon her karşılıklı küme grubu için farklı çıkabileceğinden parçalı bir fonksiyon yapısında olacaktır.

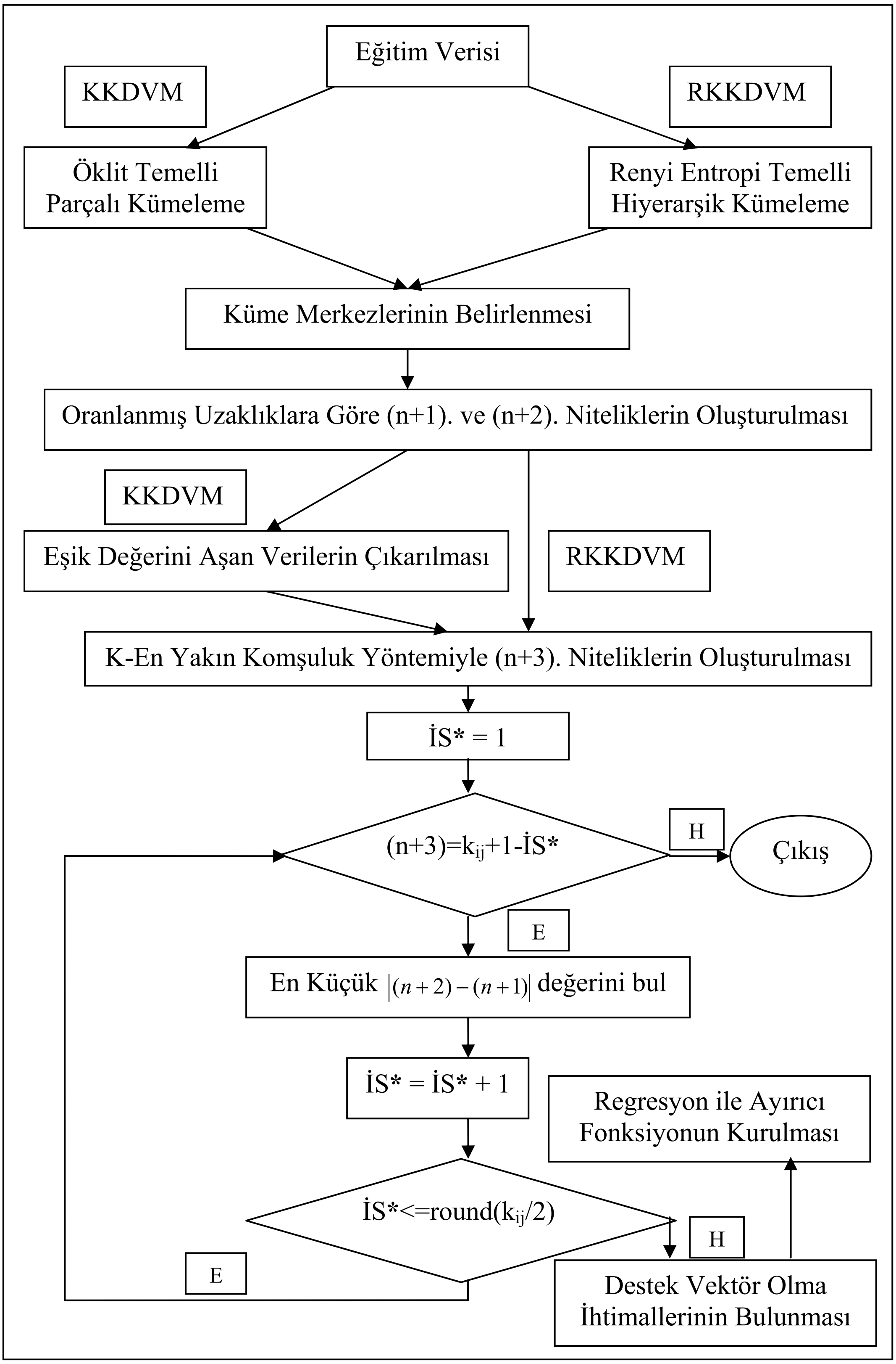
4. Aşama (Test Aşaması): KKDVM test aşamasında standart DVM den farklı olarak, öncelikle alınan verinin hangi küme grubuna ait olduğu belirlenir. Bu belirlenme esnasında, test verisinin sırayla karşılıklı küme merkezlerine olan uzaklık değerleri çarpılmaktadır. Çarpım sonucu en küçük olan küme grubuna atanmaktadır. Küme grubu belirlendikten sonraki işlemler standart DVM ile aynıdır.

KKDVM deki bir diğer yenilik, çok sınıflı sınıflandırma aşamasındadır. Bölüm 2.5.4.1 ve 2.5.4.2 de anlatılan doğrudan ve dolaylı çok sınıflı sınıflandırma yaklaşımlarından farklı yeni bir yaklaşım geliştirilmiştir. Bu gerçekleştirilirken destek vektörler ve test verisi arasındaki öklit uzaklığına dayanan ağırlıklandırılmış ortalama yaklaşımı kullanılmıştır. Bu yaklaşım eşitlik (4.3)'e dayanmaktadır.

$$ao = \frac{1}{n_{sv}} \sum_{i=1}^{n_{sv}} w_i EU(x_i, x) \qquad (4.3)$$

Burada *ao* ağırlıklandırılmış ortalama değeri, n_{sv} çalışılan kümedeki destek vektörlerin sayısını, $EU(x_i,x)$ test verisi ile *i* indisli destek vektörü arasındaki öklit mesafesini ve w_i de *i* indisli destek vektörün "destek vektör olma ihtimali" ni yansıtmaktadır. Bu ihtimalin çok sınıflı sınıflandırmada kullanılması, sınıflandırılamayan bölge problemini de çözmektedir. Ayrıca öklit uzaklıkları ile destek vektörlerin temsil kabiliyeti oranında değerlendirilmesi sağlanmaktadır. Test verisi standart dolaylı DVM'deki *+1* ve *-1* çıkışlarıyla işlem yapmak yerine bu ağırlıklı çıkışlarla işlem yapmaktadır. Etiketi bilinmeyen test verisi sınıflar arasındaki en küçük ağırlıklandırılmış ortalama değerine sahip sınıf etiketine atanmaktadır.

Şekil 4.1'de bu bölümde anlatılan KKDVM ve sonraki bölümde anlatılan RKKDVM eğitim algoritmalarının akış şeması verilmektedir.



Şekil 4.1 KKDVM ve RKKDVM Eğitim Algoritmalarının Akış Şeması

* : İS, İterasyon Sayısını temsil etmektedir.

4.2 Renyi Kümeleme Tabanlı K En Yakın Komşuluk Destek Vektör Makinesi

RKKDVM eğitim metodu, önceki bölümde anlatılan metottan farklı olarak sadece Renyi entropi temeline göre kümeleme yapmaktadır. Önceki metotta sadece sınıf verileri arasındaki öklit uzaklıkları hesaplanarak mesafeye dayalı bir kümeleme işlemi yapılmıştır. Renyi entropi temelli kümelemenin önceki kümelemeye göre önemli avantajları bulunmaktadır. Bölüm 3.2.3 te Renyi entropi kümeleme metodu anlatılırken detaylı olarak açıklanan bu avantajlar, sınıflandırma işleminin de daha güvenilir olmasını sağlamıştır. Bu sistemin tek dezavantajı ise, KKDVM deki kümelemeye oranla biraz daha fazla zaman almasıdır.

4.2.1 RKKDVM Öğrenme Algoritması

Bölüm 4.1.1 de açıklanan eğitim algoritmasından tek farkı 1. aşamada kümeleme için Renyi entropinin kullanılmış olmasıdır. RKKDVM algoritmasının KKDVM ye göre avantajları şunlardır;

- KKDVM de kullanılan kümeleme sürecinde, küme sayısı başlangıçta belirtilmelidir. Kümeleme işlemi bu sayıya göre gerçekleştirilir ve kümeleme boyunca değiştirilemez. RKKDVM ise, başlangıçta küme sayısının verilmesine ihtiyaç duymamaktadır. Başlangıçta, sadece veri kümesinin genelini temsil edebilecek sayıda K_{init} ve N_{init} değerinin Renyi entropi metodu çalıştırılırken belirtilmesi yeterlidir.

- KKDVM'de kullanılan kümeleme sürecinde, küme sayısını sabit alarak işlem yapıldığından parçalı bir kümeleme metodudur. RKKDVM de kullanılan ise, hiyerarşik bir kümelemedir. Çünkü başlangıçta verilen büyük değerdeki küme sayısını entropi değerlerine göre otomatik olarak birer birer azaltarak en iyi küme sayısını bulunur. Bu özelliği de, RKKDVM nin veri setine en uygun kümeleme dağılımını daha doğru bir şekilde seçebilmesini sağlamaktadır.

- Renyi entropi kümelemede, KKDVM kümeleme sürecinde olduğu gibi sadece veriler arasındaki ikinci dereceden istatistiksel ilişkilere (öklit uzaklıkları) bakılmamaktadır. Küme içi entropi, kümeler arası entropi ve fark temelli entropi ölçülerine göre değerlendirme yapılmaktadır. Böylece, kümeleme işlemi daha sağlıklı yapılacağından küme elemanları ve küme merkezleri daha gerçekçi bir şekilde bulunacaktır. Bu sayede, eğitim algoritması da daha iyi sonuçlar vermektedir.

- RKKDVM'nin bir diğer avantajı da, KKDVM'de bazı verileri elemek amacıyla kullanılan eşik değerlerinin kullanımına ihtiyaç duyulmamasıdır.

4.3 DVM Öğrenme Parametrelerinin Renyi Kümeleme Tabanlı Bir Yaklaşımla Düzenlenmesi

Genel olarak DVM iki çeşit öğrenme parametresine sahiptir. Bunlardan bir tanesi, giriş uzayından nitelik uzayına geçişi sağlayan çekirdek fonksiyonlarında kullanılan parametredir. Diğeri de, DVM ayırıcı düzleminin karmaşıklığı ile sınıflandırma başarımı arasında bir tercih yapılmasını sağlayan C parametresidir. Bu parametrelere farklı değerlerin verilmesi, ayırıcı fonksiyonun yapısını ve dolayısıyla da sınıflandırma başarımını etkilemektedir. Bu yüzden bu parametrelerin en iyi sınıflandırma başarımını verecek değerlerinin bulunması, DVM metodu için son yıllarda önemli bir araştırma konusu haline gelmiştir (Lin ve ark., 2007).

Bu bölümde, yardımcı metotlarla birlikte en uygun DVM öğrenme parametrelerini bulan bir yaklaşım önerilmiştir. Bu amaçla kullanılan metotlar, farklı bir yönden ele alınan Renyi entropi temelli kümeleme ve üssel regresyon tabanlı eğri uydurma metotlarıdır. Parametrelerin düzenlenmesini sağlayan yaklaşım iki adımdan oluşur. İlk adımda Renyi entropi temelli kümeleme işlemindeki entropi ölçüleri yardımıyla çekirdek fonksiyonunun en uygun parametre değeri bulunur. Sonraki adımda DVM sınıflandırıcısı bulunan σ değeri ile farklı birtakım C değerleri kullanılarak çalıştırılır. Bu C ve sınıflandırma doğruluğu değerlerine göre üstel bir regresyon eğrisi yapılandırılır. Böylece en iyi C değeri de bu regresyon fonksiyonu ile doğrudan bulunabilir.

Tez çalışmasında, çekirdek fonksiyonu olarak YTF seçilmiştir. Bunun sebebi de YTF nunun hem en çok kullanılan çekirdek fonksiyonlarının başında gelmesi hem de Renyi entropide kullanılan hesaplamalara olan benzerliğidir. YTF de kullanılan parametre σ parametresidir. σ parametre değerinin değişmesi veri noktalarının nitelik uzayındaki yerleşme şeklini etkilemektedir. Dolayısıyla destek vektör noktalarının ve buna göre kurulan ayırıcı fonksiyonun yapısıda bu durumdan etkilenmektedir. C parametresi ise, hangi çekirdek fonksiyonu seçilirse seçilsin DVM de kullanılan standart bir parametredir. Geliştirdiğimiz yaklaşım, iki aşamalı olarak her iki parametrenin de optimum değerini bulabilmektedir.

4.3.1 Parametre Düzenleyici Yaklaşım Algoritması

Gerçekleştirilen uygulamada, parametre düzenleyici yaklaşımın 1. adımı EKKDVM sınıflandırıcısı çalıştırılmadan önce *[-1,+1]* aralığında oranlanmış veriler üzerinde işlenmiştir. 2. adımda optimum σ değeri bulunduktan sonra bu bulunan değerle EKKDVM farklı C değerleri yardımıyla eğitilmiştir. Yaklaşım şu şekilde çalışır;

1. Aşama (σ değerinin belirlenmesi): Renyi entropi kümelemede de kullanılan K_{init} ve N_{init} değerleri, başlangıçta veri kümesini mümkün olduğunca iyi temsil edebilecek şekilde rasgele seçilmekte ve Renyi entropi kümeleme metodu çalıştırılmaktadır. Küme sayısı 2 olduğu andaki entropi değeri tutulmaktadır. Farklı K_{init} ve N_{init} başlangıç değerleri için bulunan değerler de aynı şekilde saklanmaktadır.

Renyi entropide de kullanılan σ değeri küçük bir sayıda (gerçekleştirilen uygulamalarda 0.001 değeri) sabit tutulmaktadır. K_{init} ve σ sabitken, N_{init} *2* den başlayarak artan değerler verildiğinde entropi değeri belirli bir *N'*e kadar azalır. Bundan sonra tekrar artışa geçer. Bu *N* değeri yerel optimum *N* değeridir. Renyi entropi kümelemede, başlangıç küme merkezleri rasgele atandığından, bu işlem farklı merkez noktaları için defalarca çalıştırılıp en iyi entropi değerini veren *N* değeri global optimum *N* değeri olarak belirlenir.

Bulunan optimum K_{init} ve N_{init} için, σ değeri 0.001 gibi bir değerden başlatılıp yine 0.001 değer aralıklarıyla artırılarak bu σ değerlerine karşılık gelen entropi ölçülerindeki değişimin aynı durum için 0.001 den az olduğu noktaya kadar artırılır. Bu aralıktaki entropi ölçüleri analiz edilir. Genel dağılıma göre farklılık gösteren entropi değerleri için ona karşılık gelen en küçük ve en büyük σ değerleri tespit edilir. En küçük σ değeri, sınıflandırma doğruluğunun artmaya başladığı noktayı işaret eder. En büyük σ değeri ise, en iyi sınıflandırma başarımının olduğu yeri işaret eder.

2. Aşama (*C* değerinin belirlenmesi): Önceki adımda bulunmuş olan en büyük σ değeriyle belirli sayıda (gerçekleştirilen uygulamada *10*) farklı ve eşit aralıklı *C* değeri için EKKDVM çalıştırılır. Bu *C* değerleri için sınıflandırma doğrulukları tutulur. *C* değerleri *x* eksenini ve doğruluk değerleri de *y* eksenini oluşturacak şekilde değer çiftleri iki boyutlu uzaya yerleştirilir. Daha sonra da, bu değerlere göre üssel tabanlı regresyon eğrisi uydurulup en iyi *C* değeri bu regresyon eğrisine göre belirlenir.

4.4 Geliştirilen Metotların Literatüre Katkıları

Bu bölümde verilen metotlar, literatürde sunulan DVM metotlarına alternatif olarak geliştirilmiş ve bu alanda üzerinde çalışılan bazı sorunların çözümüne olanak sağlamaktadır. Bu sorunlar, bölümün girişinde maddeler halinde sunulmuştur. Sözü geçen sorunlara getirmiş olduğu çözümler açısından tanıtılmış olan metotların literatüre etkisi aşağıda belirtilmiştir:

- Öncelikle bölüm 4.1 ve 4.2 de kullanılan “destek vektör olma ihtimali” fikrinin DVM alanına önemli avantajlar kazandıracağı söylenebilir. Çünkü standart DVM’de ve literatürde geliştirilen DVM metotlarının hemen hepsinde bu fikir kullanılmamaktadır. Eğitim kümesi verilerinin bir kısmı destek vektör olarak ayırıcı fonksiyonun kurulmasına tamamen katkıda bulunmaktadır, geri kalan kısmının ise bu fonksiyonun kurulmasına hiçbir katkısı yoktur. Fakat uygulamaların çoğunda, bu geri kalan grup içinde de destek vektör noktalarıyla oldukça benzer özellikler gösteren noktalar olabilmektedir. Çalışmada yapılan araştırmalara göre, bu verilerin de ayırıcı düzlem oluşturulurken incelenmesi gerçekleştirilen çalışmaların güvenilirliğini artıracaktır.

  Bu sebeplerden dolayı tez çalışmamızda olasılık tabanlı “destek vektör olma ihtimali” fikri geliştirildi. Standart DVM de, ayırıcı fonksiyonun oluşturulması hesap yükünü azaltmak için sadece destek vektör noktaları temel alınarak yapılır. Yapılan çalışmalarda, hesap yükünü azaltmak amacıyla, başlangıçta oranlanmış uzaklığı belirli bir değerden (eşik değeri) yüksek olan veriler, belirlenen sayıdan daha az komşuluk değerine sahip veriler ve ana destek vektör noktalarına belirli bir esneklikten (tolerans değeri) daha uzak olan veriler elenmiştir.

- Bölüm 4.1 ve 4.2’de sunulan metotların bir diğer katkısı, büyük veri kümeleri üzerinde standart DVM ve EKKDVM ne oranla daha kısa sürelerde eğitim yapabilmesi ve bunu yaparken de sınıflandırma başarımından da ödün vermemesidir. Bölüm 4.1 ve 4.2’de anlatılan metotlar literatürde benzeri olmayan çalışmalardır. İlk kez bu kümeleme metotlarından, DVM eğitim algoritması geliştirilmesi amacıyla faydalanılmıştır. Bölüm 4.1’de tanıtılan metot 4.2 deki metoda göre, uygulanabilirlik açısından daha kolay ve eğitim süresi daha kısadır. Fakat bölüm 4.2’deki metot ise, daha güvenilir ve kesin sonuçlar veren, öncelik bilgisi istemeyen ve hiyerarşik yapıda çalışan bir metottur. Kümeleme tabanlı DVM eğitim algoritmaları, literatürde henüz yeni çalışılmaya başlanan bir konudur. Yakın gelecekte gelişmelerin sürmesi beklenmektedir.

- Yine bölüm 4.1 ve 4.2'deki metotlar çok sınıflı sınıflandırma problemlerine uygulanmıştır. Destek vektör noktalarının olasılık tabanlı olarak belirlenmesinden bu tip problemlerde de faydalanılmıştır ve etkili olduğu görülmüştür. Çok sınıflı sınıflandırmada karşılaşılan, sınıflandırılamayan bölgeler problemi destek vektör olma ihtimallerini de kullanan ağırlıklı ortamalama yaklaşımı kullanılarak çözülmüştür. Bu yaklaşımın problem çözümünde oldukça etkili olduğu uygulama sonuçlarıyla da görülmüştür.

- Bölüm 4.3 te geliştirilen metot ise, DVM öğrenme parametrelerinin optimum değerlerinin bulunması ile ilgilidir. Sunulan metot araştırmalarımıza göre, bu tip bir problemin çözümünde entropi tabanlı kümeleme metodunu kullanan literatürdeki ilk çalışmadır. Aynı zamanda kalp kapakçığı verilerinin incelenmesi konusunda gerçekleştirilmiş ilk çalışmadır. Metodun eğitim zamanı makuldür ve tatmin edici uygulama sonuçları vermiştir. Yaklaşımımızın deneme/yanılma yoluyla defalarca deneyerek bulmuş olduğumuz parametre değerlerine, başarılı bir şekilde yakınsamakta olduğu görülmüştür.

Geliştirilen bu metotları kullanarak sınıflandırılan veri kümeleri ve elde edilen uygulama sonuçları 5. bölümde sunulmuştur.

5. ARAŞTIRMA VE UYGULAMA SONUÇLARI

Bu kısımda kullanılan veri kümelerinden ve uygulama sonuçlarından bahsedilmektedir. Bölüm 5.1 de kullanılan veri kümeleri, 5.2 de ise geliştirilen metotlarla bu veri kümeleri üzerinde elde edilen sonuçlar ele alınmıştır.

5.1 Kullanılan Veri Kümeleri

Bu bölümde, bölüm 4'te anlatılan metotların başarımını test etmek amacıyla üzerinde çalışılan veri kümelerinden bahsedilmektedir.

- **İris Veri Kümesi:** Literatürde en sık kullanılan veri kümelerinden bir tanesidir. 150 veriden ve üç farklı sınıftan oluşmaktadır. Her bir sınıf için eğitim ve test kümesi olarak ayrılacak 50 tane veri mevcuttur. Toplamda, 75 veri eğitim için 75 veri de test için kullanılmaktadır. Bu veri kümesi UCI (California Üniversitesi Bilimsel Veri Tabanı) veritabanından elde edilmiştir. Sınıflar, üç farklı iris çiçeği türünü (Setosa, Versicolor, Virginica) temsil etmektedir. Her veri 4 nitelikten oluşmaktadır. Bu nitelikler de, iris çiçeklerinin çanak yaprağı uzunluğu, çanak yaprağı genişliği, taç yaprağı uzunluğu ve taç yaprağı genişliğidir.

- **Şarap Veri Kümesi:** Bu veri kümesi 13 nitelikli, 3 farklı sınıfa ait 178 veriden oluşmaktadır. Sınıf başına eşit olmamakla birlikte verilerin 1., 2. ve 3. sınıflara göre dağılımı sırasıyla, 59, 71 ve 48 şeklindedir. 178 veriden 90 tanesi eğitme kümesi olarak kalan 88 tanesi ise test verisi olarak kullanılmaktadır. Sınıflar, İtalyanın farklı bölgelerinden elde edilen 3 çeşit şarabı simgelemektedir. Kimyasal analizler sonucu elde edilen nitelikler; alkol oranı, malik asit oranı, kül, kül alkalini, magnezyum, toplam phenols (zehirli bir asit), flavanoids (organik kimyasal bir bileşen), nonflavanoids (organik olmayan kimyasal bir bileşen), proanthocyanin (meyvelere renk veren bir kimyasal), renk yoğunluğu, renk, seyreltilmiş şarabın OD280/OD315 değeri ve amino asit oranıdır. Bu veri kümesi de UCI veritabanından elde edilmiştir.

- **Tiroid Veri Kümesi:** Tiroid verisi, özellikle çoklu sınıf sınıflandırmalı DVM problemlerinde sıkça kullanılan bir veri kümesidir. Bu veri kümesi 21 nitelikli, 3 sınıflı 7200 veriden oluşur. Verilerin %92'den fazlası birinci sınıfa aittir. Bu verilerden 3772 tanesi eğitme kümesi olarak, kalan 3428 tanesi de test kümesi olarak kullanılmaktadır. Tiroid veri kümesi de diğer iki veri kümesi gibi UCI veritabanından elde edilmiştir.

- **Araç Veri Kümesi:** Önceki üç veri kümesi kadar olmasa da literatürde çok çalışılan bir veri kümesidir ve bu veri kümesi de UCI veritabanından alınmıştır. Küme dört çeşit araçtan elde edilen görüntüler aracılığıyla çıkarılmış niteliklerden oluşur. Nitelikler, farklı açılardan görüntülerin alınmasıyla çeşitlendirilmiştir. Araç çeşitlerinden bir tanesi iki katlı otobüs, bir tanesi kamyonet ve diğer ikisi de araba grubundandır. Veriler 18 nitelikten oluşmaktadır. Veri kümesi; 1. sınıfta 212, 2. sınıfta 217, 3. sınıfta 218 ve 4. sınıfta da 199 veri olmak üzere toplam 846 veriden oluşmaktadır. Bu verilerin 424 tanesi eğitim ve 422 tanesi de test kümesini oluşturmaktadır. Aynı veri tabanından elde edilen bu dört kümenin temel özellikleri Tablo 5.1'de özetlenmiştir.

Tablo 5.1 Kullanılan UCI Veri Kümelerinin Özellikleri

Veri	Nitelikler	Çıkışlar	Eğitim Verisi	Test Verisi
Tiroid	21	3	3772	3428
Şarap	13	3	90	88
Iris	4	3	75	75
Araç	18	4	424	422

- **Doppler Kalp Verileri Kümesi:** Tez çalışmasında kullanılan diğer veri kümesi Doppler kalp verileridir. Bu veri kümesi yukarıda bahsedilen kümelerden farklı olarak UCI veritabanından değil Fırat Tıp Merkezi kardiyoloji kliniğinde muayene edilen hastalıklı ve sağlıklı kişilerden elde edilen Doppler kalp kapakçığı verilerinden oluşmaktadır. Bu veri kümesi, Aort ve Mitral kapakçığı verilerinden oluşmaktadır.

  Doppler, ilk kez 1842 yılında Christian Andreas Doppler tarafından matematiksel bir hipotez olarak önerilmiştir. Mantığı kısaca, dalga özelliği gösteren herhangi bir fiziksel varlığın frekans ve dalga boyunun hareketli bir gözlemci tarafından farklı zaman ve/veya konumlarda farklı şekillerde algılanmasına dayanmaktadır. Dalgalar için Doppler etkisi hesaplanırken, dalga kaynağı ile gözlemcinin birbirlerine göre konumları ve dalganın hareket ettiği ortamın yapısal özellikleri göz önüne alınmalıdır.

  Konunun daha iyi anlaşılması için bir örnek verecek olunursa ambulans örneği uygun olacaktır. Bu örnekte, araç bulunulan referans noktasına yaklaşırken sirenden çıkan ses dalgalarının yüksekliği ve frekansı artacak, uzaklaşırken de tam tersi olacaktır. Referans noktasında, bu etki ile ambulansın yaklaştığını ya da uzaklaştığını sadece sirenin sesini dinleyerek anlayabilirsiniz. Eğer sesteki bu değişimlerin hızını ölçebilirseniz ambulansın hızını da tahmin edebilirsiniz. Doppler ultrasonografinin çalışma mantığı da bu örneğe oldukça benzemektedir. Vücut içerisinde herhangi hareketli bir dokuya gönderilen ses dalgalarının yansıma hızları arasındaki fark o cismin proba doğru mu yoksa proba ters yönde mi hareket ettiğinin anlaşılmasını sağlar.

  Doppler incelemelerinde hedef alınan hareketli cisimler kırmızı kan hücreleri yani alyuvarlardır. Gönderilen dalganın kandaki alyuvarlardan yansıtılan kısmı algılanarak bu sinyalin frekansında ne kadar değişiklik olduğu eşitlik 5.1 yardımıyla tespit edilir.

$$f_d = f_t - f_r = \frac{2 f_t V Cos\theta}{c} \tag{5.1}$$

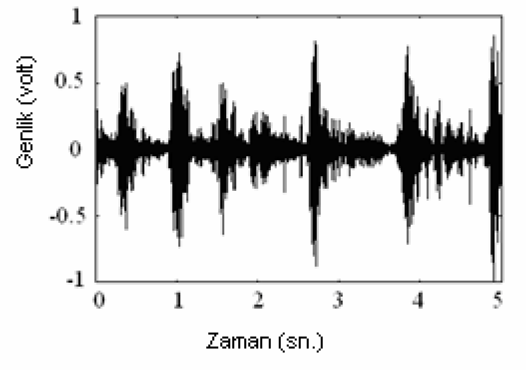
Burada f_d yansıyan ultrasonik dalganın frekansındaki kayma miktarını, f_t gönderilen dalganın frekansını, f_r yansıyan dalganın frekansını, v hareketli cismin hızını, c yansıyan ultrasonik dalganın ortamdaki hızını ve $Cos\theta$ ultrasonik dalganın yönü ile hareket yönü arasındaki açının trigonometrik değerini temsil eder.

Fırat Tıp Merkezinden elde edilen Doppler veri kümesi, (Türkoğlu, 2002) tez çalışmasında kullanılan Doppler Kalp Seslerine (DKS) aittir. Bu veri kümesine ait DKS işaretleri, Fırat Tıp Merkezi Kardiyoloji biriminden Acuson Sequoila 512 Model Doppler Ultrason cihazı aracılığı ile alınmıştır. Bu işaretler 5 saniyelik olarak SNR=0 dB kazançlı ve 20 KHz'lik örnekleme frekansı ile 16 bitlik A/D çözünürlüğe sahip olan bir ses kartı üzerinden doktor gözetiminde bilgisayar ortamına aktarılmıştır. Kullanılan Doppler ultrason dönüştürücünün çalışma modu, 2 MHz'lik sürekli dalga modudur.

Çalışılan veri kümesinde 5 sınıflı veri grubu bulunmaktadır. Bu gruplardan bir tanesi normal (herhangi bir hastalık içermeyen) verilerden oluşur. Diğer gruplardan her biri ise, 4 çeşit hastalıktan birini temsil etmektedir. Bunlar; aortik darlık (aortic stenosis), aortik yetmezlik (aortic insufficiency), mitral darlık (mitral stenosis) ve mitral yetmezlik (mitral insufficiency) şeklinde adlandırılan 4 tip kalp kapakçığı düzensizliğidir. Tez çalışması kapsamında gerçekleştirilen uygulamalarda 4 tip kalp kapakçığı düzensizliği verileri ayrı ayrı ele alınmamış, Türkoğlu ve ark. (2003) ve Uğuz ve ark. (2007)'nın çalışmalarındaki gibi sadece normal ve hastalıklı şeklinde sınıflandırılmıştır. Bu yüzden DVM sınıflandırıcı metodumuz iki sınıflı sınıflandırma yapmaktadır.

Diğer bir ifadeyle normal Doppler verileri, aortik ve mitral kalp kapakçıklarında yetersizlik ve darlık olmadığını gösterirken, hastalıklı veriler bu hastalıklardan herhangi birinin varolduğunu kendi içerisinde ayırt etmeden temsil eder.

Şekil 5.1'de hastalıklı bireyden elde edilen örnek bir Doppler kalp sesi işareti görülmektedir.



Şekil 5.1 DKS İşareti Örneği

Veri kümesi toplam 215 veriden oluşmaktadır. Bunlardan 110 tanesi aortik kalp kapakçığından alınmış bulgularla oluşturulmuştur ve doktorun görüşlerine göre 56'sı normal 54'ü hastalıklı (toplam aortik darlık ve aortik yetmezlik sayısı) olarak belirlenmiştir. Geriye kalan 105 tane veri ise, mitral kalp kapakçığı ile ilgilidir. Yine doktor görüşleri temel alınarak bunların da 39'unun normal ve 66'sının hastalıklı olduklarına karar verilmiştir.

Yukarıdaki hasta ve normal veri dağılımlarına göre, eğiticili sınıflandırma işlemi için veri kümesi üzerinde şu şekilde bir bölümleme işlemi gerçekleştirilmiştir. Aortik kalp kapakçığından 25 normal, 14 hastalıklı veri ve mitral kalp kapakçığından da 20 normal, 33 hastalıklı veri eğitim için kalan veriler ise test için ayrılmıştır (Türkoğlu ve ark., 2003). Kullanılan veri kümesinin bazı özellikleri de Tablo 5.2'de özetlenmiştir.

Tablo 5.2 DKS işaretlerinin bazı özellikleri

	Yaş		Cinsiyet		Adet	
	Değişim	Ortalama	Erkek	Bayan	Normal	Hasta
Mitral	19 - 78	47.5	58	47	39	66
Aort	15 - 80	50	74	36	56	54

5.2 Uygulama Sonuçları

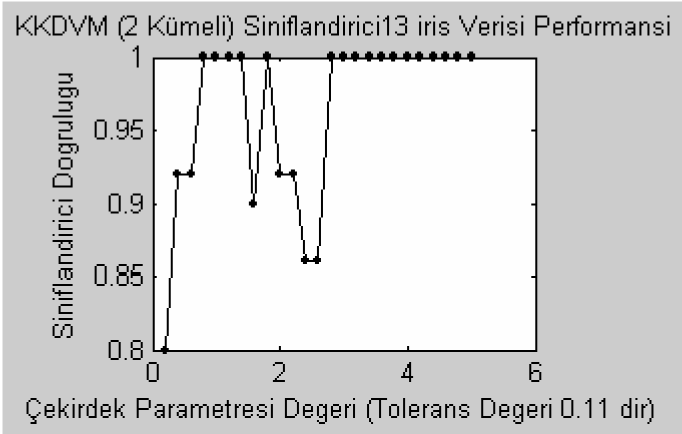
Bu bölümde verilen uygulama sonuçları Matlab 6.5 programlama dili kullanılarak elde edilmiştir. Çalışılan bilgisayar; intel pentium M 1.70 GHz işlemci, 600 MHz ve 992 MB RAM, 2 MB ön bellek donanımı ve Microsoft Windows XP Profesyonel işletim sistemi bileşenlerine sahiptir.

5.2.1 Geliştirilen KKDVM Metodu Kullanılarak Bulunan Uygulama Sonuçları

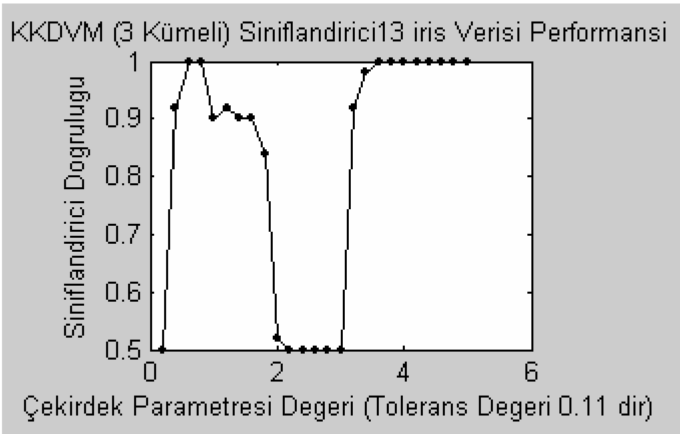
Bölüm 4.1'de detayları verilen KKDVM metodu üç farklı veri kümesi (iris, şarap ve tiroid) üzerinde test edilmiştir. Bölüm 5.1'de verilen bu veri kümelerinin sınıflandırılması işleminde, en iyi sınıflandırma başarımına ulaşılabilmesi için bunu mümkün kılan KKDVM öğrenme parametrelerinin ve sabitlerinin uygun değerlerinin bulunması gerekmektedir. Gerçekleştirilen testler sonucunda, bu parametrelerin bahsedilen değerleri Tablo 5.3'te gösterildiği gibi bulunmuştur. Tablo 5.3'te yeralan çekirdek parametresi YTF'a ait olan σ parametresidir. Tolerans değerleri, ana destek vektör ile yardımcı destek vektörler arasında oranlanmış mesafeye dayalı olarak belirlenen esneklik değeridir. Bu sayının artması, muhtemel destek vektör sayısının da artmasına sebep olacaktır. Tablonun diğer sütununda tümleşik olarak verilen değerler ise veri gruplarının ayrıldığı küme sayısını ve bu kümelerin eşik değerlerini göstermektedir. Eşik değerleri de oranlanmış uzaklıklar cinsinden alınmaktadır ve değeri aşan veriler eğitim kümesinden çıkarılıp destek vektör olma ihtimali sıfırlanmaktadır.

Tablo 5.3 Sınıflandırıcının Optimum Parametre Değerleri

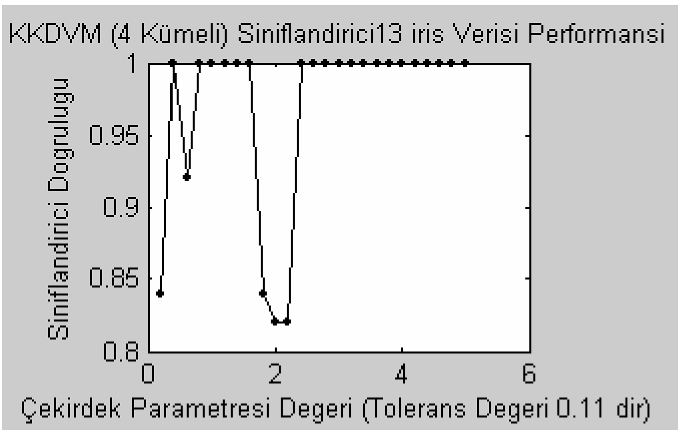
Veri	Çekirdek Par.	Tolerans	Kümelerin Eşik Değerleri ve Sayısı				
Tiroid	2.5	0.09	1.1	1.1	1.6	1.9	4 küme
Şarap	2.9	0.11	1.1	1.1	1.1	...	3 küme
İris	1.6	0.09	1.5	1.5	1.5	…	3 küme



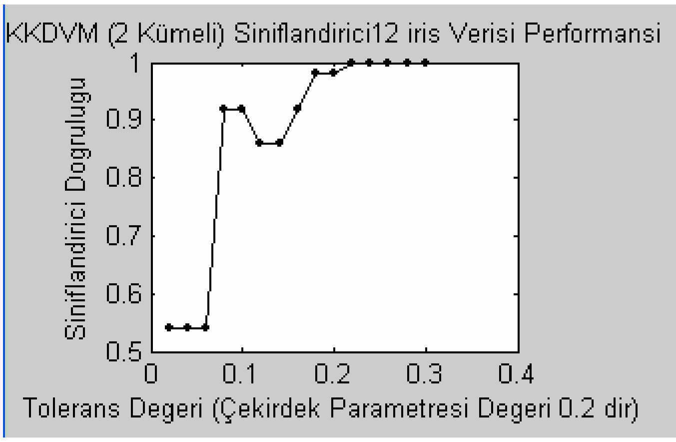
Şekil 5.2 İris Veri Kümesi İçin Çekirdek Parametresine Göre Sınıflandırma Doğruluğu Değişimi (2 Küme)



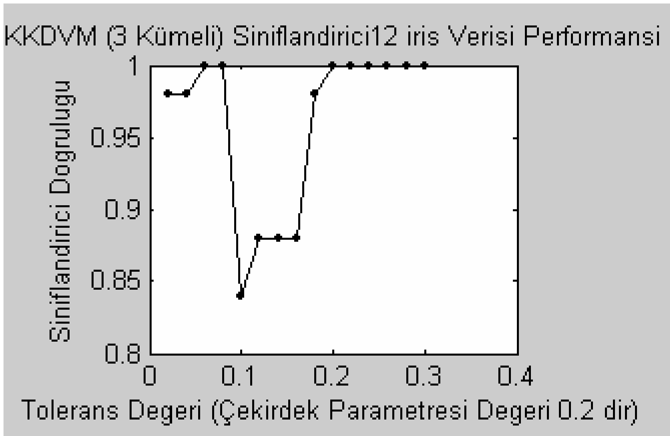
Şekil 5.3 İris Veri Kümesi İçin Çekirdek Parametresine Göre Sınıflandırma Doğruluğu Değişimi (3 Küme)



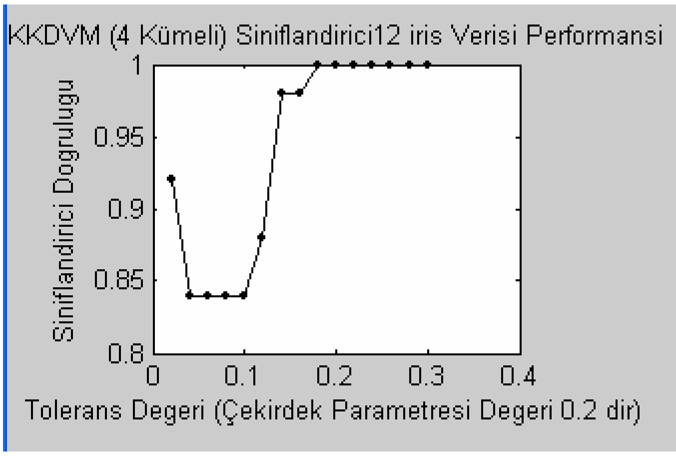
Şekil 5.4 İris Veri Kümesi İçin Çekirdek Parametresine Göre Sınıflandırma Doğruluğu Değişimi (4 Küme)



Şekil 5.5 İris Veri Kümesi İçin Tolerans Değerine Göre Sınıflandırma Doğruluğu Değişimi (2 Küme)



Şekil 5.6 İris Veri Kümesi İçin Tolerans Değerine Göre Sınıflandırma Doğruluğu Değişimi (3 Küme)



Şekil 5.7 İris Veri Kümesi İçin Tolerans Değerine Göre Sınıflandırma Doğruluğu Değişimi (4 Küme)

KKDVM nin sınıflandırma başarımı; çekirdek parametresi, tolerans ve küme sayısı değerlerine göre değişmektedir. Bu değişim şekil 5.2 – şekil 5.7'de gösterilmektedir. Şekillerin her birinde bir parametre değeri değişmekte, diğerleri ise belirli bir değerde sabit tutulmaktadır. Şekillerden de görüldüğü gibi tolerans değeri sabitken küme sayısının değiştirilmesi aynı çekirdek parametresi değerlerinde bile farklı sınıflandırma başarımlarına yol açabilmektedir. Aynı durum çekirdek parametresi değeri sabitken de geçerlidir. Şekillerin bir yarısı iris veri kümesi üzerinde sınıflandırıcı 13'ün ve diğer yarısı da sınıflandırıcı 12'nin performansını doğrudan göstermektedir. Çünkü veri kümesi çok sınıflı yapıdadır ve çok sınıflı sınıflandırma yaklaşımlarıyla işlenmelidir. Tez çalışmasında bire karşı bir yaklaşımı kullanılmıştır. Bu yüzden, bu veri kümesi için 3 adet sınıflandırıcıya gereksinim duyulmuştur. Bunlardan bir tanesi de sınıflandırıcı 13 tür.

Küme sayısının en iyi değerinin belirlenmesi de, yüksek bir sınıflandırma doğruluğuna ulaşılması açısından önemlidir. KKDVM'de küme sayısı çeşitli denemeler sonucu en iyi başarımın aranması sayesinde bulunmuştur. Bu denemeler çok uzun zaman almaktadır ve kısa değer aralıklarıyla yapılan denemeler ise en iyi sonucun atlanmasına sebep olabilmektedir. Bir sonraki bölümde uygulama sonuçlarına da değinilecek olan RKKDVM'nin ortaya çıkış sebebi de budur. Çünkü RKKDVM'de en iyi küme sayısı Renyi entropi değerleri ile otomatik olarak bulunabilmektedir. Ayrıca bu metotta, eşik değerlerinin belirlenmesine de gerek kalmamıştır.

Tablo 5.3'te belirlenen parametre değerlerine göre KKDVM eğitildiğinde alınan sınıflandırma başarımı ve aynı veri kümesi üzerinde çalıştırılan DVM tabanlı farklı sınıflandırıcıların başarımları karşılaştırmalı olarak Tablo 5.4'te gösterilmiştir. Tablonun ilk sütununda yer alan metot standart En Küçük Kareler DVM sınıflandırıcısıdır ve tez çalışmasında geliştirilmiş olan metotlarla aynı amaç için geliştirilmiş olan bir DVM çeşididir. İkinci ve üçüncü sütunlarda, sırasıyla minimum ve ortalama operatörlerini kullanan Tsujinishi ve Abe (2003) tarafından geliştirilen Bulanık En Küçük Kareler DVM sınıflandırıcılarının sonuçları yer almaktadır. Dördüncü sütun daha önceki bir çalışmada tasarlanmış olan bir metottur ve çok sınıflı sınıflandırmada oluşan sınıflandırılamayan bölgeler problemi bu çalışmada KEYK metodu yardımıyla çözülmüştür (Çomak ve Arslan, 2006). Son sütunda ise, geliştirilmiş olan KKDVM metodunun başarımı gösterilmektedir (Çomak ve Arslan, 2007). KKDVM tüm veri kümeleri için en yüksek sınıflandırma doğruluğunu vermiştir. Özellikle tiroid verisi üzerinde görülen sınıflandırma başarımı sınıflandırıcının büyük veri kümeleri üzerinde başarılı olduğunu göstermektedir.

Tablo 5.4 Sınıflandırıcı Performanslarının Karşılaştırılması

Veri	Standart EKKDVM	BEKKDVM (min) [Tsujinishi ve Abe, 2003]	BEKKDVM (ort) [Tsujinishi ve Abe, 2003]	EKKDVM + KEYK	KKDVM
Tiroid	%95.62	%95.94	%94.75	%95.74	%96.62
İris	%94.66	%96.00	%96.00	%97.72	%98.67
Şarap	%96.59	---	---	%96.00	%97.73

Bu sınıflandırma doğrulukları gözlenirken hesaplanan eğitim süreleri ise; iris, şarap ve tiroid veri kümeleri için sırasıyla Standart EKKDVM de 0.921, 1.03, 5638.26 saniye ve KKDVM de 0.155, 0.282, 1677.591 saniye olarak bulunmuştur. Bu eğitim süreleri geliştirilen metodun etkinliğinin de önemli bir göstergesidir.

Bu sınıflandırma doğrulukları ve eğitim süreleri göz önüne alındığında, geliştirilen KKDVM yönteminin EKKDVM yöntemine göre çok daha kısa sürede eğitilebildiği ve bununla da sınırlı kalmayıp sınıflandırma başarımını da artırmıştır. Çünkü EKKDVM'de orijinal DVM'ne oranla basit de olsa, optimizasyon problem sistemi kurulmaktadır. Bu sistem büyük boyutlu matrisler üzerinde matris işlemlerinin gerçekleştirilmesine ihtiyaç duymaktadır. Fakat geliştirilen KKDVM yönteminde bahsedilen optimizasyon sisteminin kurulmasına gerek yoktur. Çünkü KKDVM'de kümeleme işlemi basit bir yaklaşımla gerçekleştirilmekte ve veri noktaları üzerinde yerel kontrol özelliğinin etkili bir biçimde uygulanabilmesine olanak sağlanmaktadır. Böylece mevcut veri noktalarının ayırıcı düzlemi temsil yeteneği güvenli bir şekilde değerlendirilebilmektedir. Ayrıca EKKDVM'de eğitim süresi orijinal DVM'ne göre azaltılırken sınıflandırma doğruluğunda da küçük bir azalma görülmüştür. Fakat KKDVM'de sınıflandırma doğruluğu da artırılmıştır.

Tablo 5.4, DVM tabanlı sınıflandırma başarımlarını göstermektedir. Bölüm 5.2.2'de de aynı veri kümeleri kullanıldığı için diğer sınıflandırıcı sonuçları sonraki bölümde verilmektedir.

5.2.2 Geliştirilen RKKDVM Metodu Kullanılarak Bulunan Uygulama Sonuçları

Bölüm 4.2'de detayları verilen RKKDVM metodu ise üçü KKDVM'de de kullanılan dört farklı veri kümesi (iris, şarap, tiroid ve araç) üzerinde test edilmiştir. RKKDVM'de, KKDVM den farklı olarak Renyi entropi temelli bir kümeleme işlemi gerçekleştirilmiştir. Bu yüzden de, RKKDVM'de küme sayısını deneme/yanılma yoluyla değil otomatik olarak bulunmuştur.

Tablo 5.5 Kümele İşleminde Bulunan Optimum Değerler (İris ve Şarap)

Parametre	İris 1	İris 2	İris 3	Şarap 1	Şarap 2	Şarap 3
K_{init}	6	6	5	6	6	6
N_{init}	2	3	3	3	2	5
Alfa	2.8	2.8	2.9	3.1	3.0	3.2

Tablo 5.6 Kümele İşleminde Bulunan Optimum Değerler (Tiroid ve Araç)

Parametre	Tiroid 1	Tiroid 2	Tiroid 3	Araç 1	Araç 2	Araç 3	Araç 4
K_{init}	7	6	7	7	4	7	4
N_{init}	6	7	5	3	2	3	2
Alfa	2.3	2.6	2.0	0.6	0.9	1.2	1.2

Tablo 5.5 ve 5.6'da sınıflandırıcının kümeleme aşamasında en iyi entropi ölçülerine göre bulunan parametre değerleri sunulmuştur. Tablo 5.3'te belirtilen KKDVM metodunun parametrelerinden farklı olarak RKKDVM de küme eşik değerlerinin belirlenmesine gerek duyulmaz. Çünkü KKDVM'de, önce küme eşik değerleri belirlenip bu değeri aşan veriler veri kümesinden elenmektedir. Daha sonra, kalan veriler üzerinde öklit uzaklığına dayalı bir kümeleme işlemi gerçekleştirilmektedir. RKKDVM'de ise, kümeleme ile ilgili tüm işlemler otomatik bir şekilde Renyi entropi temelli olarak yapıldığından eşik değerlerinin belirlenmesine gerek duyulmamaktadır.

RKKDVM'de eğiticili öğrenmeden yararlanmak amacıyla, öncelikle küme sayıları her bir etiket sınıfı için bağımsız olacak şekilde hesaplanmaktadır. Sonra, tüm sınıflar için en iyi sonucu veren ortak küme sayısı değeri bulunur. Böylece, Tablo 5.7'de gösterilen optimum parametre değerleri ve küme sayıları ortaya çıkarılmıştır.

Tablo 5.7 Optimum Sınıflandırıcı Parametreleri

Veri	YTF Parametresi	Tolerans	Küme Sayısı
Tiroid	[0.001-1.9]	[0.01-0.1]	5
Şarap	[0.4-1.1]	[0.07-0.24]	4
Araç	[5.1-9.6]	[0.01-0.1]	5
İris	0.25	0.09	4

Tablo 5.7'deki değerlere göre oluşturulan sistemlerden Tablo 5.8'de gösterilen sınıflandırma başarımları elde edilmiştir. RKKDVM metodunun özellikle büyük bir veri kümesi olan tiroid verilerinde sınıflandırma başarımını önemli ölçüde artırdığı görülmüştür. İris veri kümesinde de sınıflandırma başarımını artırırken, şarap veri kümesinde KKDVM ne göre bir miktar düşüş gözlemlenmiştir. Bu sonuç, Renyi entropi temelli kümelemenin özellikle büyük veri kümeleri üzerinde önemli katkıları olabileceğini göstermektedir.

Tablo 5.8 RKKDVM ve Diğer DVM Tabanlı Sınıflandırıcıların Performansları

Veri	Standart EKKDVM	BEKKDVM (min) [Tsujinishi ve Abe, 2003]	BEKKDVM (ort) [Tsujinishi ve Abe, 2003]	EKKDVM + KEYK	RKKDVM
Tiroid	%95.62	%95.94	%94.75	%95.74	%97.26
İris	%94.66	%96.00	%96.00	%97.72	%100.00
Araç	%68.72 [S. Suresh ve ark., 2008]	---	---	---	%83.88
Şarap	%96.59	---	---	%96.00	%96.59

Karşılaştırılan diğer sınıflandırıcıların başarım ölçüleri D. Tsujinishi ve S. Abe (2003) ve S. Suresh ve ark. (2008)'nın çalışmalarından alınmıştır. Tarafımızdan geliştirilen sınıflandırıcıların başarımları ise, son iki sütunda gösterilmektedir. RKKDVM'nin başarımı diğer DVM tabanlı sınıflandırıcıların hiçbirisinden düşük bulunmamıştır. Eğitim süreleri açısından, RKKDVM'nin KKDVM'den daha yüksek değerlere sahip olduğu görülmektedir. Bu fark da, Renyi entropi hesaplamalarından kaynaklanmaktadır. Fakat eğitim zamanı dezavantajına rağmen RKKDVM KKDVM'nde bulunmayan önemli avantajlara sahiptir. KKDVM'de her bir sınıfın bölüneceği küme sayısı deneme yanılma yöntemi ile bulunurken RKKDVM'de entropi değerlerine göre otomatik olarak bulunabilmektedir. Bu noktada, Renyi entropinin üstün özelliklerinden faydalanılmıştır. Çünkü Renyi entropi, hem diğer entropi yaklaşımlarında bulunan belirsizlikleri temsil edebilme yeteneğine sahiptir hem de kümeleme problemlerine kolaylıkla uyarlanabilen bir yapıdadır. Renyi entropi ölçümlerinde değişimin ön yüksek olduğu geçiş noktası, çalışılan veri için en uygun yapıyı her zaman temsil etmektedir ve bu temsili her veri kümesinde başarıyla gerçekleştirilebilmektedir. Ayrıca RKKDVM'de KKDVM'de en uygun değerinin belirlenmesi gereken eşik değişkeninin kullanılmasına ihtiyaç duyulmamaktadır.

Renyi entropi yönteminin kazandırdığı avantajlar, uygulamalarda özellikle daha büyük olan Tiroid ve Araç veri kümelerinin sınıflandırma başarımlarında açıkça görülmüştür. Aşağıdaki Tablolarda, diğer sınıflandırıcıların aynı UCI veri kümeleri üzerindeki sınıflandırma başarımları KKDVM ve RKKDVM ile birlikte yer almaktadır (Draghici, 2001*, J. M. Martinez-Otzeta ve ark.**, 2006).

Tablo 5.9 Şarap Veri Kümesi Üzerinde Diğer Sınıflandırıcıların Performansları

Veri Kümesi	ID3**	IB1**	Bayes**	C 4.5**	Genetik Alg.**	KKDVM	RKKDVM
Şarap	%92.70	%95.00	%97.19	%92.75	%97.22	%97.73	%96.59

Tablo 5.10 Araç Veri Kümesi Üzerinde Diğer Sınıflandırıcıların Performansları

Veri Kümesi	ID3**	IB1**	Bayes**	C 4.5**	Genetik Alg.**	KKDVM	RKKDVM
Araç	%72.92	%71.04	%45.41	%72.22	%74.57	------	%83.88

Tablo 5.11 İris Veri Kümesi Üzerinde Diğer Sınıflandırıcıların Performansları

Veri Kümesi	CN2*	ELEM2*	CLEF*	C 4.5*	CBD*	KKDVM	RKKDVM
İris	%88.68	%94.67	%94.4	%95.3	%95.66	%98.67	%100.00

Aşağıda RKKDVM'nin iris veri kümesi üzerinde çalışmasını açıklayan sayısal bir örnek verilmektedir.

Örnek: İris İçin RKKDVM, Bu örnekte RKKDVMnin iris veri kümesinde çalışması örneklendirilmiştir. UCI veri tabanından alınan İris veri kümesinin bir kısmı (her sınıftan 10 eğitim verisi) şu şekilde gösterilebilir:

Tablo 5.12 Bir Kısım İris Veri Kümesi Verilerinin Sayısal Değerleri

Veri sırası/ Nitelikler	**Çanak Yaprağı Uzunluğu**	**Çanak Yaprağı Genişliği**	**Taç Yaprağı Uzunluğu**	**Taç Yaprağı Genişliği**	**Çıkış(Sınıf) Etiketi**
1	5.1	3.5	1.4	0.2	1
2	4.9	3	1.4	0.2	1
3	4.7	3.2	1.3	0.2	1
4	4.6	3.1	1.5	0.2	1
5	5	3.6	1.4	0.2	1
6	5.4	3.9	1.7	0.4	1
7	4.6	3.4	1.4	0.3	1
8	5	3.4	1.5	0.2	1
9	4.4	2.9	1.4	0.2	1
10	4.9	3.1	1.5	0.1	1
11	7	3.2	4.7	1.4	2
12	6.4	3.2	4.5	1.5	2
13	6.9	3.1	4.9	1.5	2
14	5.5	2.3	4	1.3	2
15	6.5	2.8	4.6	1.5	2
16	5.7	2.8	4.5	1.3	2
17	6.3	3.3	4.7	1.6	2
18	4.9	2.4	3.3	1	2
19	6.6	2.9	4.6	1.3	2
20	5.2	2.7	3.9	1.4	2
21	6.3	3.3	6	2.5	3
22	5.8	2.7	5.1	1.9	3
23	7.1	3	5.9	2.1	3
24	6.3	2.9	5.6	1.8	3
25	6.5	3	5.8	2.2	3
26	7.6	3	6.6	2.1	3
27	4.9	2.5	4.5	1.7	3
28	7.3	2.9	6.3	1.8	3
29	6.7	2.5	5.8	1.8	3
30	7.2	3.6	6.1	2.5	3

Tablo 5.12'de 3 çeşit sınıftan da bir takım veriler alınmıştır. Örnek ile ilgili bundan sonraki kısımlarda ise 1 ve 2 numaralı sınıf verileri temel alınmıştır. Çünkü RKKDVM de, standart DVM gibi iki sınıflı sınıflandırma işlemlerini birleştirerek daha çok sınıflı sınıflandırmaları gerçekleştirir. Tablo 5.13, 1 ve 2 numaralı sınıfların ilk 10 eğitim verisinin YTF çekirdek fonksiyonuyla işlenip nitelik uzayına aktarılan değerlerini göstermektedir. Ayrıca Tablo 5.12'den farklı olarak çıkış/etiket sütunu da artık tutulmamaktadır.

Tablo 5.13 Bir Kısım İris Veri Kümesi Verilerinin Nitelik Uzayındaki Değerleri

Veri sırası/ Nitelikler	**Çanak Yaprağı Uzunluğu**	**Çanak Yaprağı Genişliği**	**Taç Yaprağı Uzunluğu**	**Taç Yaprağı Genişliği**
1	0.95938	0.99681	0.97161	0.98174
2	0.87715	0.15831	0.97161	0.98174
3	0.42288	0.53409	0.81481	0.98174
4	0.23097	0.31499	0.98728	0.98174
5	0.99375	0.89119	0.97161	0.98174
6	0.33053	0.24385	0.63078	0.83124
7	0.23097	0.95009	0.97161	0.9786
8	0.99375	0.95009	0.98728	0.98174
9	0.042635	0.0678	0.97161	0.98174
10	0.87715	0.31499	0.98728	0.83926
11	3.083×10^{-14}	0.53409	3.3698×10^{-37}	2.45×10^{-5}
12	2.883×10^{-7}	0.53409	7.7874×10^{-33}	3.5804×10^{-6}
13	6.676×10^{-13}	0.31499	7.6889×10^{-42}	3.5804×10^{-6}
14	0.16826	1.453×10^{-5}	3.8446×10^{-23}	0.00014286
15	2.963×10^{-8}	0.024743	5.5493×10^{-35}	3.5804×10^{-6}
16	0.02698	0.024743	7.7874×10^{-33}	0.00014286
17	2.390×10^{-6}	0.77167	3.3698×10^{-37}	4.4587×10^{-7}
18	0.87715	8.861×10^{-5}	1.7267×10^{-12}	0.010845
19	2.595×10^{-9}	0.0678	5.5493×10^{-35}	0.00014286
20	0.78925	0.00769	2.0657×10^{-21}	2.45×10^{-5}

Tablo 5.13'deki şekle getirilen verilere Renyi entropi metodu uygulanarak kümeleme işlemi gerçekleştirilmiştir. Uygulamamızda en iyi entropi değerlerine göre

N_{init}=3 ve K_{init}=6 durumunda bulunmuş olan 4 kümeli yapı seçilmiştir. Bu yapıya göre, örneğin 1 etiketli sınıfın 23, 3 ve 6 numaralı verilerinin 1. sınıfın 1. kümesini oluşturduğu gözlenmiştir. Aynı şekilde 1. sınıfın 2., 3. ve 4. kümeleri ile 2. sınıfın 1., 2., 3. ve 4. kümelerini oluşturan veriler de bulunmuştur. Daha sonra her bir kümenin merkez noktası (oluşan 8 küme için bağımsız olarak) hesaplanmıştır.

Sınıfların karşılıklı küme merkezleri arası mesafe 1 olacak şekilde her verinin 1. sınıfa ait küme merkezine ve 2. sınıfa ait küme merkezine olan normalleştirilmiş uzaklıkları bulunur. Bu iki değer veriye (n+1). ve (n+2). (örnekte 5. ve 6.) nitelik olarak eklenir.

Sonraki adımda, her bir verinin *k* en yakın komşusu arasından kendisiyle aynı sınıfa ait olanlarının sayısı bulunur ve veriye (n+3). boyut olarak (örnekte 7) eklenir. Başlangıçta 4 niteliğe sahip olan veriler Tablo 5.14'deki gibi 7 nitelikli hale gelir.

Tablo 5.14 2. Sınıf 3. Küme İris Veri Kümesi Verilerinin Normalleştirilmiş Uzaklık ve Komşuluk Sayısı Eklenmiş Formu

2. Sınıf 3.Küme	**1- Çanak Yaprağı Uzunluğu**	**2- Çanak Yaprağı Genişliği**	**3- Taç Yaprağı Uzunluğu**	**4- Taç Yaprağı Genişliği**	**5- 1.Sınıf Merkezine Uzaklık**	**6- 2. Sınıf Merkezine Uzaklık**	**7- K**
1	0.000102	0.024743	$3.84*10^{-23}$	0.0001428	1.0121	0.024847	3
2	$2.88*10^{-7}$	0.0678	$9.48*10^{-29}$	0.0001428	1.0011	0.001777	3
3*	$\mathbf{1.93*10^{-10}}$	**0.31499**	$\mathbf{9.31*10^{-31}}$	$\mathbf{2.45*10^{-5}}$	**0.947**	**0.13912**	**3**
4	0.02698	0.024743	$7.78*10^{-33}$	0.0001428	1.0048	0.02824	3
5	0.000102	0.024743	$3.36*10^{-37}$	0.0007098	1.012	0.024849	3
6	$2.59*10^{-9}$	0.0678	$5.54*10^{-35}$	0.0001428	1.0011	0.001777	3
7	$2.39*10^{-6}$	0.00046	$7.68*10^{-42}$	$3.58*10^{-6}$	1.0186	0.038542	3
8	$2.96*10^{-8}$	0.024743	$5.54*10^{-35}$	$3.58*10^{-6}$	1.0122	0.024851	3
9	0.000102	0.0678	$3.36*10^{-37}$	$2.45*10^{-5}$	1.0011	0.001723	3

Veriler bu forma dönüştürüldükten sonra ana destek vektör ve yardımcı destek vektör noktaları bulunmalıdır. Bu aşamada en önemli şart komşuluk sayısı (7. nitelik) değeridir. Aranacak maksimum komşuluk sayısı 3'tür. Ana veya yardımcı destek vektörlerin komşuluk sayısı en az 2 olmalıdır. Bu değer de, maksimum komşuluk sayısının baskın değeridir.

Tablo 5.15 İris Veri Kümesi İçin Sınıflandırıcı 12'nin Gerçek Değerli Çıkışları

1. Çıkış	2. Çıkış
0.65307	1.1403
0.24166	2.0637
0.22644	2.1369
0.20672	2.1165
0.32911	1.053
0.42334	1.0611
0.41667	1.7772
0.59538	1.0505
0.47959	1.1096
0.58896	1.128
0.39714	1.8291
0.58969	1.4805
0.58896	1.128
0.27367	1.0476
0.24301	2.2113
0.26746	2.1387
0.30316	1.039
0.51709	1.0768
0.47766	1.8217
0.63895	0.99012
0.48827	1.0741
0.34676	1.9375
0.65794	1.1458
0.27793	1.876
0.22785	2.1553
1.4835	0.18281
1.5334	0.31867
1.4835	0.18281
1.5264	0.28802
1.1746	0.055066
1.1404	0.087046
1.1364	0.087547
1.1833	0.060616
1.185	0.063285
1.1236	0.33454
1.6638	0.47253
1.6127	0.13076
1.1852	0.062982
1.567	0.14258
1.1418	0.086935
1.1413	0.086947
1.4834	0.18281
1.1835	0.060182
1.2017	0.52384
1.1676	0.051091
1.4732	0.18276
1.5135	0.28246
1.5264	0.28791
1.1961	0.49664
1.1792	0.058247

Daha sonra, 5. ve 6. nitelikleri arası fark değeri en küçük olan veri ana destek vektör olarak belirlenir. Örnekte 3. veri bu şartı sağlamaktadır. Ana destek vektör bulunduktan sonra, aynı şartları sağlayan fakat 5. ve 6. nitelikleri arası fark değeri ana destek vektörden tolerans değeri aralığı kadar müsaade edilen oranda büyük olan noktalar belirlenir. Bu noktalara da, e^{-x} fonksiyonu yapısında azalan oranda destek vektör olma ihtimalleri verilir. Örnekte tolerans değeri 0.09 verilmiştir ve bu şartı sağlayan veriler bulunamamıştır. Böyle bir durumda ilgili veri kümesi yalnız bir destek vektör (ana destek vektör) ile temsil edilir. Destek vektörlere göre ayırıcı fonksiyon bulunur ve eğitim aşaması tamamlanmış olur.

Test aşamasında ise, ilk olarak karşılaşılan verinin hangi kümeye ait olduğu bulunur. Sonra bu küme için bulunan ayırıcı fonksiyondan ait olduğu sınıf belirlenir. Ayrıca çok sınıflı sınıflandırmaya geçerken de, destek vektör noktalarının olasılık değerleriyle ağırlıklandırılmış gerçek sayılar kümesine ait çıkışlar (-1, +1 çıkış değerleri yerine) kullanılarak verinin ait olduğu sınıfa karar verilir. Örneğin, 1 ve 2 numaralı sınıfları ayıran sınıflandırıcı çıkışları Tablo 5.15'deki gibi bulunmuştur.

5.2.3 Geliştirilen Parametre Düzenleyici Yaklaşım Metodu İle Bulunan Uygulama Sonuçları

Bu bölümde, bölüm 4.3'te tanıtılan metot uygulanarak elde edilen uygulama sonuçları anlatılmaktadır. Geliştirilen metot Dopler kalp kapakçığı verilerine uygulanmıştır. Literatürde yapılan araştırmalara göre herhangi bir parametre düzenleyici yaklaşımlı bir DVM metodunun bu alanda bir veri kümesine uygulandığına rastlanılmamıştır. Geliştirilen parametre düzenleyici yaklaşımdan bu yeni alanda iyi sonuçlar elde edilmiştir.

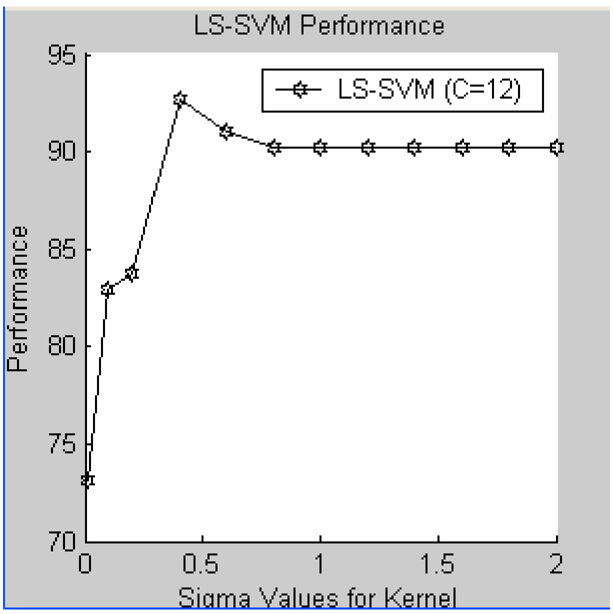
Çekirdek parametresi *(σ)* ve C parametresi değerlerinin sınıflandırma başarımını etkilediğinden daha önceki kısımlarda bahsedilmişti. Şekil 5.8'de, C değeri sabit tutularak σ değerinin değişimine göre elde edilen sınıflandırma başarımları gösterilmektedir.

Tablo 5.16 Optimum Sınıflandırıcı Parametreleri

Veri	K_{init}	N_{init}	YTF Parametresi	C Parametresi
Aort Kapakçığı	4	3	6.383	4.0
Mitral Kapakçığı	4	4	9.572	2.0

Tablo 5.17 Aort Kapakçığı *Kinit*=4 ve *Ninit*=3 İçin Renyi Entropi Değerleri

YTF Par. Değeri	Entropi Değeri (K=4, N=3)	Entropi Değeri (K=3, N=3)	Entropi Değeri (K=2, N=3)
6.1	-3.535	-3.762	-4.012
6.2	-3.537	-3.763	-4.013
6.3	-3.538	-3.764	-4.014
6.4	**-3.540**	**-3.761**	**-4.029**
6.5	-3.541	-3.762	-4.03
6.6	-3.543	-3.763	-4.031
6.7	-3.544	-3.764	-4.032
6.8	-3.545	-3.765	-4.032
6.9	-3.546	-3.766	-4.033



Şekil 5.8 Sigma Değerlerine Göre Performans Değişimi

Tablo 5.16’da, düzenleyici yaklaşımı tarafından bulunan optimum parametre değerleri gösterilmektedir. Bu parametreler Renyi entropi yaklaşımı kümeleme modunda çalıştırılarak elde edilmiştir. Bu değerlerle eğitilen EKKDVM nin en iyi sınıflandırma sonuçlarını verdiği deneysel olarak ta görülmüştür. Elde edilen sonuçların en iyi sonuçlar olduğundan emin olmak için EKKDVM sınıflandırıcısı deneme yanılma yoluyla değiştirilen parametre değerlerine göre de defalarca çalıştırılmıştır. Bu denemelerle de, sonucun en iyi sonuç olduğu doğrulanmıştır.

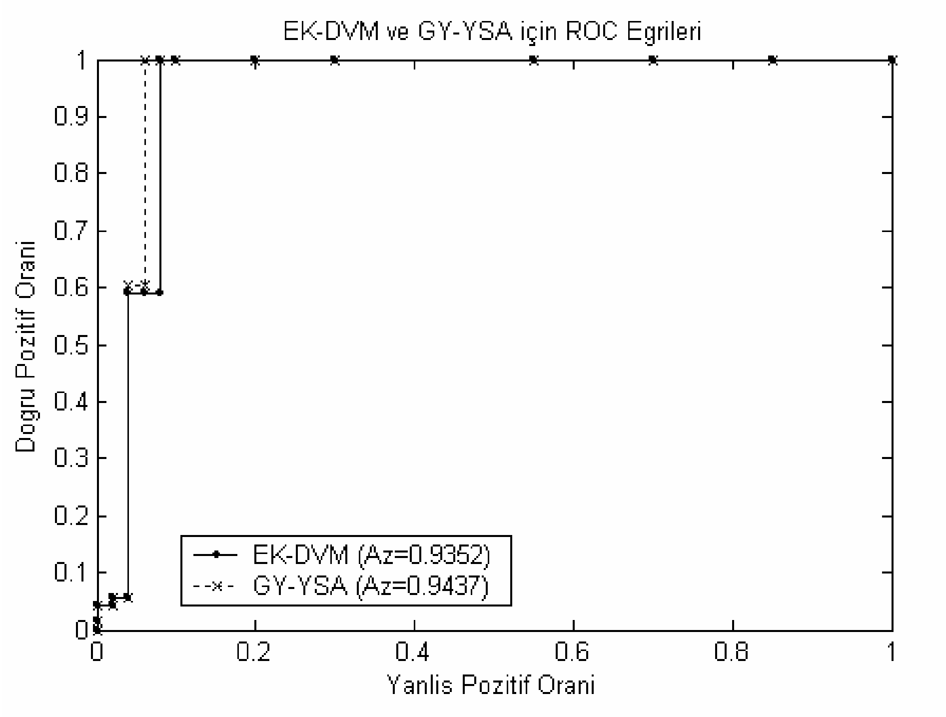
Tablo 5.18’de, aynı veri kümesi üzerinde çalıştırılan üç farklı sınıflandırıcının performansları karşılaştırmalı olarak görülmektedir. EKKDVM nin performansı biraz daha düşük çıkmasına rağmen, sahip olduğu bazı avantajlar nedeniyle tercih edilebilir etkili bir yöntemdir. Bu avantajlar; DVM sınıflandırıcısının yapısından kaynaklanır. DVM her çalıştırıldığında aynı sınıflandırma doğruluğunu verirken, diğer metotlar farklı sonuçlar sergilemektedir. Çünkü DVM de herhangi bir rasgelelik unsuru bulunmamaktadır. DVM fazla uyumluluk problemine yakalanmaz ve genelleme yeteneği oldukça iyidir. Ayrıca, DVM nin eğitim süresi, diğer metotlardan düşüktür. DVM nin en önemli eksikliklerinden olan eğitim parametrelerinin en iyi değerlerinin bulunması ise, bu çalışmada gerçekleştirilen bir yaklaşım ile çözülmüştür. Böylece, DVM metodu hem tam otomatik hem de sürekli aynı sınıflandırma sonucunu veren bir metot haline getirilmiştir. DVM nin daha popüler bir metot olmasının sebebi de bu özelliklere sahip olmasıdır.

Tablo 5.18 EKKDVM, BCO/SSMM ve GY-YSA Sınıflandırıcı Performansları

Sınıflandırıcı	Veri Tipi	Kişi Sayısı	Anor.	Nor.	X=Duyarlılık Y=Seçicilik
BCO/SSMM	Anor.	73	71	2	X=97.3%
	Nor.	50	4	46	Y=92.0%
GY-YSA	Anor.	73	70	3	X=95.9%
	Nor.	50	3	47	Y=94.0%
EKKDVM	Anor.	73	69	4	X=94.5%
	Nor.	50	5	45	Y=90.0%

Anor: Anormal (kalp kapakçığı düzensizliği var), X: Duyarlılık

Nor: Normal (kalp kapakçığı düzensizliği yok), Y: Seçicilik



Şekil 5.9 EKKDVM Ve GY-YSA Sınıflandırma Performanslarının ROC Eğrisi

Şekil 5.9, iki farklı sınıflandırıcının Dopler kalp sesleri veri kümesi üzerindeki performansını göstermektedir. ROC eğrisi ve bu eğrinin altında kalan alan ölçüsü, sınıflandırma işleminin sonunda sınıflandırıcı performanslarını karşılaştırmak için kullanılan güvenilir bir tekniktir. ROC eğrisi, doğru pozitif ve yanlış pozitif oranı değerlerine göre işletilir. Bu yüzden, bu teknik duyarlılık ve seçicilik arasında bir tercih yapılmasını sağlar. Eğrinin altında kalan alan değeri sınıflandırıcının başarımı ile doğru orantılıdır. Daha büyük bir alan değerine sahip olan sınıflandırıcının sınıflandırma doğruluğu diğerinden fazladır (Osareh ve ark., 2002), (Centor 1991). Şekil 5.3'teki A_z değerleri, bu alanın büyüklüğünü yansıtmaktadır. YSA temelli sınıflandırıcının sınıflandırma doğruluğu biraz daha yüksek bulunmuştur. ROC eğrisi, aort ve mitral kalp kapakçığı verilerinin her ikisi de düşünülerek toplamda çizilmiştir.

Bölüm 4.3'te verilen Renyi entropi temelli parametre değeri düzenleyici sistem, güvenilir uygulama sonuçları vermektedir. Sistemin güvenilirliğini gösteren iki temel durum söz konusudur.

Birinci durum; literatürde Huang ve Wang (2006)'ın geliştirdikleri genetik algoritmalar temelli, Shih-Wei Lin ve ark. (2007)'nın geliştirdikleri kısmi sürü optimizasyonu temelli ve Gold ve ark. (2005)’nın geliştirdikleri bayes temelli yaklaşımlardan ilk ikisinin rasgelelik içermesi ve sadece belirli bir uygunluk fonksiyonuna göre işlem yapmasıdır. Üçüncüsü ise sadece eğitim verilerine göre olasılık çıkarımlarının yapıldığı ve büyük veri kümelerinde büyük eğitim zamanları tüketen bir yöntemdir. Fakat Renyi entropi’de merkez noktaları farklı seçilerek defalarca işlem yapıldığı için rasgelelik unsurunun etkisi hemen hiç gözlenmemektedir. Ayrıca veri kümesini temsil eden en küçük veri alt kümesi ustalıkla seçildiği için büyük eğitim zamanı da tüketmemektedir.

İkinci durum ise deneme yanılma yoluyla uzun uğraşlar sonucu elde edilen değerlerle karşılaştırma yapıldığında, eğitim parametrelerin ulaşılabilecek en iyi değerlerinin Renyi entropi ile de bulunduğunun gözlenmesidir. En önemlisi bu değerlere tüm veri kümeleri üzerinde ulaşılabilmektedir.

6. SONUÇLAR VE ÖNERİLER

Bu tez çalışmasında, DVM sınıflandırıcısının zayıf bir takım yönlerini kuvvetlendirmek amacıyla yerel kontrol özelliğine sahip bazı yöntemler DVM eğitim algoritmasına uyarlanarak yeni DVM eğitim algoritmaları geliştirilmiştir. Bu bölümde, tezden elde edilen sonuçlar yorumlanmıştır ve gelecek çalışmalar için bazı öneriler sunulmuştur.

6.1 Sonuçlar

DVM, genellikle sınıflandırma ve regresyon işlemlerinde kullanılan başarılı bir makine öğrenmesi metodudur. DVM, güçlü bir teorik alt yapıya ve iyi bir genelleme yeteneğine sahiptir. Yine bu tez çalışmasında kullanılan KEYK ve Renyi entropi metotları ise, genellikle veriler üzerindeki yerel kontrol kabiliyeti yüksek olan metotlardır. Bu yüzden, daha çok kümeleme ve yerel sınıflandırma gibi makine öğrenmesi işlemlerinde kullanılmaktadırlar.

Bu metotlar ne kadar başarılı metotlar olsalar da, bazı zayıf yönleri bulunmaktadır. Örneğin, bölüm 1.1 ve 4.4 te de bahsedilen DVM nin büyük veri kümelerinde yetersiz kalması problemi mevcuttur. Tez çalışmasında, bu problemin verilerin yerel kontrolü ile aşılacağı düşünülmüştür ve KEYK ile Renyi entropi metotları DVM eğitim algoritmasına adapte edilmiştir. Tezin ana amacı, bu metotlar arası adaptasyonun en iyi şekilde sağlanmasıdır. Ayrıca bu uyarlama sağlanırken, eğitim zamanının da mümkün olduğunca azaltılması amaçlanmıştır. Uygulama sonuçlarına baktığımızda, başlangıçta düşündüğümüz sonuçlara büyük ölçüde ulaşıldığı görülmektedir.

Uygulama sonuçlarında özellikle EKKDVM metodu ile yapılan kıyaslamalar büyük bir öneme sahiptir. Çünkü bu metot da, tezde geliştirilen metotlardaki amaç çerçevesinde düşünülerek tasarlanmıştır. EKKDVM ayrıca, birçok uygulama alanına defalarca uygulanmıştır. DVM eğitim süresini bir miktar düşürmüş ve sınıflandırma doğruluğundan da fazla taviz vermemiştir. Fakat geliştirdiğimiz metotların daha iyi sınıflandırma doğruluğuna daha kısa eğitim süresinde ulaştığı görülmektedir.

Tez çalışmasında, üç çeşit metot önerilmiştir. Birincisi, basit bir kümeleme planı ile KEYK metodunu kullanan KKDVM eğitim algoritmasıdır. İkincisi, Renyi entropi temelli kümeleme ve KEYK metotlarını birlikte kullanan RKKDVM eğitim algoritmasıdır. Son olarak ta, Renyi entropi metodunu DVM eğitim parametrelerinin optimizasyonu için kullanan bir algoritma önerilmiştir.

KKDVM ve RKKDVM yöntemleri dört aşamadan oluşan yöntemlerdir. Bu aşamaların ilk üçü eğitim sonuncusu ise test aşamasıdır. Her iki yöntem de, veriler üzerinde yerel kontrol kabiliyetine sahiptir, destek vektör noktalarını olasılık tabanlı değerlerle temsil edebilmektedirler ve uygulama sonuçlarına göre orijinal DVM ile EKKDVM'ne göre kısa eğitim zamanı tüketmektedirler. Aynı zamanda sınıflandırma başarımları da DVM ve EKKDVM'ne oranla yüksektir. Bu özellikler de KKDVM ile RKKDVM'nin etkinliğini kanıtlamaktadır. Ayrıca çok sınıflı sınıflandırma uygulamalarında da bulunan destek vektör noktalarının olasılık tabanlı değerlerinden yararlanılarak çok sınıflı sınıflandırma problemleri aşılmaktadır.

KKDVM ile RKKDVM aralarında mukayese edildiğinde, KKDVM'nin daha kısa eğitim zamanına sahip olduğu uygulama sonuçlarından görülmektedir. Ancak RKKDVM, KKDVM'ne oranla genelde daha güvenilir ve yüksek başarımlı sınıflandırma yapabilmektedir. Ayrıca RKKDVM'de kümeleme işlemi tam otomatik olarak yapılabilmekte ve eşik değerine gerek duyulmamaktadır. İki yöntem arası farklılıklar RKKDVM'de Renyi entropi yönteminin kulanılmasından kaynaklanmaktadır.

Öğrenme parametresi değerlerinin optimizasyonunu sağlayan sistem ise tüm veri kümeleri üzerinde çalıştırılabilecek şekilde tasarlanmıştır ve en iyi parametre değerlerine başarıyla ulaşılabilmektedir. Böylece DVM temelli sınıflandırıcıların en iyi sınıflandırma başarımına ulaşılabilmektedir. Bölüm 5.2'de belirtildiği gibi, bu sistemin literatürdeki alternatiflerine kıyasla, rasgelelikten kurtulma ve eğitim zamanın kısalığı gibi avantajları bulunmaktadır.

Önerilen metotların ortak yönü, hesap yükünün fazla olmaması ve uygulanma kolaylığıdır. Bu avantajlar KEYK ve Renyi entropi metotlarının yapısından kazanılmıştır. Genel olarak bu metotların zayıf kaldığı yönler de DVM nin güçlü olduğu yönlerdir. Geliştirdiğimiz metotların diğer bir avantajı, genel yapıda tasarlanmasıdır. Sadece bir veya birkaç veri kümesi üzerinde değil tüm sınıflandırma problemlerine uygulanabilirler. Ayrıca, destek vektör noktalarını olasılık tabanlı olarak ve yerel kabiliyetleri yüksek metotların yardımıyla hesaplayan literatürdeki ilk çalışmadır. Bu yenilikler DVM ne hem ikili sınıflandırmada hem de çok sınıflı sınıflandırmada önemli avantajlar sağlamıştır.

6.2 Öneriler

DVM metodunun daha etkin çalışması amacıyla geliştirilen metotlar iyi uygulama sonuçları verse de, kuvvetlendirilebilecek yönleri bulunmaktadır. Bunlardan en önemlisi kümeleme kısmındadır. Yeni geliştirilen birlikte öğrenme tabanlı öğrenme algoritmalarıyla da kümeleme veya yerel değerlendirme yapılabilmektedir. Faming ve ark. (2006) çalışmalarında KSO (Kısmi Sürü Optimizasyonu) kullanarak DVM'ni eğittiler. Fakat bu çalışmada, KSO sadece DVM öğrenme probleminde yer alan optimizasyon probleminin çözümü için kullanılmıştır. Geliştirilen çalışmalarda ise doğrudan optimizasyon probleminin çözülmesi yerine, KSO'dan elde edilebilen değerlerle eğitim verileri analiz edilerek destek vektörler bulunup ayırıcı fonksiyon hesaplanabilmektedir.

Yukarıdaki paragrafta bahsedilen sisteme alternatif olarak, KSO metodu KEYK ve/veya Renyi entropi ile birleştirilerek buradan gelecek değerler üzerinde ortak bir analiz yapılabilir. Bu işlem süresini uzatacaktır. Fakat sınıflandırma doğruluğu da muhtemelen artırılmış olacaktır.

KKDVM ve RKKDVM'de eğitim verileri üzerinde olasılık temelli olarak destek vektör olma ihtimalleri arandığı için bu noktada bulanık sistemlerin kullanılması uygun olabilecektir.

Yine KEYK ve Renyi entropi gibi yerel metotlardan herhangi birisi kullanılarak, literatürde yapay sinir ağları ile kullanılmış olan B-Spline Bulanık üyelik fonksiyonları oluşturulup bu fonksiyon yardımıyla da destek vektör olma ihtimalleri hesaplanabilir.

Ayrıca literatürde, sınıflandırıcı birleştirme adı verilen yaklaşımlar kümeleme algoritmalarına uygulanarak birden fazla kümeleme metodunun sonucu birleştirilip bu yaklaşımlarla en iyi küme grupları bulunabilir.

Bu tez SCI ve SCI Expended türü indekslerde taranan aşağıdaki çalışmalarla,

E. Çomak, A. Arslan. A support vector machine using lazy learning approach for multi-class classification, Medical Engineering & Technology, Taylor & Francis, 30(2) (2006), 73-77.

E. Çomak, A. Arslan. A new training method for support vector machines: Clustering k-NN support vector machines. Expert Systems with Applications, In Press, Corrected Proof, Available online 15 August 2007.

E. Çomak, A. Arslan, İ. Türkoğlu. A decision support system based on support vector machines for diagnosis of the heart valve diseases. Computers in Biology and Medicine 37 (2007-a), 21-27.

E. Çomak, K. Polat, S. Güneş, A. Arslan. A new medical decision making system: Least square support vector machine (LSSVM) with Fuzzy Weighting Pre-processing. Expert Systems with Applications 32 (2) (2007-b), 409-414.

Ayrıca aşağıdaki uluslar arası konferans çalışmasıyla sonuçlandırılmıştır.

O. Fındık, M. Bayrak, İ. Babaoğlu, E. Çomak. Color image watermarking scheme based on efficient preprocessing and support vector machines. LNCS, International Conference on Intelligent Computing (ICIC) 2008 (kabul edildi).

KAYNAKLAR

A. Ben-Hur, D. Horn, H. T. Siegelmann, V. Vapnik. Support Vector Clustering. Journal of Machine Learning Research, 2 (2001), 125-137.

A. K. Jain, R. C. Dubes. Algorithms for Clustering Data. Prentice-Hall, Englewood Cliffs, NJ, 1988.

A. Kulkarni, V.K. Jayaraman, B.D. Kulkarni. Support vector classification with parameter tuning assisted by agent-based technique. Computers and Chemical Engineering 28 (2004), 311-318.

A. Osareh, M. Mirmehdi, B. Thomas, R. Markham. Comparative exudate classification using support vector machines and neural Networks. In: T. Dohi, R. Kikinis (Eds.), Fifth International Conference on Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science, 2489 (2002), 413-420.

A. Renyi. Some Fundamental Questions of Information Theory. Selected Papers of Alfred Renyi, 2 (1976), Akademia Kiado, Budapest, 526-552.

A. Widodo, B. S. Yang, T. Han. Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors. Expert Systems with Applications 32(2) (2007), 299-312.

A. Pozdnoukhov, S. Bengio. Invariances in kernel methods: From samples to objects. Pattern Recognition Letters 27(10) (2006), 1087-1097.

B. Samanta. Gear fault Detection Using Artificial Neural Network and Support Vector Machines with Genetic Algorithms. Mechanical Systems and Signal Processing, 18(3) (2004), 625-644.

B. Schölkopf, A. Smola. Learning with kernels. MIT Press, 2002.

C. Angulo, X. Parra, A. Catala. K-SVCR, A support Vector Machine for Multi-Class Classification. Neurocomputing 55(1-2) (2003), 57-77.

C. Cachin. Smooth Entropy and Renyi Entropy. In: Lecture Notes in Computer Science, (ed. Walter Fumy), Advances in Cryptology: Eurocrypt'97, 1233 (1997), 193-208.

C. Carpineto, G. Romano. A Lattice Conceptual Clustering System and its Application to Browsing Retrieval. Machine Learning, 24 (2) (1996), 96-122.

C. E. Shannon. A mathematical theory of communication. Bell System Technical Journal, 27 (1948), 379-423, 623-656.

C. F. Lin, S. D. Wang. Training algorithms for fuzzy support vector machines with noisy data. Pattern Recognition Letters 25 (2004), 1647-1656.

C. Gold, A. Holub, P. Sollich. Bayesian approach to feature selection and parameter tuning for support vector machine classifiers. Neural Networks 18 (2005), 693–701.

C. J. Du, D. W. Sun. Pizza sauce spread classification using colour vision and support vector machines. Journal of Food Engineering 66 (2005), 137–145.

C. L. Huang, C. J. Wang. A GA-based feature selection and parameters optimization for support vector machines. Expert Systems with Applications 31(2) (2006), 231-240.

C. M. Vong, P. K. Wong, Y. P. Li. Prediction of automotive engine power and torque using least squares support vector machines and Bayesian inference. Engineering Applications of Artificial Intelligence 19 (2006), 277–287.

D. Erdogmus, J. C. Principe. Lower and upper Bounds for Misclassification Probability Based on Renyi's Information. Journal of VLSI Signal Processing special Issue on Wireless Communications and Blind Signal Processing, 37 (2004), 305-317.

D. H. Hong, C. Hwang. Support vector fuzzy regression machines. Fuzzy Sets and Systems 138(2) (2003), 271-281.

D. Judd, P. McKinley, A. K. Jain. Large-Scale Parallel Data Clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (8) (1998), 871-876.

D. Kim, H. J. Lee, S. Cho. Response modeling with support vector regression. Expert Systems with Applications 34(2) (2008), 1102-1108.

D. Tsujinishi, S. Abe. Fuzzy least squares support vector machines for multi-class problems. Neural Networks 16 (2003), 785-792.

D. Xu. Energy, Entropy and Information Potential for Neural Computation. Doktora Tezi, University of Florida, 1999.

E. Çomak, 2004. Destek Vektör Makineleri Çoklu Sınıf Problemleri için Çözüm Önerileri. Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği A.B.D.,Yüksek Lisans Tezi.

E. Çomak, A. Arslan. A support vector machine using lazy learning approach for multi-class classification, Medical Engineering & Technology, Taylor & Francis, 30(2) (2006), 73-77.

E. Çomak, A. Arslan. A new training method for support vector machines: Clustering k-NN support vector machines. Expert Systems with Applications, In Press, Corrected Proof, Available online 15 August 2007.

E. Çomak, A. Arslan, İ. Türkoğlu. A decision support system based on support vector machines for diagnosis of the heart valve diseases. Computers in Biology and Medicine 37 (2007-a), 21-27.

E. Çomak, K. Polat, S. Güneş, A. Arslan. A new medical decision making system: Least square support vector machine (LSSVM) with Fuzzy Weighting Pre-processing. Expert Systems with Applications 32 (2) (2007-b), 409-414.

E. D. Übeyli. Combining eigenvector methods and support vector machines for detecting variability of Doppler ultrasound signals. Computer Methods and Programs in Biomedicine 86 (2) (2007), 181-190.

E. Gokcay, J. Principe. Information Theoretic Clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2) (2002), 158-170.

E. H. S. Han, G. Karypis, V. Kumar. Text categorization using weight adjusted k-nearest neighbor classification. Lecture Notes in Computer Science, 2035 (2001), 53-67.

G. Bontempi, H. Bersini, M. Birattari. The local paradigm for modeling and control: from neuro-fuzzy to lazy learning. Fuzzy Sets and Systems 121 (2001), 59-72.

G. Bontempi, M. Birattari, H. Bersini. Recursive lazy learning for modeling and control. In Machine Learning: ECML-98 (10 th European Conference on Machine Learning), (1998), 292-303.

G. Camps-Valls, J. D. Martin-Guerrero, J. L. Rojo-Alvarez, E. Soria-Olivas. Fuzzy sigmoid kernel for support vector classifiers. Neurocomputing 62 (2004), 501–506.

H. Frigui, R. Krishuapuram. A Robust Competitive Clustering Algorithm with Applications in Computer Vision. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21 (5) (1999), 450-465.

H. Ince, T. B. Trafalis. Kernel methods for short-term portfolio management. Expert Systems with Applications 30 (2006), 535–542.

H. M. Abbas, M. M. Fahmy. Neural Networks for Maximum Likelihood Clustering. Signal Processing, 36 (1) (1994), 111-126.

H. Uğuz, A. Arslan, İ. Türkoğlu. A biomedical system based on hidden Markov model for diagnosis of the heart valve diseases. Pattern Recognition Letters, 28 (4) (2007), 395 – 404.

İ. Türkoğlu. Durağan olmayan işaretler için zaman-frekans entropilerine dayalı akıllı örüntü tanıma, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, 2002.

İ. Türkoglu, A. Arslan, E. İlkay. An expert system for diagnosis of the heart valve diseases. Expert Systems with Applications, 23 (2003), 229–236.

J. A. K. Suykens, J. Vandewalle. Least squares support vector machine classifiers. Neural Processing Letters, 9 (3) (1999), 293-300.

J. H. Hong, J. K. Min, U. K. Cho, S. B. Cho. Fingerprint classification using one-vs-all support vector machines dynamically ordered with naive Bayes classifiers. Pattern Recognition, 41(2) (2008), 662-671.

J. K. Kim, G. P. S. Raghava, S. Y. Bang, S. Choi. Prediction of subcellular localization of proteins using pairwise sequence alignment and support vector machine. Pattern Recognition Letters, 27(9) (2006), 996-1001.

J. M. Lerski. On support vector regression machines with linguistic interpretation of the kernel matrix. Fuzzy Sets and Systems, 157(8) (2006), 1092-1113.

J. M. Martinez-Otzeta, B. Sierra, E. Lazkano, A. Astigarraga. Classifier hierarchy learning by means of genetic algorithms. Pattern Recognition Letters 27 (2006), 1998–2004.

J. Principe, D. Xu, J. Fisher. Unsupervised Adaptive Filtering. 1, chapter 7, Information Theoretic Learning, John Wiley & Sons, 2000.

J. Shen. Fusing Support Vector Machines and Soft Computing for Pattern Recognition and Regression. Kansas State University Manhattan, Kansas, Doktora Tezi, 2005.

J. Shen, Z. J. Pei, E. S. Lee. Support Vector Regression in the Analysis of Soft-Pad Grinding of Wire-Sawn Silicon Wafers. CITSA 2004/ISAS 2004, International Conference on Cybernetics and Information Technologies, Systems and Applications/The 10th International Conference on Information Systems Analysis and Synthesis, (2004), 19-24.

J. Suykens, T. Gestel, J. Brabanter, B. Moor, J. Vandewalle. Least Squares Support Vector Machines. World Scientific Pub.Co.,Singapore, 2002.

J. Weston, C. Watkins. Multi-class support vector machines. Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, 1998.

J. Zhang, Y. Liu. SVM decision boundary based discriminative subspace induction. Pattern Recognition, 38 (2005), 1746–1758.

Jayadeva, R. Khemchandani, S. Chandra. Fast and robust learning through fuzzy linear proximal support vector machines. Neurocomputing, 61 (2004), 401–411.

K. Coussement, D. V. Poel. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. Expert Systems with Applications 34(1) (2008), 313-327.

K. Crammer, Y. Singer. On the algorithmic implementation of multi-class kernel-based vector machines. Machine Learning Research, 2 (2001), 265-292.

K. Lu, J. Zhao, D. Cai. An algorithm for semi-supervised learning in image retrieval. Pattern Recognition, 39 (2006), 717–720.

K. S. Goh, E. Chang, K. T. Cheng. SVM Binary Classifier Ensembles for Image Classification. CIKM'01. Atlanta, Georgia, USA, 2001.

K. S. Chua. Efficient computations for large least square support vector machine classifiers. Pattern Recognition Letters, 24 (2003), 75–80.

K. Y. Chen, C. H. Wang. A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. Expert Systems with Applications 32(1) (2007), 254-264.

L. H. Chiang, M. E. Kotanchek, A. K. Kordon. Fault diagnosis based on Fisher discriminant analysis and support vector machines. Computers and Chemical Engineering 28 (2004), 1389–1401.

M. H. Yang. Gentle Guide to Support Vector Machines. Presentation 2002.

N. Acır, Ö. Özdamar, C. Güzeliş. Automatic classification of auditory brainstem responses using SVM-based feature selection algorithm for threshold detection. Engineering Applications of Artificial Intelligence 19 (2006), 209–218.

N. V. Vapnik. The Nature of Statistical Learning Theory, Springer-Verlag, NY, 1995.

N. V. Vapnik. Statistical Learning Theory, John Wiley & Sons, Toronto, CA, 1998.

N. V. Vapnik, A. Y. Chervonenkis. On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities. Theory of Probability and Its Applications, 16 (2) (1971), 264-280.

N. V. Vapnik, A. Y. Chervonenkis. The necessary and sufficient conditions for consistency in the empirical risk minimization method. Pattern Recognition and Image Analysis, 1(3) (1991), 283-305.

P. Hong, D. Zhang, T. Wu. An Intrusion Detection Method Based on Rough Set and SVM Algorithm. IEEE Transactions on Communications, Circuits and Systems 2 (2004), 1127-1130.

P. Mantero, G. Moser, S. B. Serpico. Partially supervised classification of remote sensing images using SVM-based probability density estimation. IEEE Transactions on Geoscience and Remote Sensing, 44(8) (2006), 327-336.

P. Sahoo, C. Wilkins, J. Yeager. Threshold Selection Using Renyi's Entropy. Pattern Recognition, 30 (1997), 71-84.

P. Shih, C. Liu. Face detection using discriminating feature analysis and Support Vector Machine. Pattern Recognition 39 (2006), 260–276.

R. Bellman. Adaptive Control Processes: A Guided Tour. Princeton University Press, New Jersey, 1961.

R. Burbidge, B. Buxton. An Introduction to Support Vector Machines for Data Mining, Young OR 12, University of Nottingham, (2001), 3-15.

R. Jenssen, K. E. Hild, D. Erdogmus, J. C. Principe, T. Eltoft. Clustering using Renyi's entropy. Neural Networks, 2003. Proceedings of the International Joint Conference. 1 (2003), 523-528.

R. M. Centor. Signal detectability: the use of ROC curves and their analysis. Medical Decision Making (1991), 102–106.

S. Abe, T. Inoue. Fuzzy Support Vector Machines for Multiclass Problems. ESANN'2002 proceedings-European Symposium on Artificial Neural Networks, Bruges-Belgium, d-side publication, (2002), 113-118.

S. C. Chapra, R. P. Canale (Çevirenler: Hasan Heperkan ve Uğur Kesgin). Yazılım Ve Programlama Uygulamalarıyla Mühendisler İçin Sayısal Yöntemler. Literatür Yayıncılık, ISBN: 975-8431-83-8, 2003.

S. Draghici. The constraint based decomposition (CBD) training architecture. Neural Networks 14 (2001), 527-550.

S. H. Kim, S. W. Shin. Identifying the impact of decision variables for nonlinear classification tasks. Expert Systems with Applications 18 (2000), 201-214.

S. H. Min, J. Lee, I. Han. Hybrid genetic algorithms and support vector machines for bankruptcy prediction. Expert Systems with Applications, 31(3) (2006), 652-660.

S. Peddabachigari, A. Abraham, C. Grosan, J. Thomas. Modeling intrusion detection system using hybrid intelligent systems. Journal of Network and Computer Applications, 30(1) (2007), 114-132.

S. Suresh, N. Sundararajan, P. Saratchandran. A sequential multi-category classifier using radial basis function networks. Neurocomputing, 71, 2008, 1345–1358.

S. W. Lin, K. C. Ying, S. C. Chen, Z. J. Lee. Particle swarm optimization for parameter determination and feature selection of support vector machines, Expert Systems with Applications, In Press, Corrected Proof, Available online 17 September 2007.

T. B. Trafalis, H. Ince. Support vector machine for regression and applications to financial forecasting. Neural Networks. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, 6 (2000), 348-353.

T. Eitrich, B. Lang. Efficient optimization of support vector machine learning parameters for unbalanced datasets. Journal of Computational and Applied Mathematics, 196 (2006), 425–436.

T. Eltoft, R. J. P. Figueiredo. A New Neural Network for Cluster-Detection-and-Labeling, IEEE Transactions on Neural Networks, 9 (5) (1998), 1021-1035.

T. Faming, C. Mianyun, W. Zhongdong. New approach to training support vector machine. Journal of Systems Engineering and Electronics, 17(1) (2006), 200-205.

T. G. Dietterich, G. Bakiri. Solving multiclass learning problems via error-correcting output codes. Journal of artificial intelligence research 2 (1995), 263-286.

T. Kikuchi, S. Abe. Comparison between error correcting output codes and fuzzy support vector machines. Pattern Recognition Letters, 26 (2005), 1937–1945.

T. Li, Q. Li, M. Ogihara. Gene Functional Classfication by Semi-supervised Learning from Heterogenous Data. In ACM SAC Bioinformatics Track, Melbourne, Florida, USA, 2003.

UCI Repository of Machine Learning Databases. ftp://ftp.ics.uci.edu /pub/machine-learning-databases.

V. Cherkassky. Learning from Data, John Wiley and Sons, 23/0-471-15493-8, 1998.

V. Cherkassky, Y. Ma. Multiple Model Regression Estimation. IEEE Transactions On Neural Networks, 16 (2005), 785-798.

V. D. A. Sanchez. Advanced support vector machines and kernel methods, Neurocomputing, 55 (2003), 5-20.

V. Kecman. Learning and Soft Computing: Support Vector Machines. Neural Networks and Fuzzy Logic Models, MIT Press, Cambridge, Massachusetts, 2001.

W. H. Chen, J. Y. Shih. A study of Taiwan's issuer credit rating systems using support vector machines. Expert Systems with Applications 30 (2006), 427–435.

W. Lu, W. Wang, A. Leung, S. Lo, R. Yuen, Z. Xu, H. Fan. Air Pollutant Parameter Forecasting Using Support Vector Machines. IJCNN '02,Proceedings of the 2002 International Joint Conference on Neural Networks, 1 (2002), 630-635.

W. Zhong, J. He, R. Harrison, P. C. Tai, Y. Pan. Clustering support vector machines for protein local structure prediction. Expert Systems with Applications 32(2) (2007), 518-526.

Y. Chen, G. Wang, S. Dong. Learning with progressive transductive support vector machine, Pattern Recognition Letters 24 (2003), 1845–1855.

Y. Dong, Z. Xia, Zu. Xia. A two-level approach to choose the cost parameter in support vector machines. Expert Systems with Applications 34(2) (2008), 1366-1370.

Y. Wang, S. T. Huang. Training TSVM with the proper number of positive samples. Pattern Recognition Letters, 26 (2005), 2187–2194.

Y. Zhan, D. Shen. An adaptive error penalization method for training an efficient and generalized SVM. Pattern Recognition, 39 (2006), 342–350.

Y. Zhan, D. Shen. Design efficient support vector machine for fast classification. Pattern Recognition, 38 (2005) 157–161.

Y. Zhang, C. H. Chu, Y. Chen, H. Zha, X. Ji. Splice site prediction using support vector machines with a Bayes kernel. Expert Systems with Applications 30 (2006), 73–81.

Z. Sun, G. Bebis, R. Miller. Object detection using feature subset selection. Pattern Recognition 37 (2004), 2165 – 2176.