



LORENTZ UZAYINDA SINIFLANDIRMA

Yerzhan KERİMBEKOV

DOKTORA TEZİ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

HAZİRAN 2017

Yerzhan KERİMBEKOV tarafından hazırlanan “LORENTZ UZAYINDA SINIFLANDIRMA” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Bilgisayar Mühendisliği Anabilim Dalında DOKTORA TEZİ olarak kabul edilmiştir.

Danışman : Doç. Dr. Hasan Şakir BİLGE

Elektrik-Elektronik Mühendisliği, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum

.....

Başkan : Prof. Dr. Mehmet Ali AKCAYOL

Bilgisayar Mühendisliği, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum

.....

Üye : Doç. Dr. Fırat HARDALAÇ

Elektrik-Elektronik Mühendisliği, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum

.....

Üye : Doç. Dr. Muhammed Fatih DEMİRCİ

Bilgisayar Mühendisliği, TOBB Ekonomi ve Teknoloji Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum

.....

Üye : Yrd. Doç. Dr. Ahmet Murat ÖZBAYOĞLU

Bilgisayar Mühendisliği, TOBB Ekonomi ve Teknoloji Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum

.....

Tez Savunma Tarihi: 02/06/2017

Jüri tarafından kabul edilen bu tezin Doktora Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Prof. Dr. Hadi GÖKÇEN

Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Yerzhan KERİMBEKOV

02/06/2017

LORENTZ UZAYINDA SINIFLANDIRMA

(Doktora Tezi)

Yerzhan KERİMBEKOV

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Haziran 2017

ÖZET

Bu çalışmada Lorentz uzayının sınıflandırma probleminde kullanımı konusu ele alınmıştır. Bu doğrultuda ilk olarak Lorentz uzayındaki noktalar arası komşuluk kavramı ele alınmıştır. Buna göre Lorentz uzayında noktaların dağılımı ve pozisyonunun sınıflandırma açısından önemli olduğu tespit edilmiştir. Bu bağlamda Destek Vektör Makinesi veya Doğrusal Diskriminant Analizi yöntemlerinin karar doğruları ile Lorentz metriği açısından en iyi sınıflandırma başarısını üretecek en iyi pozisyona taşıma yöntemi geliştirilmiştir. Ayrıca, Lorentz uzaklığının sınıflandırma probleminde ayırt edici ölçüt olarak kullanım tekniği ortaya koyulmuştur ve matris çarpımı (sıkıştırma), döndürme ve öteleme gibi temel matematik işlemlerden yararlanarak Lorentz uzayında sınıflandırma başarısını arttıran önışlem adımı önerilmiştir. Böylece, iki boyutlu Lorentz uzayında sınıflandırma yapan Lorentz uzaklığı temelindeki yeni Sıkıştırma ile Lorentz Sınıflandırıcısı (S-LS) geliştirilmiştir. Bu çalışmada önerilen S-LS yönteminin önışlem adımı döndürme ve sıkıştırma matrisinin en iyi parametrelerini hesaplama adımları ile genişletilerek sırasıyla Döndürme ile Lorentz Sınıflandırıcısı (D-LS) ve Optimal Parametrelili Lorentz Sınıflandırıcısı (OP-LS) yöntemleri geliştirilmiştir. Bu tez araştırmasında ikiden çok özneteliği sahip veri kümelerinin Lorentz uzayında sınıflandırılması amacıyla yeni Çok Boyutlu Lorentz Sınıflandırıcısı (ÇBLS) önerilmiştir. Ayrıca, çok boyutlu verilerde boyut indirgeme ve öznetelik seçme amaçlı Lorentz metriği temelli Lorentz Öznetelik Seçme yöntemi geliştirilmiştir. Bu çalışmada önerilen ÇBLS ve genişletilmiş versiyonu olan Optimal Parametrelili Çok Boyutlu Lorentz Sınıflandırıcısı (OP-ÇBLS) ve Döndürme ile Çok Boyutlu Lorentz Sınıflandırıcısı (D-ÇBLS) yöntemlerinin sınıflandırma performansını test etmek amacıyla denemelerde GESTURE, LVST, MADELON, RELAX, SONAR, SEEDS, TELESCOPE, WINE ve WISCONSIN veri kümeleri kullanılmıştır. Daha sonra, önerilen ÇBLS yöntemi çok boyutlu yüz resimleri için uyarlanarak yeni Lorentz Yüz Doğrulama (LYD) yöntemi geliştirilmiştir. Geliştirilen yeni LYD algoritması Lorentz uzaklık ölçütü ile doğru resim ile yanlış olanlarını benzerlik değerine göre ayırt etmektedir. Bunlara ek olarak, bu çalışmada Lorentz metriği temelinde sınıflandırma yapan yeni Lorentz Destek Vektör Makinesi yöntemi geliştirilmiştir. Bu yöntemlerin hepsinde klasik sınıflandırıcılara göre daha başarılı sonuçlar elde edilmiştir.

Bilim Kodu : 92431

Anahtar Kelimeler : Lorentz uzayı, Lorentz metriği, sınıflandırma, yüz doğrulama

Sayfa Adedi : 113

Danışman : Doç. Dr. Hasan Şakir BİLGE

CLASSIFICATION IN LORENTZIAN SPACE
(Ph. D. Thesis)

Yerzhan KERİMBEKOV

GAZİ UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
June 2017

ABSTRACT

In this study the use of Lorentzian space in classification problem was investigated. For this purpose, first of all, the neighborhood of points in Lorentzian space was presented. According to this neighborhood, it was found that the position and distribution of points are significant in terms of classification in Lorentzian space. In this regard, the separate lines of Support Vector Machine and Linear Discriminant Analysis methods were used to transform the points to optimal position that produces best classification results according to Lorentzian metric. Also, the technique of using Lorentzian distance in classification problem as discrimination criterion was suggested and a pre-processing step based on simple mathematical operations as matrix multiplication (compression), rotation and shifting that increases classification rate in Lorentzian space was developed. Thus, the Lorentzian Classifier via Compression (LC-C) method based on the Lorentzian distance in two dimensional Lorentzian space was developed. In this study, the pre-processing step of suggested LC-C method reorganized by rotation and computing the optimal parameters of compression matrix and the extended methods Lorentzian Classifier via Rotation (LC-R) and Lorentzian Classifier with Optimal Parameters (LC-OP) were obtained respectively. Subsequently, in this study, for purpose of classification of the multidimensional data in Lorentzian space the new Multidimensional Lorentzian Classifier (MLC) was developed. Also, to reduce the dimension in multidimensional data and feature selection the Lorentzian Feature Selection (LFS) method based on Lorentzian metric was proposed. In this study, the proposed MLC and the extended Multidimensional Lorentzian Classifier via Rotation (MLC-R) and MLC with Multidimensional Lorentzian Classifier with Optimal Parameters (MLC-OP) methods were tested by GESTURE, LVST, MADELON, RELAX, SONAR, SEEDS, TELESCOPE, WINE and WISCONSIN data sets. Subsequently, the suggested MLC was applied to face data sets and the new Lorentzian Face Recognition (LFR) method was developed. The proposed LFR method separates the right face image from wrong one according to similarity value with Lorentzian distance. Additionally, in this study, the new Lorentzian Support Vector Machine method based on Lorentzian metric was developed. In all of this methods, better results are obtained according to classical classifiers.

Science Code : 92431

Key Words : Lorentzian space, Lorentzian metric, classification, face verification

Page Number : 113

Supervisor : Assoc. Prof. Dr. Hasan Şakir BİLGE

TEŞEKKÜR

Doktora eğitimim süresince, bilgisi ve değerli yorumları ile her zaman yanımda olan, çalışma disiplini hep örnek alacağım çok değerli hocam ve danışmanım Doç. Dr. Hasan Şakir BİLGE' ye çok teşekkür ederim.

Doktora tezi çalışmamı hazırlama sürecinde TİK üyesi olarak değerli yorumlarını ve düzeltmelerini esirgemeyerek tezimi daha değerli kılan Prof. Dr. M. Ali AKCAYOL ve Doç. Dr. Muhammed F. DEMİRCİ hocalara teşekkür ederim.

Ankara'ya ilk geldiğim günden itibaren bana ağabey ve aile olan, eğitim boyunca maddi ve manevi katkılarını, dostluklarını eksik etmeyen, sevgili yol arkadaşlarım Emre ÖZER, İsmail BAYRAK, Dinmuhammed KELESBAYEV, Mirzahan BAYBOLOV'a teşekkür ederim.

Doktora eğitimi için Türkiye'ye gelmeme sebep olan ve doktora eğitimi süresince burs olarak maddi destek veren Ahmet Yesevi Üniversitesi Mütevelli Heyet Başkanlığına ve çalışanlarına Mütevelli Heyet Başkanı Prof. Dr. Musa YILDIZ' ın şahsında teşekkür ederim. Bu tez çalışmasının konusu TÜBİTAK 3001 Başlangıç Ar-Ge projesi kapsamında desteklenen 115E181 numaralı proje ile örtüşmektedir. Bu bağlamda projede araştırmacı olarak görev yapma fırsatını sağladığı ve burs desteği verdiği için TÜBİTAK'a teşekkür ederim.

TÜBİTAK 3001 projesinde proje danışmanı olarak görev alan Prof. Dr. Hasan Hüseyin UĞURLU hocaya bilgi ve tecrübelerini paylaştığı için teşekkür ederim.

Ayrıca, doktora eğitimi boyunca hep yanımda olan, destek veren ve her konuda elinden geldiği kadar yardımcı olan eşim Alina ZHUZBAYEVA'ya teşekkür ederim

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	xv
SİMGELER VE KISALTMALAR.....	xviii
1. GİRİŞ.....	1
2. LORENTZ UZAYI	7
2.1. Lorentz Metriği	7
2.2. Lorentz Uzayının Özellikleri	9
2.3. Lorentz Uzayında Noktalar Arası Komşuluk Tanımı	11
2.4. Uygulamada Kullanılan Veri Kümeleri	12
3. LORENTZ UZAYININ SINIFLANDIRMADA KULLANIMI.....	17
3.1. Önişlem	17
3.1.1. Öteleme, sıkıştırma ve döndürme	17
3.1.2. DVM ve DDA karar çizgilerinin kullanımı	20
3.2. Lorentz Uzaklığı Temelli Ayırt Etme Ölçütü	27
3.3. İki Boyutlu Veriler İçin Lorentz Sınıflandırıcısı	29
3.3.1. Lorentz uzaklığı temelli sınıflandırma	29
3.3.2. İki boyutlu Lorentz Sınıflandırıcısı.....	29
3.3.3. Karar çizgisi temelli Lorentz Sınıflandırıcısı.....	30
3.3.4. Sıkıştırma tabanlı Lorentz Sınıflandırıcısı	31

	Sayfa
3.4. Çok Boyutlu Veriler İçin Lorentz Sınıflandırıcısı	37
3.4.1. İkili öznitelik grupları	38
3.4.2. Lorentz metriği temelli öznitelik seçme	38
3.4.3. Çok Boyutlu Lorentz Sınıflandırıcısı (ÇBLS)	40
3.5. ÇBLS İle Yüz Doğrulama	43
3.6. Lorentz Uzayında Destek Vektör Makinesi.....	44
3.6.1. Destek Vektör Makinesi	45
3.6.2. Destek Vektör Makinesi ve Lorentz uzayı	49
3.6.3. Lorentz metriği temelli Destek Vektör Makinesi	51
4. DENEYSEL SONUÇLAR VE BULGULAR	53
4.1. İki Boyutlu Veriler İçin Lorentz Sınıflandırıcısı Sonuçları	53
4.2. Çok Boyutlu Lorentz Sınıflandırıcısı Sonuçları.....	73
4.3. Yüz Doğrulama Sonuçları	89
4.4. Lorentz Metriği Temelli Destek Vektör Makinesinin Sonuçları.....	95
5. SONUÇLAR VE ÖNERİLER	103
KAYNAKLAR.....	107
ÖZGEÇMİŞ.....	111

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. Veri kümeleri istatistiği	13
Çizelge 2.2. Yüz veri tabanları istatistiği	14
Çizelge 4.1. LUTS yönteminde kullanılan veri kümeleri ile ilgili istatistiksel bilgiler ..	54
Çizelge 4.2. LUTS yönteminin en iyi sınıflandırma sonuçlarının karşılaştırması	57
Çizelge 4.3. LUTS yönteminin k değerine göre sınıflandırma başarılarının karşılaştırması	58
Çizelge 4.4. KÇ-GALS yönteminde kullanılan veri kümeleri ile ilgili istatistiksel bilgiler	59
Çizelge 4.5. RELAX için k komşu sayısına göre sınıflandırma sonuçları.....	61
Çizelge 4.6. SEEDS için k komşu sayısına göre sınıflandırma sonuçları.....	62
Çizelge 4.7. TELESCOPE için k komşu sayısına göre sınıflandırma sonuçları.....	63
Çizelge 4.8. VERTEBRAL için k komşu sayısına göre sınıflandırma sonuçları	64
Çizelge 4.9. WISCONSIN için k komşu sayısına göre sınıflandırma sonuçları.....	64
Çizelge 4.10. S-LS, D-LS ve OP-LS yöntemlerinin sınıflandırma sonuçlarının kıyaslanması. (Öznitelik ikilileri parantez içinde verilmiştir)	67
Çizelge 4.11. OP-LS denemelerinde kullanılan w ve q değerleri.....	68
Çizelge 4.12. CLIMATE veri kümesinden elde edilen 8 ikili öznitelik grubu için sınıflandırma sonuçları.....	70
Çizelge 4.13. PARKINSON veri kümesinden elde edilen 8 ikili öznitelik grubu için sınıflandırma sonuçları.....	70
Çizelge 4.14. RELAX veri kümesinden elde edilen 8 ikili öznitelik grubu için sınıflandırma sonuçları	70
Çizelge 4.15. CLIMATE veri kümesi için k-EYK, S-LS, D-LS ve OP-LS yöntemlerinin k değerinin değişimindeki sınıflandırma sonuçları	72
Çizelge 4.16. GESTURE veri kümesi için k-EYK, S-LS, D-LS ve OP-LS yöntemlerinin k değerinin değişimindeki sınıflandırma sonuçları	73

Çizelge	Sayfa
Çizelge 4.17. VERTEBRAL veri kümesi için k-EYK, S-LS, D-LS ve OP-LS yöntemlerinin k değerinin değişimindeki sınıflandırma sonuçları	73
Çizelge 4.18. Veri kümeleri istatistiği. (g :veri kümesinden seçilen ikili öznitelik grup sayısı)	75
Çizelge 4.19. ÇBLS sonucu ile Bayes, DVM, k-EYK yöntemlerinin orijinal veri kümelerinin kullanımında elde edilen sonuçların karşılaştırması	76
Çizelge 4.20. Veri kümeleri için sıkıştırma matrisi parametrelerinin en iyi değerleri.....	77
Çizelge 4.21. ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemlerinin en iyi sınıflandırma sonuçlarının karşılaştırılması	89
Çizelge 4.22. CMU-PIE yüz veri tabanından seçilen 10 ikili öznitelik grup için sınıflandırma sonuçları.....	93
Çizelge 4.23. ORL yüz veri tabanından seçilen 10 ikili öznitelik grup için sınıflandırma sonuçları.....	94
Çizelge 4.24. Yale B yüz veri tabanından seçilen 10 ikili öznitelik grup için sınıflandırma sonuçları.....	94
Çizelge 4.25. Yüz veri kümeleri için LYD yönteminin en iyi yüz doğrulama oranlarının karşılaştırması.....	95
Çizelge 4.26. Klasik sınıflandırıcıların veri kümelerinin yalın halindeki sınıflandırma sonuçları	96
Çizelge 4.27. LVST kümesinden seçilen 10 ikili öznitelik grubu için LDVM yönteminin sınıflandırma sonuçları	100
Çizelge 4.28. SONAR kümesinden seçilen 10 ikili öznitelik grubu için LDVM yönteminin sınıflandırma sonuçları	101
Çizelge 4.29. TELESCOPE kümesinden seçilen 10 ikili öznitelik grubu için LDVM yönteminin sınıflandırma sonuçları	101
Çizelge 4.30. WISCONSIN kümesinden seçilen 10 ikili öznitelik grubu için LDVM yönteminin sınıflandırma sonuçları	101
Çizelge 4.31. LDVM yönteminin en iyi sınıflandırma sonuçları	102

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. İki boyutlu Lorentz uzayı.....	8
Şekil 2.2. Lorentz ve Öklid metriklerinin karşılaştırılması.....	10
Şekil 2.3. Öklid uzayında (a) ve Lorentz uzayında (b) komşu noktaların dağılımı.....	12
Şekil 2.4. Yale B yüz veri tabanından örnek yüz görüntüleri.....	14
Şekil 2.5. ORL yüz veri tabanından örnek yüz görüntüleri	14
Şekil 2.6. CMU PIE yüz veri tabanından örnek yüz görüntüleri	15
Şekil 3.1. İki sınıflı Gauss dağılımı için: a) orijinal hali, b) sınıfların λ matrisi ile sıkıştırılmış hali.....	19
Şekil 3.2. Döndürme işleminin sonucu	20
Şekil 3.3. DVM karar çizgisi	22
Şekil 3.4. Karar doğrusunun y eksenine paralel hali	22
Şekil 3.5. Karar doğrusunun x eksenine paralel hali	23
Şekil 3.6. Karar doğrusunun null ekseninin üzerindeki hali	24
Şekil 3.7. a) İki sınıflı veri kümesi için DVM yardımıyla hesaplanan karar çizgisi, b) orijine ötelenmiş karar çizgisi, c) karar çizgisi ile null eksen arasındaki açı kadar karar çizgisini ve d) verileri döndürme	26
Şekil 3.8. Karar çizgileri a) DDA ve b) DVM.....	27
Şekil 3.9. CLIMATE için k-EYK ve S-LS algoritmalarının başarı oranları	33
Şekil 3.10. CLIMATE veri kümesi için S-LS algoritmalarının a) w ve b) q parametrelerinin farklı değerleri için ürettiği sınıflandırma sonuçları.....	34
Şekil 3.11. CLIMATE veri kümesi için k-EYK, S-LS ve OP-LS algoritmalarının sınıflandırma sonuçlarının karşılaştırması	35
Şekil 3.12. VERTEBRAL veri kümesi için döndürme.....	36
Şekil 3.13. Destek Vektör Makinesinin ayırt edici çizgisi.....	45
Şekil 3.14. Sınıfları ayıran hiperdüzlem ve null doğrusu ile dik açıyla kesişen doğru...	50

Şekil	Sayfa
Şekil 3.15. Lorentz uzayındaki diklik	51
Şekil 4.1. k parametresinin değişimine göre ECOLI veri kümesi için LUTS ve diğer sınıflandırıcıların başarı grafiği.....	55
Şekil 4.2. k parametresinin değişimine göre ILPD veri kümesi için LUTS ve diğer sınıflandırıcıların başarı grafiği.....	56
Şekil 4.3. k parametresinin değişimine göre SEEDS veri kümesi için LUTS ve diğer sınıflandırıcıların başarı grafiği.....	56
Şekil 4.4. k parametresinin değişimine göre VERTEBRAL veri kümesi için LUTS ve diğer sınıflandırıcıların başarı grafiği	57
Şekil 4.5. RELAX veri kümesi için k değerlerine göre sınıflandırma oranları	60
Şekil 4.6. SEEDS veri kümesi için k değerlerine göre sınıflandırma oranları	62
Şekil 4.7. CLIMATE için k değerinin değişimindeki sınıflandırma sonucu	72
Şekil 4.8. ÇBLS yönteminin GESTURE veri kümesi için sınıflandırma sonuçları	78
Şekil 4.9. ÇBLS yönteminin LVST veri kümesi için sınıflandırma sonuçları	80
Şekil 4.10. ÇBLS yönteminin MADELON veri kümesi için sınıflandırma sonuçları ...	81
Şekil 4.11. ÇBLS yönteminin RELAX veri kümesi için sınıflandırma sonuçları	82
Şekil 4.12. ÇBLS yönteminin SEEDS veri kümesi için sınıflandırma sonuçları	83
Şekil 4.13. ÇBLS yönteminin SONAR veri kümesi için sınıflandırma sonuçları.....	84
Şekil 4.14. ÇBLS yönteminin TELESCOPE veri kümesi için sınıflandırma sonuçları .	86
Şekil 4.15. ÇBLS yönteminin WINE veri kümesi için sınıflandırma sonuçları	87
Şekil 4.16. ÇBLS yönteminin WISCONSIN veri kümesi için sınıflandırma sonuçları .	88
Şekil 4.17. CMU-PIE için yüz doğrulama oranları.....	91
Şekil 4.18. ORL için yüz doğrulama oranları	92
Şekil 4.19. Yale B için yüz doğrulama oranları	93
Şekil 4.20. LVST veri kümesi için LDVM yönteminin sınıflandırma sonuçlarının karşılaştırması	97

Şekil	Sayfa
Şekil 4.21. SONAR veri kümesi için LDVM yönteminin sınıflandırma sonuçlarının karşılaştırması	98
Şekil 4.22. TELESCOPE veri kümesi için LDVM yönteminin sınıflandırma sonuçlarının karşılaştırması	99
Şekil 4.23. WISCONSIN veri kümesi için LDVM yönteminin sınıflandırma sonuçlarının karşılaştırması	100



SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

Açıklamalar

C	sınıf etiketi dizisi
c	ikili öznitelik grubu için sınıf etiketi
<i>g</i>	LÖS ile seçilen ikili öznitelik sayısı
<i>g_{opt}</i>	en iyi ikili öznitelik grup sayısı
k	en yakın komşu sayısı

Kısaltmalar

Açıklamalar

ÇBLS	Çok Boyutlu Lorentz Sınıflandırıcısı
DDA	Doğrusal Diskriminant Analizi
D-ÇBLS	Döndürme ile Çok Boyutlu Lorentz Sınıflandırıcısı
DVM	Destek Vektör Makinesi
D-LS	Döndürme ile Lorentz Sınıflandırıcısı
k-EYK	k En Yakın Komşu
LDVM	Lorentz Destek Vektör Makinesi
LÖS	Lorentz Öznitelik Seçme
LS	Lorentz Sınıflandırıcısı
LUTS	Lorentz Uzaklığı Temelli Sınıflandırma
LYD	Lorentz Yüz Doğrulama
OP-LS	Optimal Parametrelili Lorentz Sınıflandırıcısı
S-LS	Sıkıştırma ile Lorentz Sınıflandırıcısı

1. GİRİŞ

Günümüzde her alanda devasa boyutta veriler üretilmektedir. Bu verilerden anlamlı ve önemli olanlarını elde etmek için veri madenciliği gibi teknikler kullanılmaktadır [1]. Elde edilen anlamlı verilerden olumlu veya olumsuz sonuçlara ulaşmak için ve ortak özelliklerine göre farklı gruplara ayırma gibi önemli işlemler sınıflandırma yöntemleri ile yapılmaktadır [2]. Sınıflandırma yöntemleri veri analizi, yüz ve el yazısı tanıma, örüntü tanıma ve internet ortamında saldırı tespiti gibi birçok alanda yaygın olarak kullanılmaktadır [2, 3]. Sınıflandırma yöntemlerinde farklı uzaklık ölçütleri kullanılabilir. Bunlardan bazıları Öklid uzaklığı, Bhattacharyya metriği [1], Hamming uzaklık ölçütü [4], Mahalanobis uzaklığı [5], Kullback Leilber (KL) uzaklık ölçütü [6], Image To Class (I2C) uzaklık ölçütü [7]. Bu tez çalışmasında sınıflandırma probleminde farklı bir bakış açısıyla Lorentz metriğinin kullanılması incelenmiştir.

Lorentz uzayı matematikte uzaylar teorisi ve diferansiyel geometride yaygın kullanılan bir uzaydır [8]. Ancak mühendislik alanındaki uygulamaları üzerine yeterli çalışma yapılmamıştır. Bu nedenle literatürde mevcut olan Lorentz uzayı ile ilgili matematiksel bilgilerin yeni bir uygulama alanı olarak örüntü tanıma ve sınıflandırma gibi alanlarda nasıl kullanılabileceği incelenmeye değer bir problemidir.

Sınıflandırma ve uzay kavramları birbiriyle ilgili ifadelerdir. Farklı sınıftaki örneklerin farklılıklarını görebilmek için ayrı gruplara ayrılması gerekir; örneğin hasta veya sağlıklı, olumlu veya olumsuz. Veriler çok boyutlu ise çeşitli doğrusal veya doğrusal olmayan yöntemlerle daha küçük uzaylara taşınarak sınıflandırma gerçekleştirilir. Yeni uzaylara projeksiyon yapılırken orijinal örneklerin yapısı ve özellikleri korunabilir (izometrik taşıma). Sınıflandırmada karşılaşılan başka bir sorun da verilerin eğitimidir. Eğer veriler küçük boyutlu uzaylara taşınırken sistem yeni uzayın ölçümleri ve yapısına göre eğitilmemişse yeni gelen örnekler yanlış taşınabilir. Bu nedenle sınıflandırma işleminin doğru ve anlamlı yapılabilmesi için yeni uzayın metriği ve geometrisi çok önemlidir [9].

Lorentz uzayı ile ilgili literatürde çoğunlukla matematik ve fizik alanlarındaki bilim adamlarının çalışmaları bulunmaktadır. Lorentz uzayının fizikteki en çok bilinen uygulaması Einstein'ın Görecelik Kuramıdır [10]. Bu uzayın bir boyutu zamandır ve bu eksen diğer uzaysal eksenlere göre farklı bir özelliğe sahiptir; zaman eksenini vektörlerin iç

çarpımı tanımında negatif bir ağırlığa sahiptir. Bundan dolayı bu uzaydaki uzaklık ölçütü Öklid uzayından farklılık gösterir. Öklid uzayında noktaların her eksenindeki farkları uzaklık formülüne pozitif katkı sağlarken, Lorentz uzayında zaman eksenini negatif bir katkı sağlar. Bu şu anlama gelir; Lorentz uzayında farklı koordinatlara sahip 2 farklı noktanın arasındaki Lorentz uzaklığı sıfır çıkabilir ve hatta (karekök alınmadığında) negatif olabilir. Farklı koordinatlardaki noktaların tanımsal olarak sıfır uzaklığına sahip olabilmeleri oldukça ilginç bir durumdur. Bu bilgiden yola çıkarak Lorentz uzayının sınıflandırma problemlerinde önemli bir uygulama potansiyeli olduğu düşünülmüştür.

Lorentz uzayının örüntü tanıma problemlerinde kullanılması ile ilgili literatürde sadece bir grup tarafından hazırlanmış birkaç bilimsel yayın bulunmaktadır; bu makale ve bildirilerde Lorentz manifoldundan yararlanılarak boyut indirgeme problemi ele alınmıştır. Ayrıca literatürde verileri Lorentz veya benzeri uzaylara taşıma, aktarma ve boyut indirgeme işlemlerinde zeki optimizasyon tekniklerinin kullanılabilme konusu açıktır.

Bu tezin amacı, fizik ve matematik alanlarında kullanılan Lorentz uzayının sınıflandırma problemlerindeki potansiyel uygulama imkânlarının ve getireceği avantajların incelenmesidir. Bu tezde Lorentz uzayının sınıflandırma probleminde en başarılı bir şekilde nasıl kullanılacağı araştırılmış ve uygun metodolojiler geliştirilmiştir.

Lorentz uzayının tüm matematiksel özellikleri, bu uzaya ilişkin tüm kuramsal bilgiler ve yapısı [8] numaralı kaynakta detaylıca açıklanmıştır. Lorentz uzayının Riemann ve semi-Riemann uzayları ile karşılaştırılması ve Lorentz uzayının matematiksel yapısının net olarak tanımı [11] numaralı kaynakta verilmiştir. Lorentz geometrisi ile ilgili temel kaynaklar olarak [12, 13] araştırmaları da bulunmaktadır. Riemann ve Lorentz uzaylarının uzaklık, metrik ve metrik imzası gibi önemli bileşenlerinin detaylıca açıklanması [13] numaralı kaynakta verilmiştir. Bu araştırmada ayrıca Genel Görecelik kuramında kullanılan diğer metrik ve uzaklıklar da yer almaktadır. Riemann geometrisinden Lorentz geometrisine geçiş şeklinde iki uzay arasındaki farklılıkları detaylı bir şekilde karşılaştırmalı olarak açıklaması [14] numaralı kaynakta anlatılmıştır. Yazarlar Genel Görecelik Kuramındaki bazı önemli problemlere değinerek ve Lorentz uzayının özellikleri bakımından yeni fikirler sunmuşlardır. Lorentz uzayı ile ilgili diğer bir teorik kaynakta [8] Afine ve Minkowski uzaylarının özellikleri incelenerek Lorentz uzayı ile ilişkileri araştırılmıştır. Lorentz uzayındaki dönme ve dönüştürme işlemleri matematiksel ifadelerle gösterilmiştir [11].

Lorentz uzayının sınıflandırma problemlerinde kullanılıp kullanılamayacağı üzerine yapılan akademik bilimsel yayınları tarama neticesinde sınırlı sayıdaki bilimsel çalışmaya rastlanmıştır. Bu araştırmalar bir grup araştırmacılar tarafından yapılmıştır. Onlardan ilkinde diskriminant analizi temelinde sınıflandırma problemi için Lorentz metrik tensörleri ile boyut indirgeme yapan LDP (*Lorentz Discriminant Projection*) yöntemini önermişlerdir [15]. Önerdikleri yaklaşımla orijinal uzaydaki örneklerin yapısını manifold ile modellemişlerdir. Önerilen yöntemde veri kümelerinin yerel ve global yapısını modellemek amacıyla her bir örneğin aynı sınıfa ait diğer örnekler ile arasındaki uzaklık ve global geometrik merkez noktası ile arasındaki uzaklık hesaplanmaktadır. Bu adım önerilen yöntemin literatürdeki diskriminant analizi yapan yöntemlerden farkını ortaya koymaktadır. Böylece veri kümesindeki örneklerin sınıflar kümesi ve global yapı geometrisi Lorentz metrik tensör ile eğitilebilmektedir. Yüz tanıma ve el yazısı veri kümeleri üzerinde yapılan sınıflandırma denemelerinde bu yaklaşımın başarılı olduğu vurgulanmıştır. Aynı ekip daha sonra gerçekleştirdikleri çalışmalarında ise Lorentz metrik tensörleri ile öznitelik çıkarımı üzerine kernel, tensör ve smooth LDP ile yaptıkları çalışmayı genişletmişlerdir [16].

Lorentz uzayının sınıflandırmada kullanımı ile ilgili diğer bir çalışmada, yazarlar denetimli öğrenme algoritmalarının yerine sınıf etiketi belli olmayan örneklerin yerel yapısını koruyan yarı-denetimli algoritmaların kullanılmasının daha isabetli olacağını vurgulayarak yeni bir yaklaşım geliştirmişlerdir [17]. Yazarların önerdikleri yöntemde yarı gözetimli boyut indirgeme işlemi özel Lorentz metrik tensörü ile yapılmaktadır. Yazarlar Lorentz metrik tensörü ile yapılan eğitimde etiketi belli olmayan örneklerin hem yerel hem de global geometrik yapısının korunduğunu savunmaktadırlar. Deneyleri yüz veri kümeleri ile gerçekleştirmişlerdir. Böylece, yazarlar önerdikleri özel Lorentz matrisi ile gözetimli algoritmaları yarı gözetimli hale dönüştürmüşlerdir. Araştırmada sunulan sınıflandırma sonuçlarına göre önerilen yaklaşımın yüz tanımda başarılı bir şekilde kullanıldığı görülmektedir.

Lorentz uzayının görsel izleme problemindeki uygulanmasını anlatan diğer bir çalışmada yazarlar yeni Incremental Lorentz Discriminant Projection (ILDLP) yöntemini önermişlerdir. Yazarlar görsel izleme probleminde Lorentzian Discriminant Projection yöntemini kullanarak Lorentz uzayında verilerin boyutunu artırımlı bir şekilde indirmişlerdir. Makalede geliştirilen yaklaşımın büyük varyanslı hareketleri başarılı bir şekilde tespit edebildiği görülmüştür [18].

Riemann ve Lorentz uzayları semi-Riemann uzayının alt uzayları olduğu için benzer özellikleri taşımaktadır. Bu nedenle Riemann geometrisini kullanarak yapılan akademik çalışmalar da incelenmiştir. Riemann uzayı Lorentz uzayı ile bazı aynı özellikleri taşımasından dolayı bu tez araştırması için önemlidir. Riemann uzayını kullanarak yüz tanıma, sınıflandırma ve benzeri problemleri inceleyen birkaç akademik makale mevcuttur.

Örneğin bu çalışmaların birinde yazarlar boyut indirgeme problemini Riemann geometrisinin klasik problemi kabul ederek modellemişler, yani Riemann manifoldunun normal koordinatlarını oluşturmayı ele almışlardır. Bunun için de Riemann geometrisinde sık kullanılan normal koordinatları karışık verileri örneklemek için kullanmışlardır. Neticede k komşular sayısı ve d indirgenecek boyut şeklinde iki parametreyi kullanarak manifoldun koordinatlarını hesaplamışlardır [19].

Diğer bir çalışmada ise yazarlar Semi-Riemann uzayının yerel metrikleri ile orijinal örneklerin sınıf içi ve sınıflar arası dağılım ilişkisini yapısını modellemişlerdir [20]. Böylece sınıfların geometrik yapısını oluşturmuşlar ve bu yapıyı yeni uzayda optimize etmek için semi-Riemann uzayındaki metrik tensörlerin karesel değerlerini maksimize etmişlerdir. Bu modelleme stratejisini kullanarak yazarlar alt uzaylar temelinde sınıflandırma yapan SRDA (*Semi-Riemannian Discriminant Analysis*) yöntemini geliştirmişlerdir [20]. Önerilen yöntemin yüz tanıma ve el yazısı veri kümeleri üzerinde yapılan sınıflandırma denemelerinde başarılı sonuç ürettiği görülmüştür. Deneylerde elde edilen sayısal sonuçlara göre yüz tanıma ve sınıflandırma probleminde önerilen yöntemin diğer yöntemlere göre daha başarılı olduğu sonucuna varılmıştır.

Riemann uzayının sınıflandırmada kullanımı ile ilgili diğer bir çalışmada yazarlar yarı-Öklid manifoldu üzerinde boyut indirgeme yapan S^3RMM (*Semisupervised Semi-Riemannian Metric Map*) yöntemi ile yarı gözetimli sınıflandırma algoritmasını geliştirmişlerdir [21]. Yazarların önerdikleri S^3RMM yaklaşımı semi-Riemann metrik tensörü ile modellenen ayırt edicilik ve benzerlik ölçümlerini maksimize etmektedir. Böylece, her bir örneğin metrik tensörü yarı gözetimli regresyon ile elde edilmektedir. Sonuç olarak çalışmada, yüz ve el yazısı veri kümeleri ile yapılan denemelerde elde edilen sayısal sonuçlara göre yarı gözetimli öznetelik çıkarımı için S^3RMM yönteminin başarılı bir yaklaşım olduğu vurgulanmıştır [21]. Diğer bir çalışmada Riemann geometrisini kullanarak elektroensefalografi sinyallerini sınıflandırarak hastanın uyku durumunu belirlemek amaçlı bir bilimsel araştırma

yapmışlardır. Sinyalleri birbirinden farklı sınıflara ayırmak için ayırt edicilik ölçütü olarak Riemann uzaklık ölçütünü kullanmışlardır [22].

Başka bir çalışmada [23] yazarlar elektroansefalografi sinyallerini sınıflandırarak beyin-bilgisayar ara yüzlerini iyileştirmeye çalışmışlardır. Bu çalışmada iki yöntem geliştirilmiştir. Birincisi, Riemann geodezik uzaklık ölçütünü kullanarak Riemann değerini oluşturan pozitif tanımlı simetrik matrisini tanımlamışlar ve Riemann Değeri için Minimum Uzaklık seçen algoritmayı geliştirmişlerdir. İkincisi Riemann Tanjant Uzayını kullanarak sinyalleri sınıflandıran Tanjant Uzay Haritalama yöntemini geliştirmişlerdir. Bu yöntemde sinyallerin kovaryans matrislerini Riemann tanjant uzayına taşıyarak işlem yapılmaktadır.

Literatürdeki akademik yayınların incelemesi sonucunda Lorentz uzaklık ölçütünden yararlanılarak sınıflandırma yapma konusunda şimdiye kadar hiçbir bilimsel çalışma yapılmadığı tespit edilmiştir. Bu nedenle seçilen konu özgünlüğünü ve önemini korumaktadır. Bu tezin amacı Lorentz uzayının kendine özgü karakteristiklerinden yararlanılarak veri kümelerinin daha iyi sınıflandırılması için yeni bir veri gösterim şeklinin, yeni bir uzaklık ölçütünün ve yeni bir sınıflandırma sürecinin tanımlanmasıdır. Bu amaçla ilk önce veriler Öklid uzayında bir ön işleme aşamasından geçirilmektedir. Bu aşamada, verilerin Lorentz uzayına en uygun bir şekilde aktarılmasına hazırlık yapılmaktadır.

İkinci aşamada, sınıfları birbirinden en fazla ayırt etmek üzere verilerin en uygun şekilde Öklid uzayından Lorentz uzayına aktarılması hedeflenmektedir. Üçüncü aşamada ise, Lorentz uzayında gösterilen verilerin Lorentz uzaklık ölçütünden yararlanılarak sınıflandırılması yapılmaktadır.



2. LORENTZ UZAYI

2.1. Lorentz Metriği

Genel olarak Genel Görecelik Kuramında kullanılan Lorentz uzayı Öklid dışı uzay olarak da bilinmektedir [10]. Lorentz uzayında iç çarpım Öklid uzayındaki iç çarpımdan farklıdır ve pozitif tanımlılığı sağlamayacak şekildedir. Lorentz uzayında iç çarpım $X = x_1, x_2, \dots, x_n$ ve $Y = y_1, y_2, \dots, y_n \in \mathbb{R}^n$ olmak üzere Eş. 2.1'de verilmiştir.

$$(X, Y) \rightarrow \langle X, Y \rangle_{I_L} = \sum_{i=1}^{n-1} x_i y_i - x_n y_n \quad (2.1)$$

Bu eşitlik ile tanımlanan iç çarpım $\mathbb{R}^n$ üzerinde bilineer, simetrik, non-dejenere metrik tensördür ve Lorentz metriği olarak tanımlanmaktadır [9].

Lorentz uzayı matematiksel olarak semi-Riemann uzayının alt kümesi olarak ifade edilir [10, 12]. Riemann uzayının genellemesi semi-Riemann uzayıdır. Semi-Riemann uzayındaki metrik diyagonal matrisi olarak tanımlıdır. Diyagonal metrik matrisinin elemanları ise sıfıra eşit değildir [12]. Semi-Riemann uzayında p pozitif, q negatif olmak üzere metrik $p + q = n$ olacak şekilde diyagonal elemana sahip ise metrik imza (p, q) 'dir. M semi-Riemann uzayı g metrik tensör ile donatılmış bir uzaydır [12]. M uzayının *indeksi (metrik imzadaki negatif değerler sayısı)* s ($0 \leq s \leq \dim M$) değerini tanımlarsak ve eğer:

- $s = 0$ ise Riemann uzayı, pozitif tanımlı, simetrik
- $s = 1$ ise Lorentz uzayı olur, pozitif tanımlı değil, simetrik, dejenere olmayan
- $s = \frac{1}{2} \dim M$ ise tarafsız metriğe sahip olan semi-Riemann uzayıdır, simetrik, pozitif tanımlı olmayan, dejenere olmayan [11, 12].

Lorentz uzayındaki metrik matrisi tek negatif elemana sahip olmasından dolayı metriğin işareti $(n - 1, 1)$ 'dir. Böylece, Lorentz uzayı için tanımlanan metrik matrisi Eş. 2.2'de verilmiştir [13].

$$G = \begin{pmatrix} \hat{\Lambda}_{(n-1) \times (n-1)} & 0 \\ 0 & -\check{\lambda} \end{pmatrix} \quad (2.2)$$

Buradaki $\hat{\Lambda}$ - pozitif ve $\check{\lambda}$ - negatif öz değerleri ifade etmektedir. $g(r, r)$ metrik tensörü Eş. 2.3 ile tanımlanmaktadır.

$$l(\gamma) = \int_0^1 \sqrt{-\left\langle \frac{d\gamma}{dt}, \frac{d\gamma}{dt} \right\rangle} dt \quad (2.4)$$

Uzaysal bir eğri için $l(\gamma) = 0$. Böylece $p, q \in M$ iki nokta arasındaki Lorentz uzaklığı, eğer $p < q$ ise, yani *geleceğe yönelmiş uzaysal olmayan* eğriler kümesi Γ boş değilse Eş. 2.5 ile hesaplanabilir.

$$d(p, q) = \sup_{\gamma \in \Gamma} l(\gamma) \quad (2.5)$$

Diğer durumlarda Lorentz uzaklığı sıfıra eşittir [24].

2.2. Lorentz Uzayının Özellikleri

Lorentz uzayının Eş. 2.1 ile verilen metriğine göre iki (X ve Y) nokta arasındaki Lorentz uzaklığı d Eş. 2.6 ile hesaplanmaktadır [4, 11]:

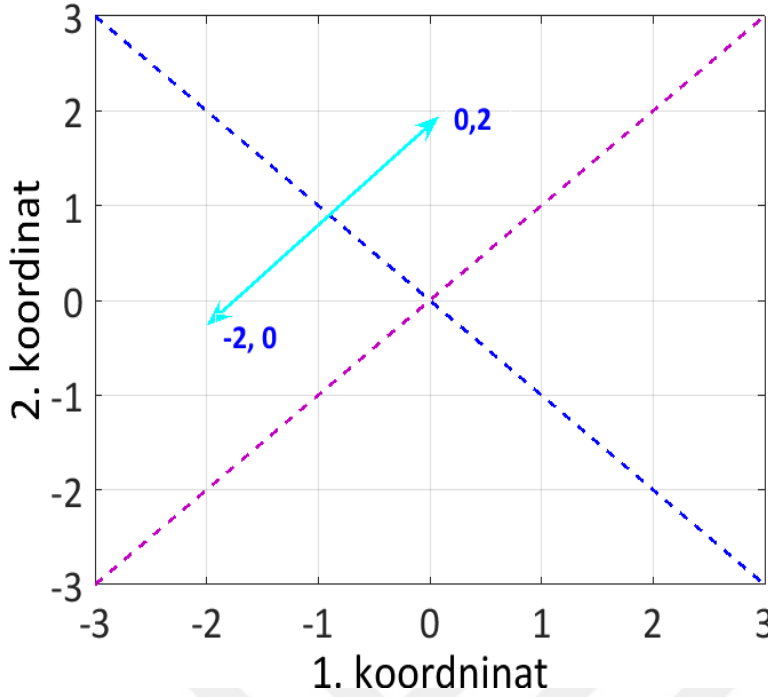
$$d_L(X, Y) = \sqrt{\sum_{i=1}^{n-1} |x_i - y_i|^2 - |x_n - y_n|^2} \quad (2.6)$$

Buradaki, n Lorentz uzayının boyutudur. Sonuncusu Lorentz metriğine göre negatif imzalıdır, yani eksi işaretlidir. Bu durumu örneklemek için sırasıyla Öklid ve Lorentz uzayındaki $(-2, 0)$ ve $(0, 2)$ iki nokta arasındaki Öklid ve Lorentz uzaklıkları sırasıyla Eş. 2.7 ve Eş. 2.8 ile verilmiştir.

$$d_0 = \sqrt{(-2 - 0)^2 + (0 - 2)^2} = 2\sqrt{2} \quad (2.7)$$

$$d_L = \sqrt{(-2 - 0)^2 - (0 - 2)^2} = \sqrt{0} \quad (2.8)$$

Lorentz ve Öklid metriklerini kullanarak iki nokta arasındaki uzaklık hesaplamalarının karşılaştırılması görsel olarak Şekil 2.2’de yer almaktadır.



Şekil 2.2. Lorentz ve Öklid metriklerinin karşılaştırması

Eş. 2.8’den görüldüğü gibi, Lorentz metriğine göre iki nokta arasındaki uzaklık sıfır veya sıfıra yakın değer çıkabilmektedir. Böylece, Lorentz uzayının sahip olduğu özel metriğinden dolayı Lorentz uzaklığını sınıflandırma probleminde en yakın komşu tekniği ile örnekleri bir birinden ayırt etmede ölçüt olarak kullanmanın doğru olacağı görülmektedir.

Bununla birlikte, Lorentz uzayının sınıflandırma probleminde kullanımında yararlı olabilecek diğer özellikleri de incelenerek, analiz edilmiştir. Bu bağlamda, Lorentz uzayında bu uzayın sahip olduğu özel yapısı ve metriğinden dolayı matris çarpımı işleminin Öklid uzayından farklı olduğu görülmüştür [25, 26]. Matris çarpımı işlemi uzaylar arası örneklerin taşınması ve bazı sınıflandırıcıların içerdiği kernel fonksiyonlarda sık kullanılan bir işlemdir.

Lorentz uzayında iki $X = [x_{ij}] \in \mathbb{R}_u^t$ ve $Y = [y_{jl}] \in \mathbb{R}_p^u$ matrislerin çarpma işlemi Eş. 2.9 ile hesaplanmaktadır [25]:

$$X \cdot_L Y = \left[\sum_{j=1}^{u-1} x_{ij}y_{jl} - x_{iu}y_{ul} \right] \quad (2.9)$$

Buradaki, ‘ $\cdot_L$ ’ notasyonu Lorentz uzayındaki matris çarpımını ifade etmektedir. Örneğin, Lorentz uzayındaki 2×2 boyuttaki iki X, Y matrislerin çarpımı Eş. 2.10 ile verilmiştir.

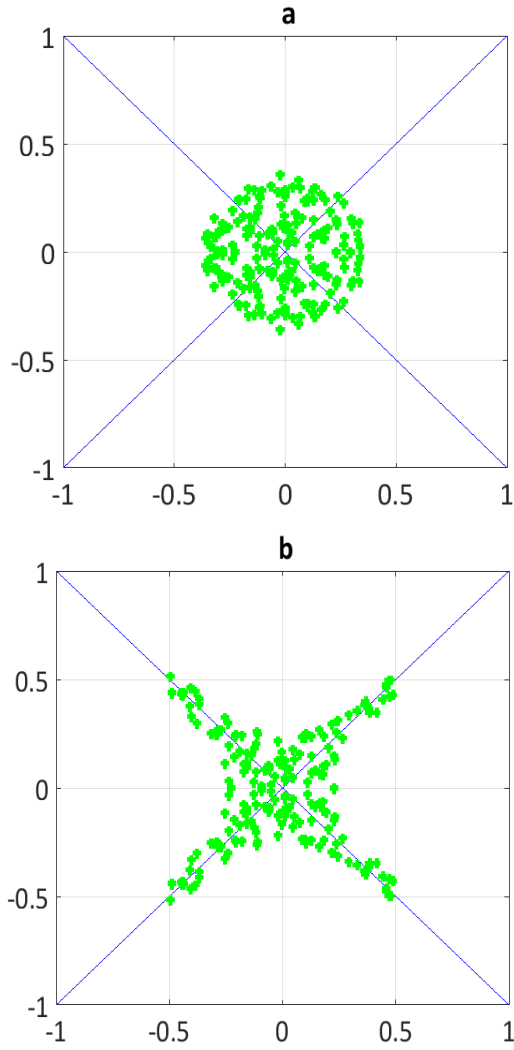
$$X \cdot_L Y = \begin{bmatrix} x_{11}y_{11} - x_{12}y_{21} & x_{11}y_{12} - x_{12}y_{22} \\ x_{21}y_{11} - x_{22}y_{21} & x_{21}y_{12} - x_{22}y_{22} \end{bmatrix} \quad (2.10)$$

2.3. Lorentz Uzayında Noktalar Arası Komşuluk Tanımı

Genel olarak, uzaklık ölçütü sınıflandırma probleminde bir örneğe en yakın uzaklıktaki örnekleri belirlemede sık kullanılan kavramdır. Bu bağlamda, ilk olarak Lorentz uzayının yapısı ve metriği hesaba katılarak noktalar arasındaki komşuluk ilişkisi tanımlanmıştır. Lorentz uzayında komşuluk kavramının tanımlanabilmesi için önce bu uzayda herhangi iki noktanın arasındaki uzaklık kavramının iyi anlaşılması gerekmektedir. Matematikten bilindiği gibi her uzayda uzaklık hesaplanabilmesi için metrik tanımlı olması gerekmektedir [27]. Lorentz uzayının metriği Eş. 2.1'den görüldüğü gibi Öklid'e göre farklıdır. Bu nedenle, Lorentz uzaklığı d_L da Eş. 2.6'dan görüldüğü gibi literatürde geniş tanınan ve sık kullanılan Öklid uzaklığından farklıdır. Lorentz uzaklığının Öklid uzaklığından en önemli farkı da, iki nokta arasındaki uzaklığın sıfır çıkabilmesidir.

Lorentz uzayının genelinde koni çizgilerine, yani Şekil 2.1'de gösterilen null eksenine paralel olan tüm doğrular üzerindeki noktalar arası uzaklık Lorentz uzaklığına göre sıfırdır. Bu yapısal özelliğinden dolayı Lorentz uzayında iki noktayı birleştiren doğru null eksenine göre açısı sıfır veya sıfıra yakın ise, bu iki nokta arasında uzaklık sıfır veya sıfıra yakın değere sahip olur, yani iki nokta birbirine yakın ve komşudur. Bu teorik bilgiyi görsel olarak canlandırmak adına Gauss dağılımına göre rasgele 500 nokta oluşturulmuştur. Sonra (0,0) orijin noktasına en yakın 100 nokta Öklid ve Lorentz metriklerine göre seçilmiştir. Elde edilen komşu noktalarının dağılımları görsel olarak Şekil 2.3'te yer almaktadır.

Şekil 2.3'ten görüldüğü gibi Öklid uzayında (0,0) noktasının komşu noktaları çember oluştururken, Lorentz metriğine göre seçilen noktalar null eksenini etrafına dağılmış durumdadır.



Şekil 2.3. Öklid uzayında (a) ve Lorentz uzayında (b) komşu noktaların dağılımı

2.4. Uygulamada Kullanılan Veri Kümeleri

Bu araştırmada, Lorentz uzayının sınıflandırmada kullanımı konusu araştırılmıştır ve bu bağlamda geliştirilen sınıflandırma algoritmalarının doğruluğunu ve geçerliliğini test etmek amacıyla genel kullanıma açık olan CLIMATE, ECOLI, GESTURE, ILPD, LVST, MADELON, PARKINSON, RELAX, SEEDS, SONAR, TELESCOPE, VERTEBRAL, WINE, WISCONSIN veri kümeleri seçilmiştir [28]. Bu veri kümelerinde öznitelik sayısı 6-500 arasında değişmektedir ve genellikle iki sınıflı veri kümeleri tercih edilmiştir. Denemelerde veri kümelerindeki örneklerin %30 eğitim ve kalan kısmı test için kullanılmıştır. Uygulamada kullanılan veri kümeleri ile ilgili bazı istatistikler Çizelge 2.1’de verilmiştir.

Çizelge 2.1. Veri kümeleri istatistiği

	Öznitelik sayısı	Sınıf sayısı	Örnek sayısı	Eğitim	Test
CLIMATE	8	2	540	120	420
ECOLI	7	2	220	80	140
GESTURE	18	2	448	150	298
ILPD	10	2	263	140	223
LVST	310	2	127	42	85
MADELON	500	2	200	66	134
PARKINSON	29	2	1040	200	60
RELAX	12	2	182	60	122
SEEDS	7	5	140	46	94
SONAR	60	2	208	69	139
TELESCOPE	10	2	400	134	266
VERTEBRAL	6	3	310	110	200
WINE	13	2	130	44	86
WISCONSIN	33	2	198	66	132

Yüz veri kümeleri

Algoritma-10 ile verilen Lorentz Yüz Doğrulama (LYD) yöntemini test etmek amacıyla seçilen CMU PIE, ORL ve Yale B yüz veri tabanlarındaki yüz resimleri 22x22 boyutundadırlar. Yüz veri tabanlarındaki resimler piksellerden oluştuğu için ve her bir pikselin ayrı bir renk seviyesini belirttiği için toplamda 484 özniteliğe sahip örnek olarak ele alınmaktadır. Denemede kullanılan yüz veri tabanları ile ilgili sayısal istatistikler Çizelge 2.2’de yer almaktadır.

Çizelge 2.2. Yüz veri tabanları istatistiği

	Öznitelik sayısı	Sınıf sayısı	Örnek sayısı	Eğitim	Test
CMU-PIE	484	60	11554	180	240
ORL	484	40	400	120	160
Yale B	484	38	2414	117	190

Lorentz Yüz Doğrulama algoritmasını test etmek amacıyla kullanılan Yale B, ORL ve CMU PIE yüz veri tabanlarından alınan örnek yüz görüntüleri aşağıdaki Şekil 2.4, Şekil 2.5 ve Şekil 2.6’da sırasıyla görüntülenmiştir.



Şekil 2.4. Yale B yüz veri tabanından örnek yüz görüntüleri



Şekil 2.5. ORL yüz veri tabanından örnek yüz görüntüleri



Şekil 2.6. CMU PIE yüz veri tabanından örnek yüz görüntüleri



3. LORENTZ UZAYININ SINIFLANDIRMADA KULLANIMI

Bu bölümde Lorentz uzayının sınıflandırma probleminde kullanımını gerçekleştirmek adına iki boyutlu ve çok boyutlu veri kümeleri için geliştirilen Lorentz Sınıflandırıcısı (LS) ve Lorentz metriğine göre yapılandırılmış Lorentz Destek Vektör Makinesi (LDVM) ile ilgili bilgiler verilecektir. Ayrıca, Lorentz uzayında sınıflandırma başarısını arttırmak amacıyla önerilen önişlem adımları, Lorentz metriği temelli ayırt etme ölçütü ve öznelik seçme yöntemi anlatılacaktır.

3.1. Önişlem

Lorentz uzayının özel yapısını ve özelliklerini göz önünde bulundurarak, bu uzayın sınıflandırma probleminde kullanımında ve başarı oranının artırımında yardımcı olabilecek Lorentz uzayına has önişlem ve veri gösterimi yöntemleri araştırılmıştır. Bu araştırma sonucunda bazı temel işlemlerden oluşan önişlem aşaması önerilmiştir. Bu işlemler matematikte ve cebirde çok bilinen öteleme, sıkıştırma ve döndürme yöntemleridir [27].

3.1.1. Öteleme, sıkıştırma ve döndürme

Bu araştırmada, Lorentz uzayının yapısal ve Bölüm 2.3. te tanımlanan noktalar arası özel komşuluk özelliklerinden en iyi bir şekilde istifade edebilmek için örnekler Önişlem 1'deki adımlardan geçirilerek Lorentz uzayına taşınmıştır.

Önişlem 1.

Girdi: X eğitim kümesi

Adım 1. Eğitim kümesindeki her bir sınıfın μ_i ortalaması bulunmaktadır;

Adım 2. Her sınıfın ortalaması μ_i orijine gelecek şekilde veriler ötelenmektedir;

Adım 3. Veri kümesi diyagonal λ matrisi ile çarpılarak sıkıştırılmaktadır;

Adım 4. Veri kümesi 45° açıyla döndürülmekte ve her bir sınıfın verisi ortalama noktasına göre başlangıçtaki yerine geri ötelenmektedir.

Önişlem-1' in ilk adımında bir sınıfın ortalaması (μ) literatürde çok iyi bilinen Eş. 3.1 ile hesaplanmaktadır [27].

$$\mu = \frac{1}{N} \sum_{i=1}^N A_i \quad (3.1)$$

Buradaki A bir sınıftaki veriler, N ise sınıftaki örnek sayısıdır.

2. adımdaki veri kümesinin ötelenmesi matematiksel olarak çok basit bir işlemdir ve A vektörünün δ kadar ötelenmesi için bunların toplamını almak yeterli olmaktadır. Noktaların ötelenmesini sağlayan operatör Eş. 3.2’de verilmiştir.

$$T_{\delta}f(A) = f(A + \delta), \quad (3.2)$$

3.adımdaki sıkıştırma işlemi projeksiyon almak için kullanılmıştır. Burada matris teorisinde mevcut olan diyagonal matrisin özelliğinden yararlanılmıştır [26]. Diyagonal matris matematiksel olarak $\lambda = 0, if i \neq j \forall i, j \in \{1, 2, \dots, n\}$ şeklinde ifade edilebilir [26] ve iki boyutlu veriler için sıkıştırma operatörü Eş. 3.3’te verilmiştir.

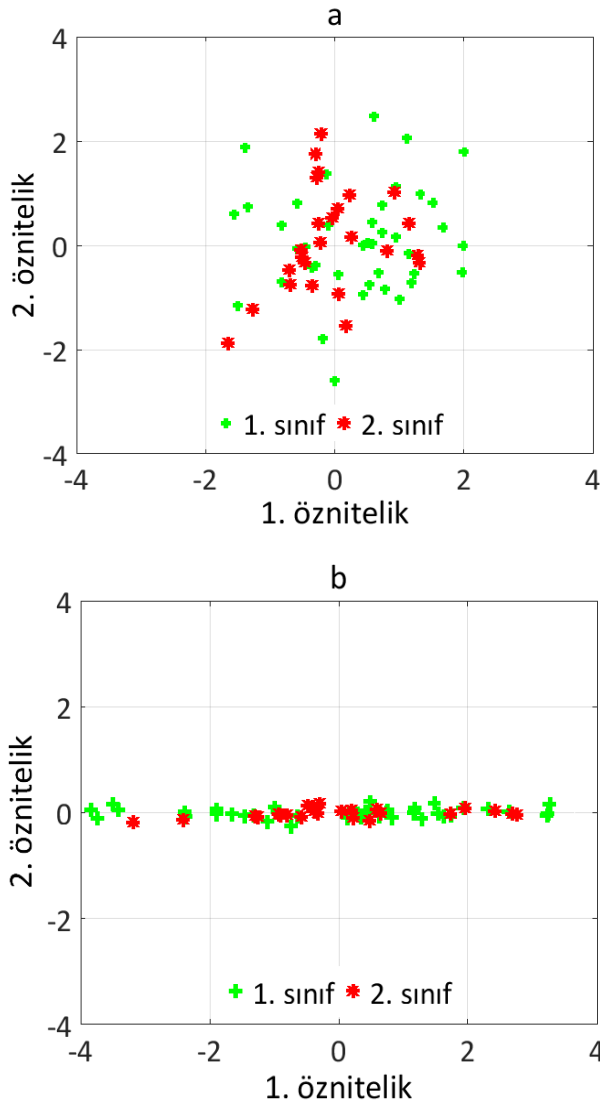
$$\lambda = \begin{pmatrix} w & 0 \\ 0 & q \end{pmatrix} \quad (3.3)$$

Buradaki $w, q \in R$ reel sayılardır. Bu yöntemin kullanılma sebebi ise verilerin ortalama dan geçen ve null eksenine paralel olan doğru etrafında dağıtılmasını sağlamaktır. Uygulamalarda λ matrisinin w, q parametreleri $w = 20 * q$ şeklindeki bir oranda kullanılmıştır ve genel durum $\lambda = \begin{pmatrix} 2 & 0 \\ 0 & 0.1 \end{pmatrix}$ olarak tanımlanmıştır. Verilerin λ matrisiyle çarpılarak sıkıştırma işlemi sınıfların μ geometrik ortalamalarına bağlı olarak yapılmaktadır. Şekil 3.1’de iki sınıflı veri kümesini Eş. 3.3 ile verilmiş olan λ matrisi ile çarpılarak gerçekleştirilen sıkıştırma işleminin görselleştirilmesi verilmiştir.

Ayrıca literatürde çeşitli veri kümeleri mevcuttur ve onların dağılım şekilleri de farklılık göstermektedir. Dolayısıyla dağılıma göre Eş. 3.4 ile verilmiş olan dikey sıkıştırma matrisi de gerektiğinde kullanılabilir.

$$\lambda = \begin{pmatrix} 0 & w \\ q & 0 \end{pmatrix} \quad (3.4)$$

Ön işlemenin son adımında veriler matematikte yaygın kullanılan ve Eş. 3.5’te verilmiş olan döndürme matrisi ile 45° açığa döndürülmektedirler ve ortalamasına göre başlangıç yerlerine geri ötelenmektedirler.

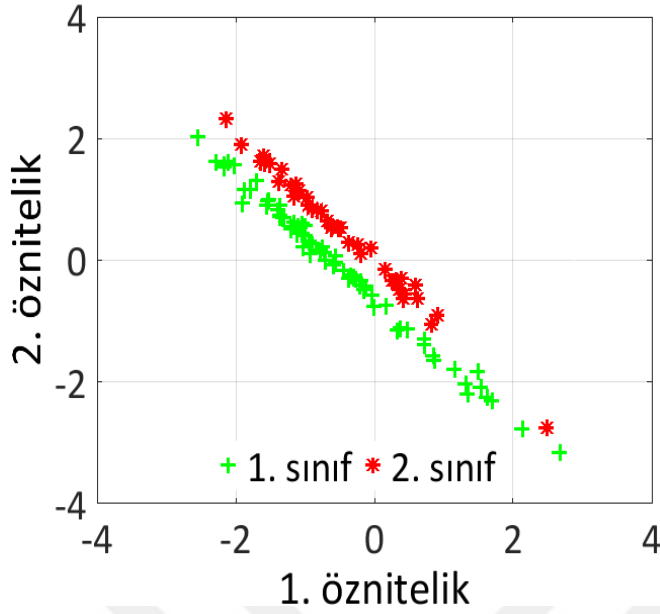


Şekil 3.1. İki sınıflı Gauss dağılımı için: a) orijinal hali, b) sınıfların λ matrisi ile sıkıştırılmış hali

Eş. 3.5 ile döndürme işleminin görselleştirilmesi Şekil 3.2’de verilmiştir.

$$R = \begin{pmatrix} \cos(45^\circ) & \sin(45^\circ) \\ -\sin(45^\circ) & \cos(45^\circ) \end{pmatrix} \quad (3.5)$$

Döndürme işlemi orijin etrafında tanımlı olduğu için bu sürecin bir parçası olarak 2. ve son adımdaki ötele işlemleri yapılmaktadır. Sıkıştırma ve 45° döndürme işlemlerinin sonucunda Öklid uzayında Gauss şeklinde dağılan noktalar Lorentz uzayında çarpı ($\times$) şeklindeki hiperbolik dağılımın bir kanadı olacak şekle dönüşmektedir. Böylece, Şekil 2.3. b deki şekle uygun bir dağılım elde edilmektedir. Bu süreci, “Öklid uzayından Lorentz uzayına verilerin taşınması” olarak adlandırabiliriz.



Şekil 3.2. Döndürme işleminin sonucu

3.1.2. DVM ve DDA karar çizgilerinin kullanımı

Lorentz uzayında sınıflandırma başarısını arttıracak olan önişlem veya verilerin uygun hale getirilmesi Lorentz metriği açısından önemlidir. Çünkü örneklerin Lorentz uzayındaki dağılımı 2.3. bölümde beyan edilen komşuluk kavramına göre sınıflandırma başarısını doğrudan etkilemektedir. Bu bağlamda örnekleri en optimal ve anlamlı bir şekilde Lorentz metriğine göre uygun hale getirerek Lorentz uzayında sınıflandırmak amacıyla Destek Vektör Makinesi ve Doğrusal Diskriminant Analizi yöntemlerinin karar çizgilerinin kullanılması araştırılmıştır.

Destek Vektör Makinesi

DVM yöntemi sınıflandırıcı olarak örnekleri farklı sınıflara ayırt etmek için destek vektörlerinden yararlanarak hiperdüzlem oluşturmaktadır [29, 30]. Bu hiperdüzlem iki boyutlu veriler için iki sınıfı en iyi şekilde ayırt eden karar çizgisi şeklinde ele alınmaktadır [31]. DVM ile hesaplanan karar çizgisini kullanarak verilerin Lorentz uzayına taşınmasında ilk olarak Lorentz uzayının özellikleri göz önünde bulundurularak, örneklerin orijinal dağılım şekli ile Lorentz uzayının arasında mantıksal bağ oluşturulmaya çalışılmıştır. Lorentz uzayının yapısı ve özelliklerinin detaylı inceleme sonucunda bazı tespitler ortaya çıkmıştır. Bu tespitler sırasıyla:

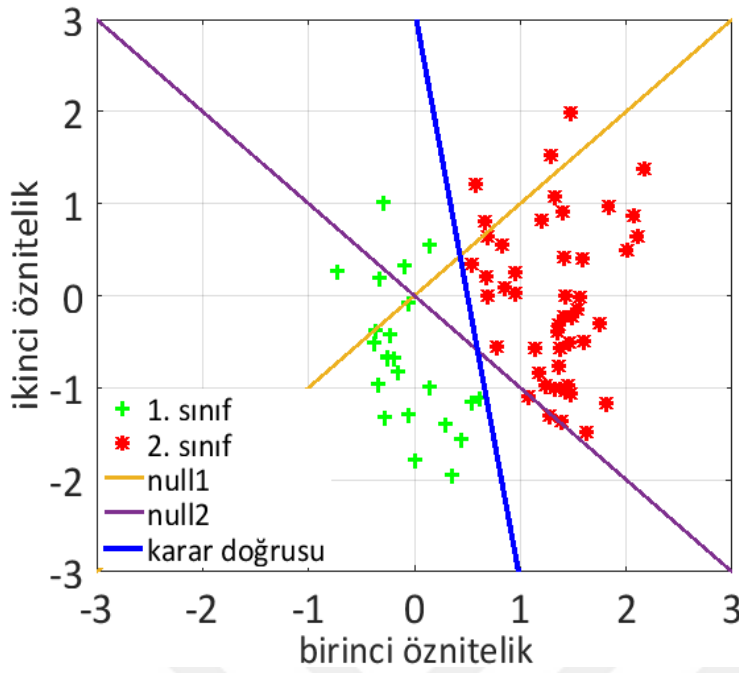
1. Sınıflandırılacak olan veri kümesinin Lorentz uzayına taşınmadan önceki dağılım şekli sınıflandırma başarısı açısından önemlidir. Bunun sebebi Lorentz uzaklık ölçütüdür.
2. 2.3. bölümde ortaya koyulan Lorentz uzayındaki noktalar arası komşuluk kavramına göre sınıf içi toplam uzaklığın azaltılması için noktalar ile sınıf ortalaması arasındaki doğruların null eksenine paralel olması lazım.
3. Sınıflar arası uzaklığın artması için farklı sınıflara ait örnekler null ekseninin iki tarafında kalmalıdır.

Son iki madde sınıflandırma probleminin çözümü oluşturan ana formüllerdir. Bu tespitlerden yola çıkarak, farklı sınıflara ait olan örneklerin null eksenine olan pozisyonunu ve en önemlisi de iki sınıfı birbirinden en doğru bir şekilde ayıran karar çizgisini belirlemek için DVM karar çizgisinin kullanımı anlamlı işlem olacağı ve örneklerin Lorentz uzayında sınıflandırılmasında başarıyı olumlu etkileyeceği düşünülmüştür. Çünkü DVM yöntemi farklı sınıflara ait örnekleri ayırt etmede benzer karar doğruları kullanmaktadır [32, 33].

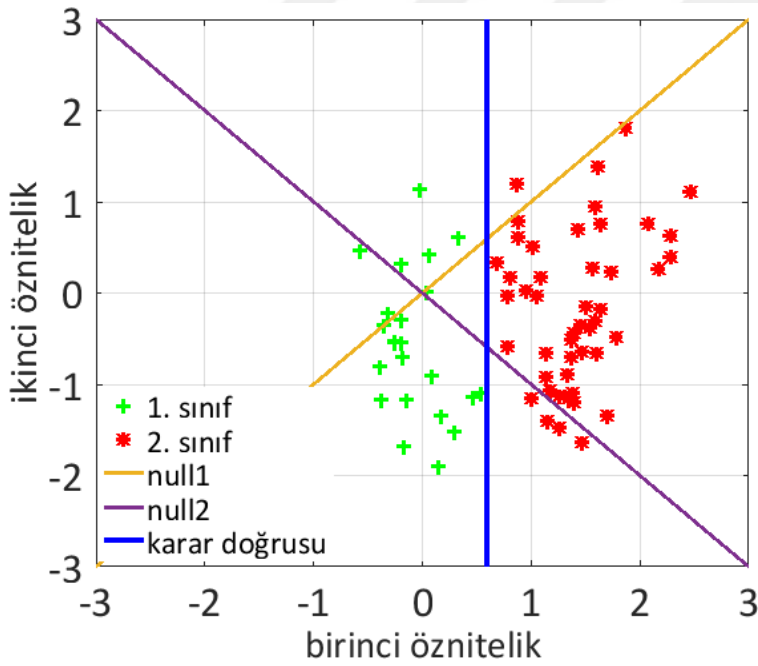
Böylece, DVM karar çizgisi temelli önileşim aşaması önerilmiştir. Bu önileşim en iyi sınıflandırma başarısını elde etmek için verilerin null eksenine olan en iyi pozisyonu belirleyerek, örnekleri Lorentz metriğine göre sınıflandırmaya uygun hale getirmektedir. Farklı sınıfları birbirinden en iyi şekilde DVM göre ayıran ve örneklerin Lorentz uzayındaki null eksenine olan pozisyonunu net bir şekilde belirleyen karar çizgisinin örneği Şekil 3.3'te görsel olarak verilmiştir.

DVM ile hesaplanan karar çizgisinin, dolayısıyla örneklerin Lorentz uzayında sınıflandırma yaparken en yüksek sonucu üretecek olan en iyi pozisyonlar incelenmiştir. Bu bağlamda üstte beyan edilen tespitler göz önünde bulundurularak karar çizgisinin Lorentz uzayının null eksenlerine göre alması gereken en iyi üç pozisyon belirlendi. Onlar:

- a) Üstte beyan edilen 2. tespite göre DVM ile hesaplanan karar çizgisinin y eksenine paralel veya tam üzerinde yer alması lazım. Karar çizgisinin x eksenine ile dik açıyla kesiştiği durum Şekil 3.4'te görsel olarak verilmiştir.



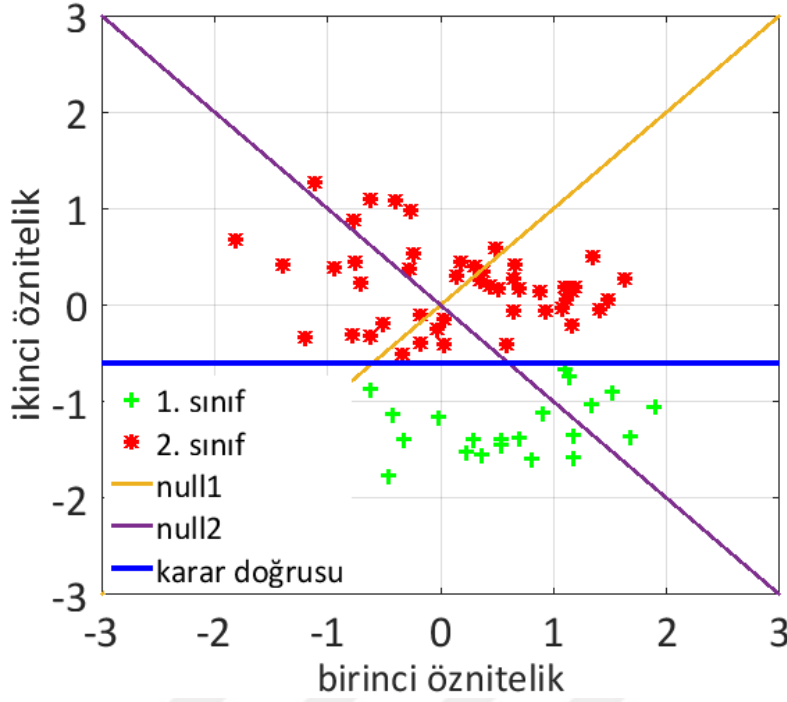
Şekil 3.3. DVM karar çizgisi



Şekil 3.4. Karar doğrusunun y eksenine paralel hali

- b) İkinci en iyi pozisyon karar çizgisinin x eksenine paralel veya tam üzerinde olmasıdır. Karar çizgisinin y eksenine dik açıyla kesiştiği durum Şekil 3.5'te görsel olarak verilmiştir. Birinci veya ikinci durumlarda DVM ile hesaplanan karar doğrusunun x veya y eksenine paralel olması ile tam eksenlerin üzerinde olmasının aralarında

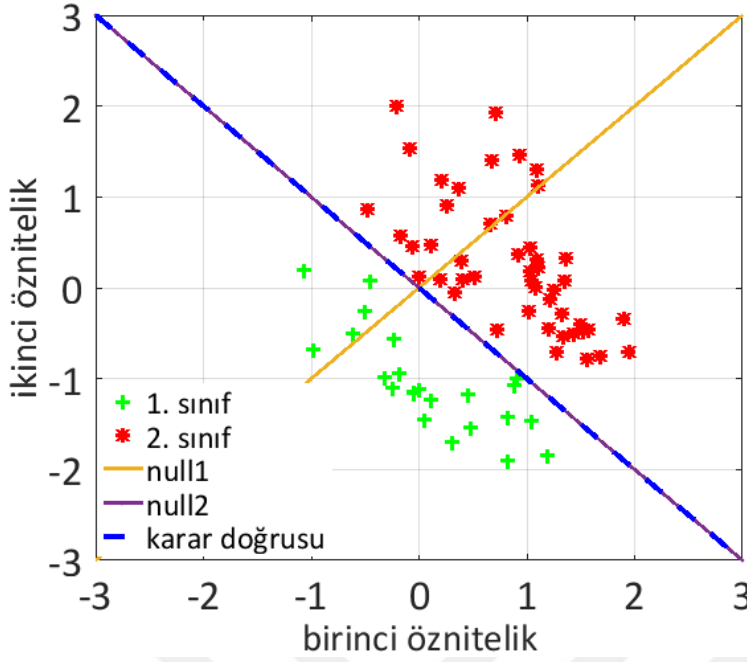
sınıflandırma başarısının artırımı bakımından fark yoktur. Aslında basit bir öteleme işlemi ile iki durum teke indirilebilir. Çünkü öteleme verilerin yapısal özelliklerini bozmamaktadır.



Şekil 3.5. Karar doğrusunun x eksenine paralel hali

- c) Üçüncü durum ise DVM ile bulunan karar doğrusunun tam null ekseninin üzerinde veya bu eksenle arasında az açığa sahip olmasıdır. Bu durumdan yararlanarak iki sınıfı en iyi şekilde doğrusal olarak ayıracak olan bir çizgiyi, herhangi bir null doğrusu üzerine getirmekle sınıf içi ile sınıflar arası verilerin daha iyi ayrılması sağlanabilmektedir. Karar çizgisinin null ekseninin tam üzerine geldiği durum Şekil 3.6'da görsel olarak verilmiştir.

DVM karar çizgisinin yardımı ile örneklerin bu en iyi pozisyonlara Lorentz uzayında eğitime sokulmadan önce getirilmesi Lorentz uzayının metriklerine göre en iyi sınıflandırma sonucunu üretmeyi sağlayacaktır. Bu işlemlerin bütünü DVM karar çizgisi temelli olan Önişlem-2'nin adımlarını oluşturmaktadır.



Şekil 3.6. Karar doğrusunun null ekseninin üzerindeki hali

Bu tez araştırmasında gerçekleştirilen ön deneme sonucunda iki sınıfı ayıran ve DVM ile hesaplanan karar çizgisinin null eksenini üzerinde olması, örneklerin Lorentz uzayında en yüksek sınıflandırma başarısını elde etme açısından eğitim öncesi alması gereken en iyi pozisyon olduğu tespit edilmiştir.

Bu bağlamda, DVM ile hesaplanan karar çizgisinin, dolayısıyla örneklerin Lorentz uzayında null eksene göre yerleştirilmesi için gerekli döndürme açısı karar çizgisinin x eksenini ile ve null eksenleri ile arasındaki $\cos(\alpha)$ açılara bakılarak hesaplanması sırasıyla Eş. 3.6, Eş. 3.7, ve Eş. 3.8’de verilmiştir.

$$a_{\text{çl}_1} = \alpha = \cos(\text{null}_1, S) \quad (3.6)$$

$$a_{\text{çl}_2} = \beta = \cos(\text{null}_2, S) \quad (3.7)$$

$$a_{\text{çl}_x} = \gamma = \cos(O, S) \quad (3.8)$$

Buradaki, null_1 null_2 -sırasıyla null eksenleri, S - DVM ile hesaplanan karar çizgisi, O – x eksenidir.

Eş. 3.6 ve Eş. 3.7 ile hesaplanan döndürme açıları ile örnekleri döndürebilmek için

döndürme yönü, yani saat yönüne doğru veya ters yönü önemlidir. Bu doğrultuda Eş. 3.9 ile verilmiş döndürme yönü belirleme koşulu oluşturulmuştur.

$$\begin{aligned}
 &\text{if } (\beta < 45^\circ \text{ ve } \gamma > 45^\circ) \\
 &R = \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \\
 &\text{elseif } (\beta < 45^\circ \text{ ve } \gamma < 45^\circ) \\
 &R = \begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} \\
 &\text{elseif } (\alpha < 45^\circ \text{ ve } \gamma > 45^\circ) \\
 &R = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \\
 &\text{elseif } (\alpha < 45^\circ \text{ ve } \gamma < 45^\circ) \\
 &R = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \\
 &\text{end}
 \end{aligned} \tag{3.9}$$

Böylece, DVM ile hesaplanan ve Lorentz uzayında en iyi sınıflandırma başarısını üretecek olan optimal pozisyonu belirleyen karar çizgisinin belirlenmesi ve null eksenine göre döndürülmesi işlemleri Şekil 3.7’de görsel olarak verilmiştir. Bu işlemlerin bütünü Önişlem-2’yi oluşturmaktadır ve verileri Lorentz uzayında eğitime hazır duruma getirmektedir.

Önişlem 2.

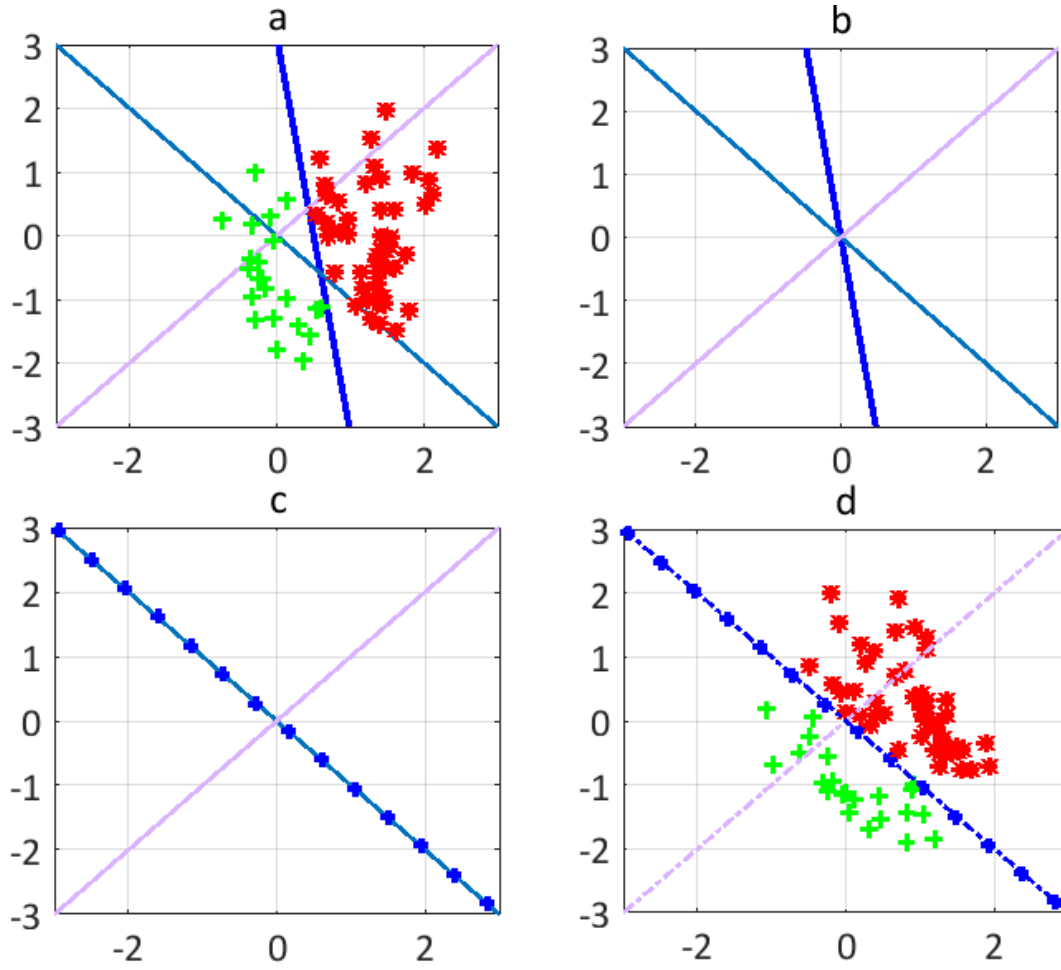
Girdi: X eğitim kümesi

Adım 1. DVM yardımı ile S karar çizgisini hesapla

Adım 2. Orijine ötele ($X' = X$)

Adım 3. Yönüne göre α veya β açılara döndür ($X'' = X'$)

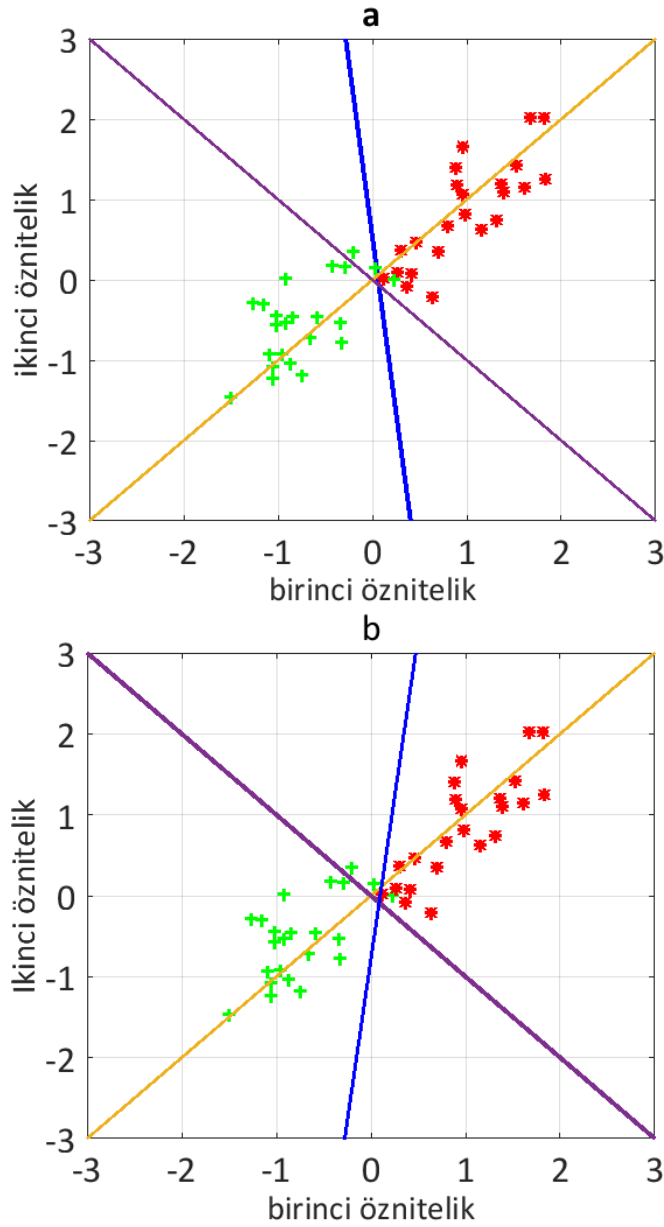
Adım 4. Orijinden geri ötele ($X''' = X''$)



Şekil 3.7. a) İki sınıflı veri kümesi için DVM yardımıyla hesaplanan karar çizgisi, b) orijine ötelenmiş karar çizgisi, c) karar çizgisi ile null eksenindeki açı kadar karar çizgisini ve d) verileri döndürme

Doğrusal Diskriminant Analizi

Doğrusal Diskriminant Analizi yöntemi ile Lorentz uzayında en iyi sınıflandırma sonucunu üreten karar çizgisini hesaplamada DVM ile karar çizgisini hesaplama yönteminde kullanılan adımlar rehber edinilmiştir. DVM ve DDA yöntemleri ile bulunan karar doğrularının görsel olarak karşılaştırılması Şekil 3.8’de verilmiştir.



Şekil 3.8. Karar çizgileri a) DVM ve b) DDA

3.2. Lorentz Uzaklığı Temelli Ayırt Etme Ölçütü

Sınıflandırma demek, örnekleri bir birinden benzer veya farklı özelliklerine göre ayırt etmek demektir. Bu bağlamda, Lorentz uzayında farklı sınıflara ait örnekleri birbirinden ayırt etmek için noktalar arası uzaklığı belirleyen Lorentz uzaklığı ayırt edici ölçüt olarak ele alınmıştır. Lorentz uzaklığı ayırt edici ölçüt olarak sınıflandırmanın en önemli adımı olan örneklerin eğitim aşamasında kullanılmıştır.

Önerilen yaklaşımın ilk aşamasında eğitim kümesi $X_i = (x_1, x_2, \dots, x_n)$ Lorentz uzayında bu tezin 3.1.1. bölümünde beyan edilen Önişlem-1 aşamalarıyla eğitim ve sınıflandırma için uygun hale $X'_i = (x'_1, x'_2, \dots, x'_n)$ getirilmektedir. Yeni gelen her bir örneğin $Y_j = (y_1, y_2, \dots, y_m)$ sınıflandırılması için eğitim kümesindeki sınıflara ait ortalamalara μ_s , ($s = \text{sınıf sayısı}$) göre ayrı ayrı projeksiyon (önişlem aşamasından geçirilir) $Y'_{j_s} = (y'_1, y'_2, \dots, y'_m)_s$ yapılmaktadır ve eğitim kümesindeki her noktası ile arasındaki uzaklık Eş. 2.6. ile hesaplanmaktadır.

$$d_L(x'_i, y'_j) = \sqrt{\sum_{i=1}^{n-1} |x_i - y_i|^2 - |x_n - y_n|^2} \quad (3.10)$$

Buradaki, n : X_i ve Y_j veri kümelerinin boyutu, yani öznitelik sayısıdır. Eğitim kümesindeki her bir örnek ile her yeni gelen test örneğinin arasındaki Lorentz uzaklıklarının tamamı Eş. 3.11 ile hesaplanmaktadır.

$$D_{(m \times n)_s} = \{d(y'_j, x'_i)_s\}, i = (1, n), j = (1, m) \quad (3.11)$$

Sonra her bir y'_j test örneği için $d(y'_j, x'_i)$ uzaklık vektöründen k komşu sayısı kadar en küçük uzaklık grubu $\min(d(y'_j, x'_i))^k$ seçilir ve eğitim kümesindeki örneklerle yakınlık derecelerine göre Eş. 3.12 ile etiketlenmektedir.

$$l_c = \min(d(y'_j, x'_i))^k, c = (1, s) \quad (3.12)$$

Bu işlem s sınıf, dolayısıyla onlara ait ortalamalar kadar her biri için ayrı ayrı yapılmaktadır. Son olarak her bir yeni gelen test örneği y'_j için belirlenen l_c etiket yığımından her sınıfa ait olma değeri Eş. 3.13 ile hesaplanmaktadır.

$$l_c = \sum_{i=1}^{k \times s} l_i \quad (3.13)$$

Son olarak her bir yeni gelen örnek y'_j için Eş. 3.14 ile verilmiş L_j sınıf değerlerinden en büyüğü hangi sınıfa ait ise o sınıfın etiketi verilmektedir.

$$\begin{aligned}
L_j &= \left\{ \min \left(d(y'_j, x'_i) \right)_1^k, \min \left(d(y'_j, x'_i) \right)_2^k, \dots, \min \left(d(y'_j, x'_i) \right)_s^k \right\} \\
&= \{l_{c_1}, l_{c_2}, \dots, l_{c_s}\} = L_{k \times s_j}
\end{aligned} \tag{3.14}$$

3.3. İki Boyutlu Veriler İçin Lorentz Sınıflandırıcısı

Bu çalışmanın 2.1. bölümünde belirtilen ve Lorentz uzayının “imza” adı altında bilinen tanımlamada iki çeşit örnek mevcuttur. Örneğin dört boyutlu Lorentz uzayında imza (-, -, -, +) veya (+, +, +, -) şeklinde verilmektedir [5]. Örnekten görüldüğü gibi biri diğerinin tersidir ve iki durumda da hesaplama yanlış olmaz. Ancak, bu tez çalışmasında araştırmalar iki boyutlu Lorentz uzayı üzerinde yapılmıştır. Dolayısıyla araştırmalarda negatif eksen olarak en sonuncu boyut veya öznitelik seçilmiştir. Böylece, her zaman en son boyutu negatif tanımlı olan uzaklık ölçütü elde edilmiştir.

3.3.1. Lorentz uzaklığı temelli sınıflandırma

Bu bölümde Lorentz uzayının sınıflandırmada kullanımı konusunda Önişlem-1 ve Eş. 3.11 ile verilmiş olan ayırt etme ölçütü temelinde Lorentz Uzaklığı Temelli Sınıflandırma (LUTS) yöntemi geliştirilmiştir ve önerilen yöntemin adımları Algoritma-1’de verilmiştir.

Algoritma 1. LUTS

Girdi: X, Y eğitim ve test veri kümeleri

Adım 1: X için Önişlem-1. adımlarını uygula ($X \rightarrow X'$)

Adım 2: Her yeni test örneği için ($y_j \in Y$)

Adım 1. uygula ve D_s örnekler arası uzaklıkları hesapla

En yakın uzaklıklara göre L_j sınıf etiketini belirle

Adım 3: Sınıflandırma oranını hesapla

3.3.2. İki boyutlu Lorentz Sınıflandırıcısı

Bu tez çalışmasının 3.3.1. bölümünde önerilen LUTS algoritması işlem açısından karmaşık olmasından dolayı bu yöntemi basitleştirmek veya alternatif sınıflandırıcı önermek adına

Lorentz uzayının sınıflandırmada kullanımı konusu derinden araştırılmıştır. Bu bağlamda, Lorentz uzayında en yakın komşu tekniği ile çalışan sınıflandırıcı geliştirme hedeflenmiştir.

Bu doğrultuda, ilk olarak Lorentz uzaklığı ile en yakın komşuları belirleyerek sınıflandırma yapan yeni Lorentz Sınıflandırıcısı (LS) önerilmiştir. Önerilen LS yöntemi Lorentz metriği ve bu araştırmanın 2.3. bölümünde belirtilen komşuluk tanımı temelinde çalışmaktadır. Benzer teknikle Öklid uzayında k En Yakın Komşu yöntemi çalışmaktadır [1].

LS yönteminin girdileri $x_i \in X, y_j \in Y$ eğitim ve test kümeleridir. Eğitim aşamasında Eş. 2.6'da verilmiş olan Lorentz uzaklığına göre $x_i \in X$ eğitim örnekleri ile en yakın komşularını belirleme tekniği ile eğitilmektedir. Test aşamasında her yeni gelen y_j test örneği ile tüm eğitim $x_i \in X$ örneklerinin arasında $d_L(X, y_j)$ Lorentz uzaklığı hesaplanmaktadır. Sonra Eş. 3.15 ile y_j test örneğine Lorentz uzaklığına göre en yakın k örnek seçilmektedir.

$$\min(d_L(X_k, y_j)), k = (1, i) \quad (3.15)$$

Sınıflandırma aşamasında ise seçilen örneklerin etiketlerindeki çoğunluğa göre yeni test örneğinin sınıfı belirlenmektedir. Önerilen LS sınıflandırıcısının adımları Algoritma-2'de verilmiştir.

Algoritma 2. LS

Girdi: X, Y eğitim ve test kümeleri

Adım 1: Her yeni test örneği için ($y_j \in Y$)

En yakın komşu örnekleri bul

Çoğunluğa göre sınıfını belirle

Adım 2: Sınıflandırma oranını hesapla

3.3.3. Karar çizgisi temelli Lorentz Sınıflandırıcısı

Geliştirilen LS yöntemi hem yapısı ve hem de işlem karmaşıklığı açısından çok basittir. Ayrıca, bu araştırmanın 3.1.1. ve 3.1.2. bölümlerinde belirtilen Lorentz uzayının özelliklerine göre sınıflandırma sonucu Lorentz metriği açısından örneklerin uzaydaki

dağılımından etkilendiği tespit edilmiştir. Yani, Lorentz uzaklık ölçütü ile çalışan sınıflandırıcı örneklerin dağılımına hassas olduğu ve doğrudan her zaman anlamlı sonuç üretemediği görülmüştür. Bu sorun genel olarak tüm ayırt edici ölçütlere has özelliktir [1]. Literatürde, bu sorunu ortadan kaldırmak için “feature representation”, yani sınıflandırma başarısını arttırmak için örnekleri anlamlı ve geçerli işlemler ile uygun hale getirme yöntemi kullanılmaktadır. Bu doğrultuda, ilk olarak bu çalışmada Lorentz uzayında en yakın komşu tekniği ile çalışan LS algoritmasına 3.1.2 bölümde önerilen Önişlem-2 adımları eklenerek yapılandırılmıştır.

Böylece, DVM veya DDA yöntemlerinin yardımı ile hesaplanan karar çizgisini kullanarak örnekleri null eksenine göre en iyi sınıflandırma sonucunu üretecek olan en iyi pozisyona taşıma işleminden oluşan Önişlem-2 ile yapılandırılan Karar Çizgisine Göre Ayarlanan Lorentz Sınıflandırıcısı (KÇ-GALS) sıralı adımlar olarak Algoritma-3’te verilmiştir.

Algoritma 3. KÇ-GALS

Girdi: X, Y eğitim ve test kümeleri

Adım 1: X için Önişlem-2 adımlarını uygula ($X \rightarrow X'$)

Adım 2: Her yeni test örneği için ($y_j \in Y$)

En yakın komşuları bul

Çoğunluğa göre sınıfını belirle

Adım 3: Sınıflandırma oranını hesapla

3.3.4. Sıkıştırma tabanlı Lorentz Sınıflandırıcısı

Bu tez araştırmasının 3.1.1. bölümünde tanımlanan Önişlem-1 adımları Lorentz uzayının özellikleri açısından tekrar değerlendirilerek işlem karmaşıklığını ve sınıflandırma başarısını arttırmak amacıyla optimizasyon işlemi gerçekleştirilmiştir. Böylece, Algoritma-2’de tanımlanan LS yöntemi sadece sıkıştırma ve öteleme aşamalarını içeren önişlemle tekrar yapılandırılmıştır. Öteleme sıkıştırma işleminin anlamlı yapılması için seçilen ek adımdır. Önerilen sıkıştırma tabanlı ve Lorentz uzayında sınıflandırma yapan yeni yöntemin sıralı adımları Algoritma-4’te verilmiştir.

Algoritma 4. Sıkıştırma ile LS (S-LS)

Girdi: X, Y eğitim ve test kümeleri

Adım 1: Sıkıştırma uygula ($X' = X\lambda$)

Adım 2: Her yeni test örneği için ($y_j \in Y$)

Sıkıştırma uygula ($Y' = Y\lambda$)

En yakın komşuları bul

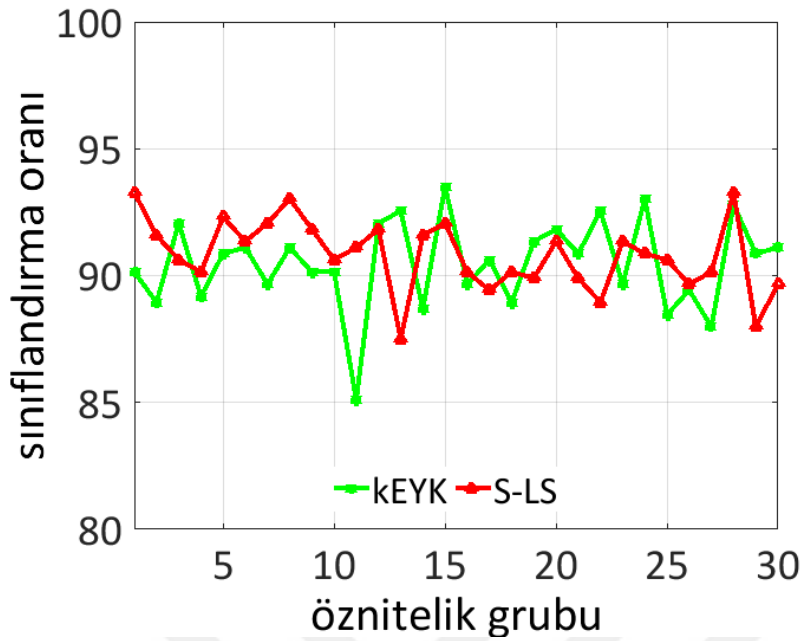
Çoğunluğa göre sınıfını belirle

Adım 3: Sınıflandırma oranını hesapla

En iyi parametre temelli Lorentz Sınıflandırıcısı

Genel olarak bu tez araştırmasında geliştirilen S-LS yönteminin ana adımlarının biri olan ve Lorentz uzaklık ölçütünü kullanarak başarıyı elde etmeyi sağlayan önişlem aşamasındaki sıkıştırma işlemi veya diyagonal matris çarpımı önemlidir. Çünkü bu matrisin parametrelerine göre projeksiyon yapılmaktadır ve örnekler Lorentz metriği ile sınıflandırma yapmak için uygun hale gelmektedir. Bu tez araştırmasında Eş. 3.3. ile verilmiş olan sıkıştırma matrisinin w, q parametreleri genel durumda $w = 2, q = 0.1$ olarak alınmıştır. Bu bağlamda, araştırmada sıkıştırma operatörü parametreleri değerlerinde yapılan değişimin ne derecede sınıflandırıcı sonucunu etkileyeceği incelenmiştir. Neticede S-LS ile en iyi sınıflandırma sonucunu üreten en iyi parametre değerlerinin veri kümelerine göre farklılık gösterdiği görülmüştür.

Örneğin, geliştirilen S-LS yöntemi ile CLIMATE veri kümesi üzerinde yapılan ön denemelerde bazı öznitelik ikili gruplarında önerilen yöntemin sınıflandırma sonucu k-En Yakın Komşu (k-EYK) yöntemine göre daha düşük olarak kaydedildiği görülmüştür. Araştırma iki boyutlu Lorentz uzayı üzerinde yapılmasından dolayı, denemelerde ikiden çok boyuta sahip veri kümeleri ikili öznitelik gruplarına ayrılarak kullanılmıştır. CLIMATE veri kümesinin 30 ikili öznitelik grubu için yapılan denemelerin görsel sonucu Şekil 3.9'da verilmiştir.

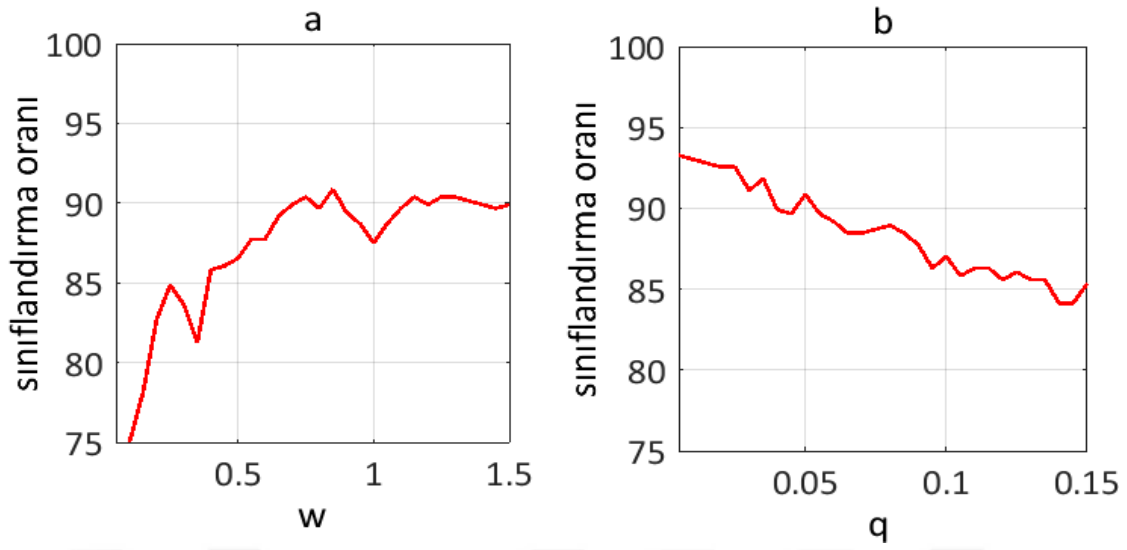


Şekil 3.9. CLIMATE için k-EYK ve S-LS algoritmalarının başarı oranları

Bu durumun aşılması ve en iyi parametrelerin bulunması için anlamlı yöntem geliştirmek adına, λ sıkıştırma matrisinin w ve q parametreleri sırayla belirli aralıklarda ve standart sabit değerleri de içerecek şekilde değiştirilerek deneme yapılmıştır. Ancak, tüm veri kümesini temsil eden en iyi parametre değerlerini her 30 ikili grubunu ayrı ayrı deneyerek bulmak çok maliyetlidir ve anlamlı bir işlem değildir. Bu doğrultuda, en iyi parametre değerleri tüm ikili öznitelik grupları içinde S-LS yöntemi ile en kötü sonuç üreten ikili öznitelik grubu ile bulunmaktadır. Sıkıştırma matrisinin w ve q parametrelerinin değişim aralıkları Eş. 3.16'da verilmiştir.

$$\begin{aligned} w &= (0, 3) \\ q &= (0.01, 0.3) \end{aligned} \tag{3.16}$$

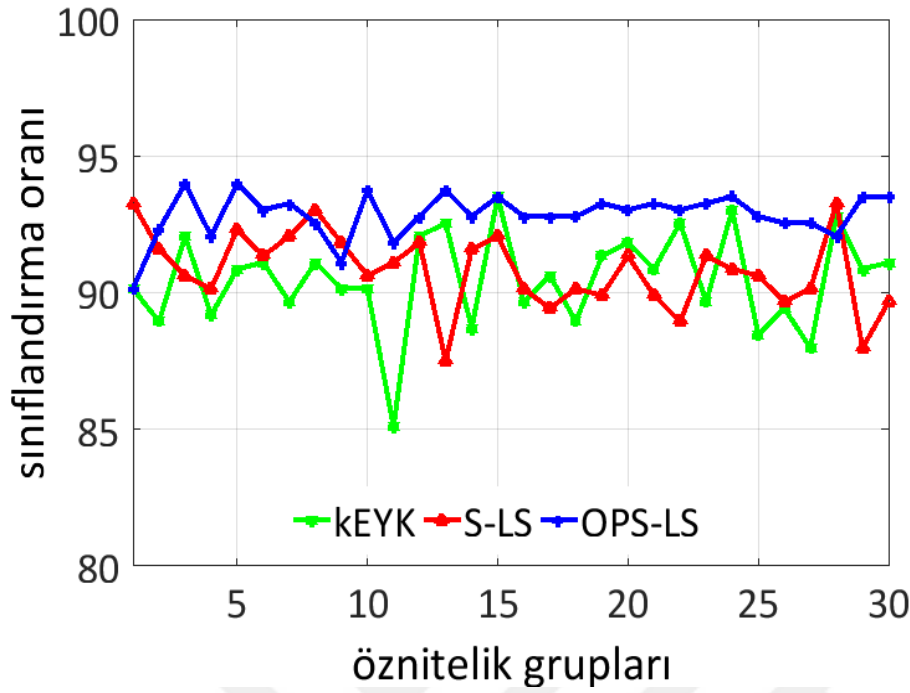
Deneme sonuçları ayrı ayrı kaydedilerek analiz edilmiştir ve CLIMATE için en iyi sınıflandırma sonucu değerleri ile $w = 1.7$ ve $q = 0.01$ elde edildiği görülmüştür. Eş. 3.16 ile verilmiş aralıkta sıkıştırma matrisi değerlerinin değişimi önerilen S-LS yöntemi sonucunu ne derece etkilediğinin görselleştirilmesi Şekil 3.10'da verilmiştir.



Şekil 3.10. CLIMATE veri kümesi için S-LS algoritmalarının a) w ve b) q parametrelerinin farklı değerleri için ürettiği sınıflandırma sonuçları

Şekil 3.10'dan görüldüğü gibi sıkıştırma matrisinin parametreleri $w=1.7$ ve $q=0.01$ değerlerine sahip olduğunda önerilen S-LS yöntemi en iyi sınıflandırma sonucunu üretmiştir. En kötü sonuç üreten ikili öznitelik grubu için elde edilen en iyi sıkıştırma parametrelerinin CLIMATE veri kümesinin diğer tüm öznitelik grupları için kullanılması durumunda önerilen S-LS yönteminin ürettiği sonuçlar Şekil 3.11'de verilmiştir. Şekil 3.11'den görüldüğü gibi en iyi parametre ile sıkıştırma temelli Lorentz Sınıflandırıcısı k-EYK ve S-LS yöntemlerine göre CLIMATE için daha başarılı sınıflandırma sonucunu üretmektedir. Bu durum en iyi parametrenin bulunması için en kötü sonucu üreten ikili öznitelik grubunu kullanma yönteminin anlamlı bir adım olduğunu göstermektedir.

Böylece, en iyi parametrelili sıkıştırma temelli Lorentz Sınıflandırıcısı önerilmiştir ve geliştirilen yeni OP-LS algoritmasının adımları Algoritma-5'te verilmiştir.



Şekil 3.11. CLIMATE veri kümesi için k-EYK, S-LS ve OP-LS algoritmalarının sınıflandırma sonuçlarının karşılaştırılması

Algoritma 5. En iyi Parametrelili-LS (OP-LS)

Girdi: X, Y eğitim ve test kümeleri

Adım 1: λ matrisi için en iyi w, q parametreleri bul

Adım 2: Sıkıştırma uygula $X' = X\lambda$

Adım 3: Her yeni test örneği için $y_j \in Y$

Sıkıştırma uygula $Y' = Y\lambda$

En yakın komşuları bul

Çoğunluğa göre sınıfını belirle

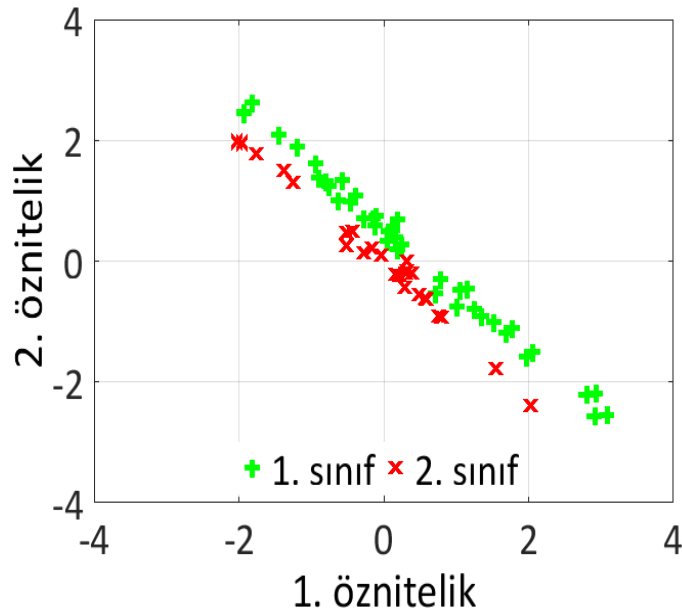
Adım 4: Sınıflandırma oranını hesapla

Döndürme temelli Lorentz Sınıflandırıcısı

Bu tez araştırmasında önerilen S-LS yönteminin sınıflandırma sonucunu arttırmada Lorentz metriği açısından yararlı olabilecek diğer özellikler ve işlemler analiz edilmiştir. Yani önışlem aşamasına eklenecek ekstra adım ile S-LS yönteminin sınıflandırma başarısını

arttırma olanağı incelenmiştir. Bu bağlamda, bu tez araştırmasının 3.1.1 bölümünde belirtilen ve Eş. 3.5 ile tanımlanan döndürme işleminin S-LS yönteminin önışlem aşamasına eklenmesi halinde sınıflandırma başarısını ne derece etkileyeceği araştırılmıştır. Çünkü bu araştırmanın 3.1.2. bölümünde döndürme işleminin örneklerin sınıflandırma açısından en iyi pozisyonu belirlemede önemli bir adım olduğu belirtilmiştir.

Böylece, önerilen S-LS algoritmasına sıkıştırma adımından sonra 45° derecelik döndürme işlemi eklenmiştir. Döndürme işlemi Eş. 3.5. ile yapılmaktadır. Döndürme işlemi önışlem adımında hem eğitim hem de test alt kümeleri için uygulanmaktadır. Eğitim veri kümesi için iki sınıfın ortalamasına göre döndürme yapılmaktadır. Test veri kümesi için eğitim veri kümesindeki iki sınıfın geometrik ortasına göre döndürme yapılmaktadır. Örneğin, VERTEBRAL veri kümesi için döndürme işlemi sonrası örneklerin pozisyonu görsel olarak Şekil 3.12’de verilmiştir.



Şekil 3.12. VERTEBRAL veri kümesi için döndürme

Son olarak, Lorentz uzayında döndürme ve sıkıştırma ile sınıflandırma yapan D-LS yöntemi önerilmiştir ve geliştirilen yöntemin adımları Algoritma-6’da verilmiştir.

Algoritma 6. Döndürme ile LS (D-LS)

Girdi: X, Y eğitim ve test kümeleri

Adım 1: Sıkıştırma uygula ($X' = X\lambda$)

Adım 2: Döndür ($X'' = X'R$)

Adım 3: Her yeni test örneği için $y_j \in Y$

Sıkıştırma uygula ($Y' = Y\lambda$)

Döndür ($Y'' = Y'R$)

En yakın komşuları bul

Çoğunluğa göre sınıfını belirle

Adım 4: Sınıflandırma oranını hesapla

3.4. Çok Boyutlu Veriler İçin Lorentz Sınıflandırıcısı

Bu tez araştırmasının 3.3. bölümü ve altbölümlerinde önerilen yöntemler iki boyutlu Lorentz uzayında sınıflandırma yapmaktadırlar. Ancak, günümüzün gerçek uygulamalarında iki boyutlu veriler çok az sayıda karşımıza çıkmaktadır [34]. Tam tersine yani çok boyutlu verilerle çalışan uygulamalar oldukça önem taşımaktadır. Örneğin, yüz doğrulama, örüntü tanıma, internet ortamında saldırı tespiti gibi uygulamalar [1, 34, 35]. Bu doğrultuda, 3.3. bölümde önerilen sınıflandırıcıların ikiden fazla özneliğe sahip veriler üzerinde geçerli ve uygulanabilir yapmak amacıyla Lorentz uzayı ile sınıflandırma konusunun kapsamı genişletilmiştir.

Çok boyutlu veya çok özneliğe sahip olan veriler üzerinde yapılan çalışmaların işlem karmaşıklığı açısından maliyetli olması sınıflandırıcının verimliliğini düşürmektedir [1]. Ayrıca, istenilen sınıflandırma başarısını elde etmek zordur. Bu bağlamda, bu sorunu aşabilmek için az sayıdaki öznelik kombinasyonları veya alt gruplar ile sınıflandırma başarısını arttırmayı hedefleyen bir araştırma yapılmıştır. Araştırmada, veri kümelerinden elde edilen bazı az sayıdaki öznelik kombinasyonları üzerinde yapılan analiz sonucunda sınıflandırma açısından bu alt grupların kullanımı tüm özneliklerin kullanımına göre daha başarılı sonuç ürettiği görülmüştür [1, 35].

3.4.1. İkili öznitelik grupları

Bu tez araştırmasında Lorentz uzayı ve özellikleri ile 3.2. ve 3.3. bölümlerindeki tespitlerden yararlanmak ve onlara uyumlu yapmak adına çok boyutlu verilerden elde edilecek öznitelik kombinasyonlarının boyutu iki olarak belirlenmiştir. İkili öznitelik kombinasyonları istatistikte çok iyi bilinen kombinasyon yöntemi ile elde edilmiştir [36]. Benzer teknik DVM algoritması ile çok sınıflı örnekleri sınıflandırmak için de kullanılmaktadır [31]. Yani, sıradaki öznitelik ve diğerleri şeklindedir. Örneğin dört özniteliğe sahip bir veri kümesi için tüm ikili kombinasyonlar $\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}$ şeklindedir. Küme içinde elemanların sırasının önemli olmadığı gibi kombinasyonlarda da sıranın önemi yoktur. n özniteliğe sahip bir veri kümesinin r boyutlu öznitelik kombinasyonlarının sayısı $S(n, r)$ şeklinde tanımlanmaktadır. Veri kümelerinden ikili öznitelik kombinasyonlarını oluşturma yöntemi Eş. 3.17 ile yapılmaktadır [36].

$$S(n, r) = \binom{n}{r} = \frac{n! (n - r)!}{r! (n - r)!} \quad (3.17)$$

Böylece, Eş. 3.17 ile orijinal veri kümesinin tüm özniteliklerini kapsayan ikili öznitelik kombinasyonları elde edilebilmektedir. Örneğin, 20 özniteliğe sahip olan veri kümesinde Eş. 3.17'e göre, seçilecek kombinasyon boyutu $r = 2$ olursa, toplamda 190 ikili öznitelik kombinasyonu elde edilmektedir. Ancak, bu kombinasyonların sayısı, eğer veri kümesinde öznitelik sayısı çok olursa kat kat artar ve elde edilen tüm ikili öznitelik gruplarının tamamı üzerinde işlem yapmak çalışma zamanı açısından çok maliyetlidir. Dolayısıyla, bu çalışmada Lorentz uzayında orijinal veri kümesini en iyi şekilde temsil edecek ve en yüksek sınıflandırma başarısını üretecek olan en iyi ikili öznitelik kombinasyonlarını seçme yöntemi araştırılmıştır.

3.4.2. Lorentz metriği temelli öznitelik seçme

Genel olarak öznitelik seçme işlemi çok boyutlu veri kümelerinde sınıflandırıcı olarak bilinen makine öğrenmesi algoritmalarının çalışma zamanını azaltmak ve en önemlisi de veri kümesini temsil eden en iyi öznitelik gruplarını ya da kombinasyonlarını seçerek sınıflandırma başarısını arttırmak amacıyla kullanılan yöntemdir [33-35]. Ayrıca, çok

boyutlu verilerde çokça karşılaşılan ilgisiz veya gürültülü bilgilerin elenmesi açısından da öznitelik seçme yöntemi önem taşımaktadır. Literatürde istatistik tabanlı çalışan ve en iyi öznitelik vektörlerinden oluşan altküme seçme yöntemleri örüntü tanımada sık kullanılmaktadır [1, 34]. Öznitelik seçme yöntemleri veri kümesinden en iyi öznitelik gruplarını seçmek için ana kriter olarak ayırt edicilik ölçütlerini kullanmaktadır [1, 35]. Bu ölçütlerden birisi ve Öklid uzayında örneklerin sınıflar içi ve sınıflar arası dağılım matrisleri temelinde hesaplanan J ayırt edicilik ölçütü Eş. 3.18 ile verilmiştir [1]

$$J = \text{trace}(S_w^{-1}S_m) \quad (3.18)$$

Buradaki, S_w M sınıflı veri kümesindeki örneklerin sınıflar içi dağılımını tanımlayan matristir. Bu matris her bir w_i sınıfının x öznitelik vektörü ile μ_h sınıf içi ortalaması ayırımının kovarians matrisi Σ_h ve P_h en yüksek olasılık değerinin çarpımından oluşmaktadır ve Eş. 3.19 ile ifade edilmektedir.

$$S_w = \sum_{h=1}^M P_h \Sigma_h \quad (3.19)$$

Buradaki, Σ_h kovaryans matrisi Eş. 3.20 ile hesaplanmaktadır.

$$\Sigma_h = E[(x - \mu_h)(x - \mu_h)^T] \quad (3.20)$$

S_m matrisi ise x öznitelik vektörü ile μ_0 genel ortalamanın ayırımının kovaryans matrisidir ve Eş. 3.21’de verilmiştir.

$$S_m = E[(x - \mu_0)(x - \mu_0)^T] \quad (3.21)$$

Literatürde S_m değeri karışık dağılım matrisi olarak bilinmektedir [1].

Örneklerin dağılım matrisi temelindeki J öznitelik ayırt edici ölçütü Öklid uzayının metriğine göre çalıştığı için Lorentz uzayında geçerli değildir ve bu tez araştırmasında Lorentz metriği temelinde öznitelik seçme yöntemini geliştirmek adına Lorentz uzayının özelliklerine göre yapılandırılmıştır.

Yapısal olarak J ayırt edici ölçütü Eş. 3.20 ve Eş. 3.21’den görüldüğü üzere kovaryans matrisi hesaplama işlemini içermektedir. Kovaryans matrisi hesaplama işlemi ise matris çarpımından oluşmaktadır [1, 25]. Bu tez araştırmasının 2.2. bölümünde belirtildiği gibi

Lorentz uzayında matris çarpımı Eş. 2.9 ile yapılmaktadır. Böylece, Öklid uzayında Eş. 3.20 ve Eş. 3.21 ile verilmiş olan öznitelik vektörlerinin kovaryans matrisini hesaplama işlemi Lorentz uzayının metriğine göre yapılandırılmıştır ve Eş. 3.22 ile Eş. 3.23'te verilmiştir.

$$(\Sigma_h)_L = E[(x - \mu_h) \cdot_L (x - \mu_h)^T] \quad (3.22)$$

$$(\Sigma_m)_L = E[(x - \mu_0) \cdot_L (x - \mu_0)^T] \quad (3.23)$$

Bununla birlikte, Eş. 3.20 ve Eş. 3.21 ifadelerini Lorentz metriğine göre yapılandırarak, bu uzayda kovaryans matrisini hesaplama işlemi tanımlanmıştır. Son olarak, Eş. 3.22 ve Eş. 3.23 ile Lorentz uzayında öznitelik gruplarını ayırt edecek olan Lorentz J (LJ) ölçütü elde edilmiştir ve Eş. 3.24'te verilmiştir.

$$LJ = trace[(S_w^{-1})_L (\Sigma_m)_L] \quad (3.24)$$

Böylece, LJ ölçütü ile Lorentz uzayının metriği ve özelliklerine göre en iyi öznitelik kombinasyonlarını seçebilen Lorentz Öznitelik Seçme (LÖS) yöntemi geliştirilmiştir.

3.4.3. Çok Boyutlu Lorentz Sınıflandırıcısı (ÇBLS)

Bu bölümde, çok boyutlu veriler için Lorentz uzayında sınıflandırma yapan yöntem anlatılmıştır. Çok boyutlu verileri sınıflandırılmasında çalışma zamanı ve işlem karmaşıklığını azaltmak için öznitelik seçimi gibi boyut indirgeme yöntemleri ve onların ana bileşeni olan ayırt edicilik ölçütünün doğru seçimi gibi meselelerin önemliliği vurgulanmıştır. Bu bağlamda, bu tez araştırmasında çok boyutlu verileri Lorentz uzayında LÖS yöntemi ile seçilen en iyi ikili öznitelik grupları temelinde sınıflandıran yeni Çok Boyutlu Lorentz Sınıflandırıcısı (ÇBLS) önerilmiştir.

Önerilen ÇBLS yönteminde ilk olarak çok boyutlu veri kümesi Eş. 3.17'e göre $S(n, 2)$ tane ikili kombinasyonlara ayrılmaktadır. Sonra, ayrılan ikili öznitelik kombinasyonlarını LJ ayırt edici ölçütünün değerine göre sıralayarak, her veri kümesi için Lorentz Öznitelik Seçme ile $k = (1, S)$ sayısı kadar en iyi ikili öznitelik kombinasyonları seçilmektedir. Buradaki, S toplam ikili öznitelik grup sayısıdır.

Önerilen ÇBLS yönteminde seçilen ikili öznitelik kombinasyonları ile eğitim yapılmaktadır. Her yeni gelen test örneği için Lorentz uzaklık ölçütü temelinde eğitim seti örnekleri arasından k en yakın komşu belirleme ve komşular arasındaki çoğunluğuna göre sınıf etiketi belirleme yapılmaktadır. Aslında, bu adımlar bu araştırmanın 3.3.2 bölümünde tanımlanan Algoritma-2'nin adımlarıyla örtüşmektedir. Ancak, Algoritma-2 Lorentz uzayında iki boyutlu verilerin sınıflandırılması için geliştirilen yöntemdir. Bu nedenle, ikiden fazla öznitelige sahip veri kümelerini Lorentz uzayında sınıflandırmak için LS yöntemine ek karar verme mekanizması eklenerek geliştirilmiştir ve Çok Boyutlu Lorentz Sınıflandırıcısı (ÇBLS) elde edilmiştir. Söz konusu ek karar verme mekanizması birden fazla olan ve orijinal çok boyutlu veri kümesini temsil edecek olan en iyi ikili öznitelik grupları için üretilen sınıf etiketlerinden anlamlı ve geçerli sonuç elde etmek amacıyla kullanılmaktadır. Önerilen ÇBLS yöntemi seçilen ikili öznitelik grubunun sayısı g kadar iterasyonla çalışmaktadır ve bu değer aynı zamanda durdurma koşulu olarak da görev yapmaktadır. Bu değer az olması çalışma zamanını önemli ölçüde azaltmaktadır.

Önerilen ÇBLS yöntemi Lorentz uzayında her yeni gelen test örneklerini sınıflandırmak amacıyla en yakın komşu tekniği temelinde seçilen her ikili öznitelik grubu için $c_i, i = (1, g)$ sınıf etiketini belirler. Üretilen c_i sınıf etiketi LÖS ile seçilen i . ikili öznitelik kombinasyonu için elde edilen sınıf etiketidir. Algoritmanın iterasyon artımında c_i sınıf etiketlerinin birleşiminden oluşan ve bütün veri kümesini temsil edecek olan C_i etiket dizisi elde edilir. Geliştirilen yeni Çok Boyutlu Lorentz Sınıflandırıcı algoritmasının i . iterasyonunda bu dizinin C_i vektörüne bakılarak test örneğinin sınıf etiketine karar verilir. İlk adımda $C_0 = [c_1]$ olarak belirlenmektedir. Sonraki adımda $C_i = [C_{i-1}, c_i]$ olarak devam eder. Sınıflandırma oranının hesaplamada ise C_i kullanılmaktadır ve sınıfları belirlemekte çoğunluk hesaba alınmaktadır. Örneğin, iki sınıflı ve çok boyutlu veri kümesinden elde edilen toplam S ikili öznitelik kombinasyonlarından LÖS ile $g=5$ oranında seçilen en iyi ikili gruplar için önerilen ÇBLS yöntemi $C_5 = [0, 1, 1, 0, 1]$ şeklinde etiket dizisi üretecek olursa, çoğunluğa göre bu test örneğinin sınıf etiketi 1 olur.

Son olarak, Lorentz uzayında çok boyutlu veriler için geliştirilen ÇBLS yönteminin adımları Algoritma-7'de verilmiştir.

Algoritma 7. Çok Boyutlu LS (ÇBLS)

Girdi: X, Y eğitim ve test kümeleri

Adım 1: $S(n, 2)$ tane X' ikili öznitelik gruplarını oluştur

Adım 2: g kadar ikili öznitelik alt grubunu X'' LÖS yöntemi ile seç

Adım 3: Sıkıştırma uygula ($X' = X\lambda$)

Adım 4: Y test örneği için Y''' oluştur

Adım 5: Her bir i için $i=(1, g)$

X''' göre y_i''' en yakın komşularını bul

Çoğunluğa göre c_i sınıf etiketini belirle

$C_i = [C_{i-1}, c_i]$ oluştur, $C_0 = [c_1]$

Adım 6: C_i sınıf etiketine göre sınıflandırma yap

En iyi parametrelili ÇBLS

Çok boyutlu veriler için en iyi parametrelili sıkıştırma matrisi ile Lorentz uzayında sınıflandırma yapan ÇBLS yönteminin adımları Algoritma 8’de verilmiştir.

Algoritma 8. En iyi Parametrelili (OP-ÇBLS)

Girdi: X, Y eğitim ve test kümeleri

Adım 1: $S(n, 2)$ tane X' ikili öznitelik gruplarını oluştur

Adım 2: g kadar ikili öznitelik alt grubunu X'' LÖS yöntemi ile seç

Adım 3: En iyi w ve q ile sıkıştırma uygula ($X''' = X''\lambda$)

Adım 4: Y test örneği için Y''' oluştur

Adım 5: Her bir i için $i=(1, g)$

X''' göre y_i''' en yakın komşularını bul

Çoğunluğa göre c_i sınıf etiketini belirle

$C_i = [C_{i-1}, c_i]$ oluştur, $C_0 = [c_1]$

Adım 6: C_i sınıf etiketine göre sınıflandırma yap

Döndürme temelli ÇBLS

Algoritma-7’de tanımlanan ÇBLS yönteminin önışlem aşamasına döndürme işleminin eklenmesi yapılandırılan çok boyutlu veriler için ÇBLS yönteminin adımları Algoritma-9’da verilmiştir.

Algoritma 9. Döndürme ile ÇBLS (D-ÇBLS)

Girdi: X, Y eğitim ve test kümeleri

Adım 1: $S(n, 2)$ tane X' ikili öznitelik gruplarını oluştur

Adım 2: g kadar ikili öznitelik alt grubunu X'' LÖS yöntemi ile seç

Adım 3: Sıkıştırma ve döndürme uygula ($X''' = (X''\lambda)R$)

Adım 4: Y test örneği için Y''' oluştur

Adım 5: Her bir i için $i=(1, g)$

X''' göre y_i''' en yakın komşularını bul

Çoğunluğa göre c_i sınıf etiketini belirle

$C_i = [C_{i-1}, c_i]$ oluştur, $C_0 = [c_1]$

Adım 6: C_i sınıf etiketine göre sınıflandırma yap

3.5. ÇBLS ile Yüz Doğrulama

Bu araştırmada Lorentz uzayının özel yapısal özelliklerinden yararlanarak yeni yüz doğrulama sistemini geliştirme hedeflenmiştir. Kimlik doğrulamada yaygın olarak parmak izi ve yüz resmi kullanılmaktadır ve bu bilgileri kullanarak bilgisayarın yetkili kişiyi tanınması ve öğrenmesi için yüz doğrulama gibi Makine Öğrenmesi yöntemleri kullanılmaktadır [37-41]. Yüz doğrulama probleminde kullanılan yüz resimleri yüksek çözünürlüklü olması nedeniyle çok boyutlu veri olarak işlenmiştir. Bu bağlamda çok boyutlu veriler için geliştirilen Algoritma-7’de tanımlanan ÇBLS yöntemi yüz doğrulama veri kümeleri açısından analiz edilerek, optimize edilmiştir. Yani, Lorentz uzayında sınıflandırma yapan Çok Boyutlu Lorentz Sınıflandırıcı yöntemi yüz veri tabanlarının sınıflandırmasında sadece LÖS öznitelik seçme ve sınıflandırma adımlarını içerecek şekilde daha basit hale getirilmiştir. Bunun için, ÇBLS yönteminin eğitim aşamasında örnekler iki

sınıflı yüz veri kümeleri halinde ele alınmıştır. Yani n sınıflı örnek için i . sınıf ($i=1, n$) ve diğerleri şeklinde eğitim yapılmıştır. Hal böyle iken bu işlem aslında yüz veri kümelerinin sınıflandırılması değil yüz doğrulama (face verification) olarak değişmektedir.

Önerilen yöntem her yeni gelen Y' test örneği için Lorentz uzaklık ölçütü temelinde eğitim kümesi örnekleri arasından k en yakın komşu belirlenerek ve komşular arasındaki çoğunluğuna göre benzerlik değeri hesaplanmaktadır. Bu benzerlik değerine göre yüz doğrulama yapılmaktadır. Benzerlik değeri 1 veya 0 şeklindedir, yani sırasıyla doğru veya yanlıştır. Böylece, Lorentz metriği temelinde yüz resimlerini kullanarak Lorentz Yüz Doğrulama (LYD) yapabilen yeni algoritma önerilmiştir ve geliştirilen yöntemin adımları Algoritma-10'da verilmiştir:

Algoritma 10. Lorentz Yüz Doğrulama (LYD)

Girdi: X, Y eğitim ve test kümeleri

Adım 1: $S(n, 2)$ tane X' ikili öznitelik gruplarını oluştur

Adım 2: g kadar ikili öznitelik alt grubunu X'' LÖS yöntemi ile seç

Adım 3: Y test örneği için Y'' oluştur

Adım 4: Her bir i için $i=(1, g)$

X'' göre y_i'' en yakın komşularını bul

Çoğunluğa göre c_i benzerlik değerini hesapla

$C_i = [C_{i-1}, c_i]$ oluştur, $C_0 = [c_1]$

Adım 5: C_i benzerlik değerine göre doğrulama yap

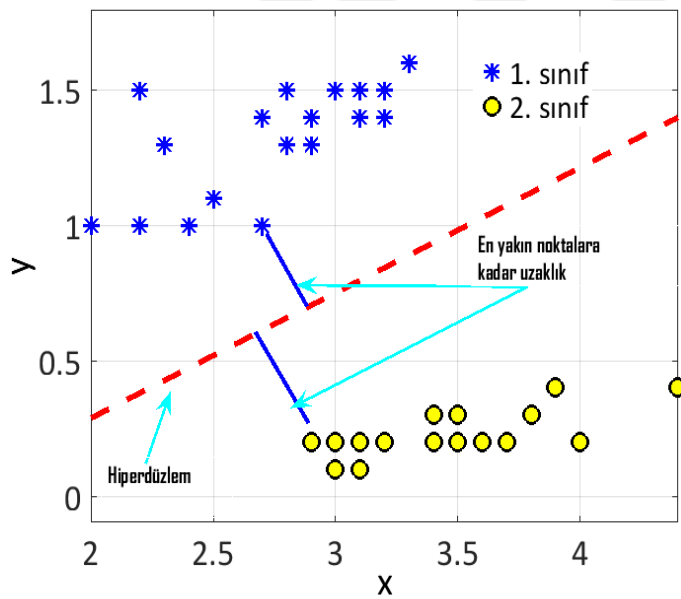
3.6. Lorentz Uzayında Destek Vektör Makinesi

Tez araştırmasının bu bölümünde Destek Vektör Makinesinin Lorentz uzayına uyarlanarak yeni bir sınıflandırıcı geliştirmesi hedeflenmiştir. Bu bağlamda Lorentz uzayının tüm yapısal özellikleri açısından DVM yönteminin içerdiği işlemler detaylı analiz edilerek Lorentz metriğine göre yeniden yapılandırılmıştır.

3.6.1. Destek Vektör Makinesi (DVM)

Geçtiğimiz bir kaç yıl içerisinde Destek Vektör Makinesi (DVM) yöntemine olan ilgi oldukça artmıştır [29]. Bu durumun ortaya çıkmasının ana nedeni DVM yönteminin el yazısı, yüz tanıma, örüntü tanıma ve diğer önemli problemlerin çözümünde yüksek başarı sergilemesidir [29-32].

Destek Vektör Makinesi sınıflandırma yöntemi olarak 1979 yılında Vladimir Vapnik tarafından ortaya atılmıştır [30]. DVM yöntemi, yapısı bakımından çok basit, lineer form olan hiperdüzlem yardımı ile pozitif ve negatif örnek kümelerini birbirinden en maksimum marjinle ayırmaktadır. Lineer durumlarda bu marjin hiperdüzlemden negatif ve pozitif örneklere kadar olan uzaklıkla tanımlanmaktadır. Bu durumun görsel olarak şekillendirilmesi Şekil 3.13'te verilmiştir.



Şekil 3.13. Destek Vektör Makinesinin ayırt edici çizgisi

Lineer DVM makinasının çıktısı Eş. 3.25'te verilmiştir.

$$u = \vec{w} * \vec{x} - b \quad (3.25)$$

Buradaki, w hiperdüzleme olan normal vektörüdür ve x ise giriş vektörüdür. Pozitif ve negatif veya farklı sınıflara ait örnekleri birbirinden ayıran hiperdüzlem $u = 0$ düzlemidir.

Hiperdüzleme en yakın noktalar $u = \pm 1$ marjinde yerleşmektedir. Bu nedenle hiperdüzlem ile destek vektörleri aralarındaki m marjin Eş. 3.26 ile tanımlanmıştır.

$$m = \frac{1}{\|w\|} \quad (3.26)$$

Bu m marjininin maksimize edilmesi farklı sınıfa ait örneklerin belirgin bir şekilde ayırt edilmesi demektir ve bu DVM yöntemi ile sınıflandırma problemini çözmedeki ana adımların biridir. Bu bağlamda, m marjininin maksimize edilmesi Eş. 3.27 ile verilen optimizasyon probleminin çözümü ile yapılabilmektedir.

$$\begin{aligned} y_i(\vec{w} * \vec{x}_i - b) &\geq 1, \forall i \text{ için} \\ \min_{w,b} \frac{1}{2} \|\vec{w}\|^2 \end{aligned} \quad (3.27)$$

Buradaki, x_i i . eğitim örneğidir ve y_i ise DVM yönteminin i . eğitim örneği için ürettiği çıktıdır. Ayrıca, y_i değişkeni pozitif örnekler için +1, negatif örnekler için -1 değerini alır.

Bu optimizasyon problemi Lagrange operatörü ile dual form olan QP problemine dönüştürülebilir. Böylece, söz konusu problemdeki Ψ amaç fonksiyonu α_i Lagrange çarpanlar kümesine doğrudan bağımlıdır ve dual form Eş. 3.28'de verilmiştir.

$$\min_{\vec{\alpha}} \Psi(\vec{\alpha}) = \min_{\vec{\alpha}} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j (\vec{x}_i \cdot \vec{x}_j) \alpha_i \alpha_j - \sum_{i=1}^N \alpha_i \quad (3.28)$$

Buradaki, N eğitim için kullanılan örnek sayısıdır ve $\alpha_i \geq 0$ eşitsizliği ile Eş. 3.29'da verilmiş koşulun geçerli olması şarttır.

$$\sum_{i=1}^N y_i \alpha_i = 0 \quad (3.29)$$

Eş. 3.28'den görüldüğü gibi, her bir eğitim örneği ile her Lagrange çarpanları arasında bire bir ilişki mevcuttur. Yani, her bir örneğe ayrı bir çarpan hesaplanmaktadır. Eğer Lagrange çarpanları belli ise ve herhangi bir $\alpha_i > 0$ için w normal vektör ile b sınır değerinin Lagrange çarpanları yardımı ile hesaplanması Eş. 3.30 ve Eş. 3.31 ile verilmiştir.

$$\vec{w} = \sum_{i=1}^N y_i \alpha_i \vec{x}_i \quad (3.30)$$

$$b = \vec{w} \cdot \vec{x}_k - y_k \quad (3.31)$$

Ancak, her örneğin lineer olarak ayırt edilemeyeceğini de unutmamak lazım. Yani pozitif ve negatif örnekleri birbirinden ayıran hiperdüzlem bulunamayabilir. Dolayısıyla üstte verilen formül ile ayırt edilemeyen durum için sonsuz çözüm üretir. Ancak, 1995 yılında Cortes ve Vapnik [41] tarafından formül Eş. 3.27 ile verilen optimizasyon ifadesini geliştirilerek ayırt edilemeyen durumlar için uygun hale getirilmiştir. Bunun için, ayırt etme marjini içerisinde kalan bazı örnekler için hata oranını hesaplayarak ceza verme yöntemi uygulanmıştır [42], [43]. Böylece, Eş. 3.27 ile ifade edilen optimizasyon problemi lineer olmayan olarak değiştirilmiştir ve Eş. 3.32'de verilmiştir.

$$y_i(\vec{w} * \vec{x}_i - b) \geq 1 - \xi_i, \forall i \text{ için}$$

$$\min_{w,b,\xi} \frac{1}{2} \|\vec{w}\|^2 + C \sum_{i=1}^N \xi_i \quad (3.32)$$

Buradaki, ξ_i değişkeni ayırt etme marjininin hata oranını verir ve C parametresi ise küçük hata oranını kayda alarak ayırt etme marjini genişletmede yardımcı olmaktadır. Eğer bu optimizasyon problemini Eş. 3.28'e göre dual forma dönüştürürsek, basit bir değişiklik yapılması gerekmektedir ve $\alpha_i \geq 0$ kısıtlama şartı aşağıdaki gibi değişmektedir.

$$0 \leq \alpha_i \leq C \quad (3.33)$$

Hata oranını belirleyen ξ_i değişkeni tüm dual formu etkilememektedir. Böylece DVM lineer olmayan sınıflandırıcı olarak da kullanılabilir. Lineer olmayan DVM çıktısı Lagrange çarpanları ile açık bir şekilde Eş. 3.34'te verilmiştir.

$$u = \sum_{j=1}^N y_j \alpha_j K(\vec{x}_j, \vec{x}) - b \quad (3.34)$$

Buradaki, K giriş vektörü, $\vec{x}$ eğitim için gelen $\vec{x}_j$ vektörü arasındaki benzerlik veya aralarındaki uzaklığı veren çekirdek (kernel) fonksiyonudur. Gauss, polinomial ve sinir ağları K kernel fonksiyonunun lineer olmayan türleridir. Eğer K fonksiyonu lineer ise DVM için Eş. 3.25 ile verilmiş formül geçerlidir [43].

Lagrange çarpanları α_i Eş. 3.28'den görüldüğü gibi QP ile hesaplanmaktadır. Lineer olmaması karesel formu değiştirir, ancak dual Ψ amaç fonksiyonu α için değişmemektedir.

$$\min_{\vec{\alpha}} \Psi(\vec{\alpha}) = \min_{\vec{\alpha}} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j (\vec{x}_i \cdot \vec{x}_j) \alpha_i \alpha_j - \sum_{i=1}^N \alpha_i \quad (3.35)$$

$$0 \leq \alpha_i \leq C$$

$$\sum_{i=1}^N y_i \alpha_i = 0$$

Her optimizasyon probleminde olduğu gibi söz konusu QP problemi için de gereklilik ve yeterlilik koşulu yerine getirilmeli ve bu şart için Karush-Kuhn-Tucker (KKT) koşulu geçerlidir [44]. Eş. 3.35 için KKT şartı çok basittir ve QP problemi her bir i için aşağıdaki sınırlama koşulları geçerli olursa gereklilik ve yeterlilik koşulu sağlanmış olur.

$$\alpha_i = 0 \Leftrightarrow y_i u_i \geq 1$$

$$0 \leq \alpha_i \leq C \Leftrightarrow y_i u_i = 1$$

$$\alpha_i = C \Leftrightarrow y_i u_i \leq 1$$

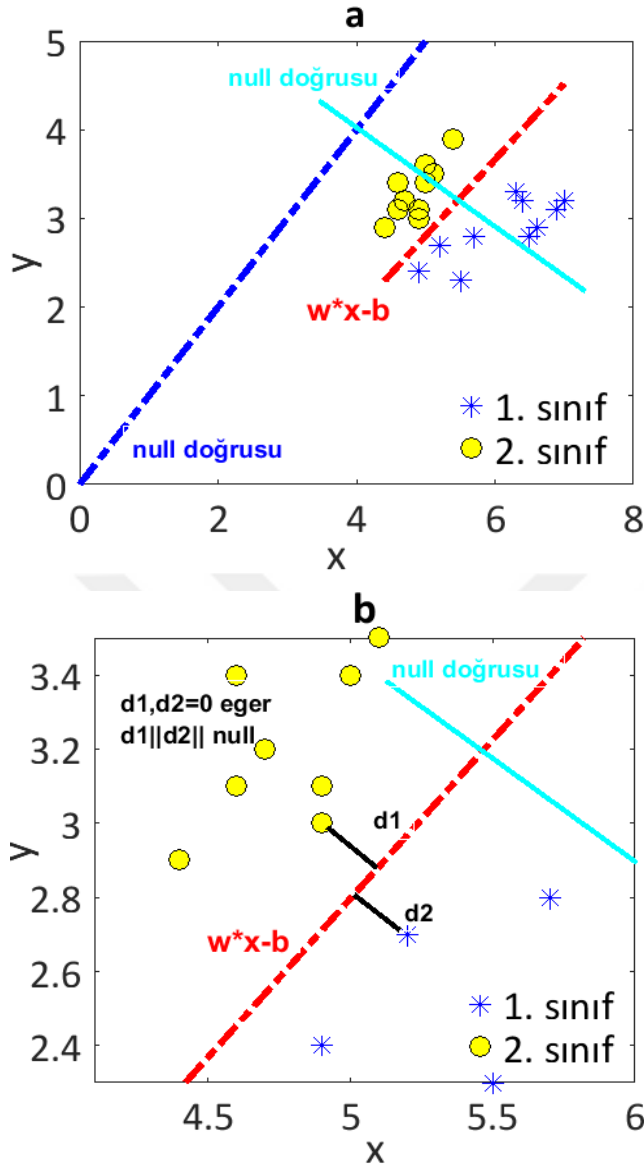
Buradaki u_i DVM'nin i . eğitim örneği için ürettiği sonuçtur. Eş. 3.35 ile verilen QP problemi Sequential Minimal Optimization (SMO) yöntemi ile çözülmektedir [41]. SMO algoritması da bu şartlar oluşana kadar belirli bir *tolerans* değerine göre en iyi çözümü bulmak için devam etmektedir.

3.6.2. Destek Vektör Makinesi ve Lorentz uzayı

Destek Vektör Makinesi iki sınıflı örnekler için çok başarılı çalışan bir yöntemdir [41]. Bu nedenle DVM algoritmasının Lorentz uzayında sınıflandırıcı olarak kullanma fikri doğmuştur. Ancak, Lorentz uzayının özel yapısal özellikleri ve Destek Vektör Makinesinin eğitim ve sınıflandırma aşamalarında kullanmış olduğu işlemler göz önünde bulundurularak inceleme yapıldığında aşağıdaki tespitler ortaya çıkmıştır:

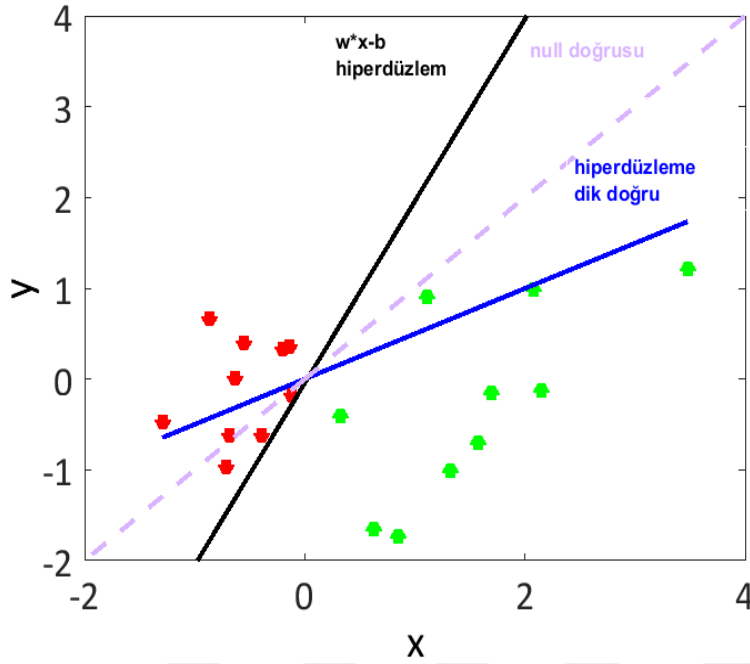
1. *Tespit:* DVM algoritması eğitim esnasında ayırt edilebilir örnekler için Eş. 3.27 ve ayırt edilemeyen örnekler için ise Eş. 3.32 formüllerle verilmiş olan optimizasyon problemini çözmeye çalışmaktadır. Yani, hiperdüzlemden en yakın destek vektörlere kadar olan $d = \frac{1}{\|w\|}$ uzaklığını maksimize etmeye çalışmaktadır. Bu uzaklık hiperdüzleme dik açıyla en yakın destek vektörü bağlayan doğrudur (normal vektör). Lorentz uzayının Eş. 2.6 ile verilmiş olan uzaklık metriğine göre bu doğrunun uzunluğu eğer null eksenine paralel veya dik açıyla kesişiyorsa sıfıra eşittir. Bu durum optimizasyon problemini, dolayısıyla DVM algoritmasını geçersiz kılmaktadır. Anlatılan durumun görselleştirilmesi Şekil 3.14'te verilmiştir.

2. *Tespit:* 1. tespitten yola çıkarak araştırmaya Destek Vektör Makinesinin içyapısını inceleyerek devam edilmiştir. Bu bağlamda Destek Vektör Makinesi eğitim aşamasında Eş. 3.27 ve Eş. 3.32 ile verilmiş olan optimizasyon problemlerini çözmek için kullandığı Eş. 3.28 ile verilmiş olan QP problem örneklemede iç çarpım içermektedir. Bu durum yeni gelen test örneklerini sınıflandırmak için DVM yönteminin çıktısı olarak Eş. 3.34'deki formülde K kernel fonksiyonu olarak verilmiştir. İç çarpım Lorentz uzayının metriğine göre iç çarpım Eş. 2.9 ile yapılmaktadır ve bu durum Eş. 3.27 ve Eş. 3.32 formüllerle verilmiş olan optimizasyon problemini Lorentz uzayında geçerli konuma getirerek DVM makinesinin kullanımı olanağını arttırmıştır.



Şekil 3.14. Sınıfları ayıran hiperdüzlem ve null doğrusu ile dik açıyla kesişen doğru

Çünkü, Lorentz uzayında diklik kavramı Öklid uzayına göre çok farklıdır, bunun sebebi ise Lorentz uzayının sahip olduğu özel yapı ve iç çarpımıdır. Lorentz uzayındaki diklik kavramı Şekil 3.15'te verilmiştir.



Şekil 3.15. Lorentz uzayındaki diklik

3.6.3. Lorentz metriği temelli Destek Vektör Makinesi

Destek Vektör Makinesinin yapısını ve çalışmasını Lorentz uzayının özelliklerine ve metriğine göre yapılandırmada bu tez araştırmasının 3.6.2. bölümünde belirtilen *Tespitleri* baz alarak Lorentz metriği ile sınıflandırma yapan DVM sınıflandırma algoritması geliştirilmiştir. Geliştirilen yöntemde Eş. 3.27 ile verilmiş olan ve eğitim aşamasının ana adımını oluşturan optimizasyon problemi Lorentz uzayının metriğine göre yapılandırılmıştır. Yani, Eş. 3.27 ile verilmiş olan optimizasyon probleminin Lagrange operatörü ile dönüştürülmüş hali olan ve Eş. 3.28 ile verilen dual formdaki Ψ amaç fonksiyonu iç çarpım içerdiği için Lorentz uzayının Eş. 2.9 ile verilen metriğine göre yapılandırılmıştır. Böylece, Lorentz uzayında α_i Lagrange çarpanlar kümesini hesaplayan Ψ_M amaç fonksiyonunun çözümü Eş. 3.36 ile edilmektedir.

$$\min_{\vec{\alpha}} \Psi_M(\vec{\alpha}) = \min_{\vec{\alpha}} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j (\vec{x}_{N-1} \cdot \vec{x}_{N-1} - \vec{x}_N \vec{x}_N) \alpha_i \alpha_j - \sum_{i=1}^N \alpha_i \quad (3.36)$$

Destek Vektör Makinesi sınıflandırma algoritmasının çıktısını üreten veya yeni test örneklerin sınıf etiketini belirleyen Eş. 3.34 ile verilmiş olan ifade iç çarpım işlemin

barındırdığından Lorentz metriğine göre yapılandırılmıştır. Böylece, her bir yeni y_j test örneğini Lorentz uzayında lineer kernel fonksiyonu ile sınıf etiketini belirleyen hesaplama Eş. 3.37’de verilmiştir.

$$u_M = \sum_{j=1}^N y_j \alpha_j (x_j \cdot x - x_N x) - b \quad (3.37)$$

Böylece, Lorentz uzayında destek vektör ve Lorentz metriği temelinde geliştirilen yeni Lorentz Destek Vektör Makinesi sınıflandırma algoritmasının adımları Algoritma-11’de verilmiştir.

Algoritma 11. Lorentz DVM (LDVM)

Girdi: Test, eğitim kümeleri $(X, Y) \in \mathbb{R}$

Adım 1. İkili öznitelik alt kümelerini $(X' = X)$ oluştur

Adım 2. X' ’yığınınından k adet öznitelik ikilisini (X'') seç

Adım 3. X'' için $\min_{w,b} \frac{1}{2} \|\vec{w}\|^2$ hesapla ve M makineyi eğit

Adım 4. Her yeni gelen test örneği $y_i \in Y$ için

M ile U_i sınıf etiketini belirle

Adım 5. Sınıflandırma oranını hesapla

4. DENEYSEL SONUÇLAR VE BULGULAR

Bu bölümde Lorentz uzayında sınıflandırma işlemini gerçekleştirmek amacıyla geliştirilen yöntemlerin deneysel sonuçları verilmiştir. Deneysel sonuçlar genel kullanıma açık olan veri kümeleri üzerinde yapılan denemelerde elde edilmiştir. Ayrıca, önerilen yöntemlerin geçerliliğini ve doğruluğunu test etmek amacıyla elde edilen sayısal sonuçlar klasik sınıflandırıcıların sonuçları ile karşılaştırılarak analiz edilmiştir.

Bu tez çalışmasında, Lorentz metriğinin sınıflandırma probleminde kullanımı ile ilgili geliştirilen ve Bölüm 3'te yer alan yöntemler “tekten genele” mantığı ile elde edilmiştir. Genel olarak, bu bölümde yer alan deneysel sonuçlar da aynı sırayı takip etmektedir. Yani, ilk önce iki boyutlu veri kümelerinin Lorentz metriği temelinde sınıflandırılması ile ilgili yöntemlerin sonuçları verilmiştir. Sonra, ikiden çok boyutlu veri kümelerinin Lorentz uzayında sınıflandırılması ile ilgili sonuçlar yer almaktadır. Onları, Lorentz Yüz Doğrulama (LYD) ve Lorentz Destek Vektör Makinesi (LDVM) yöntemleri ile elde edilen sonuçlar takip etmektedir.

4.1. İki Boyutlu Veriler İçin Lorentz Sınıflandırıcısı Sonuçları

İki boyutlu Lorentz uzayı için geliştirilen sınıflandırma algoritmalarının sınıflandırma başarılarını test etmek amacıyla genel kullanıma açık CLIMATE, ECOLI, GESTURE, ILPD, PARKINSON, RELAX, SEEDS, TELESCOPE, VERTEBRAL, WINE, WISCONSIN veri kümeleri kullanılmıştır. İki boyutlu Lorentz uzayı için veriler ikili öznitelik gruplarına ayrılarak sınıflandırma yapılmıştır. Yapılan ön denemelerde, Lorentz uzaklığı ile elde edilen sınıflandırma sonuçlarının veri kümesinin dağılımına göre değişiklik gösterdiği ortaya çıkmıştır. Yani, Lorentz uzaklık ölçütü ile çalışan sınıflandırıcının, örneklerin dağılımına hassas olduğu ve bazen diğer yöntemlere göre daha iyi sonuç üretmediği görülmüştür. Bu sorunun ortadan kaldırılması için “öznitelik yapılandırılması” (feature representation), yani örneklerin uygun hale getirilmesi için Önişlem-1 ve Önişlem-2'de belirtilen adımlar uygulanmıştır.

Lorentz uzaklığı temelli sınıflandırma

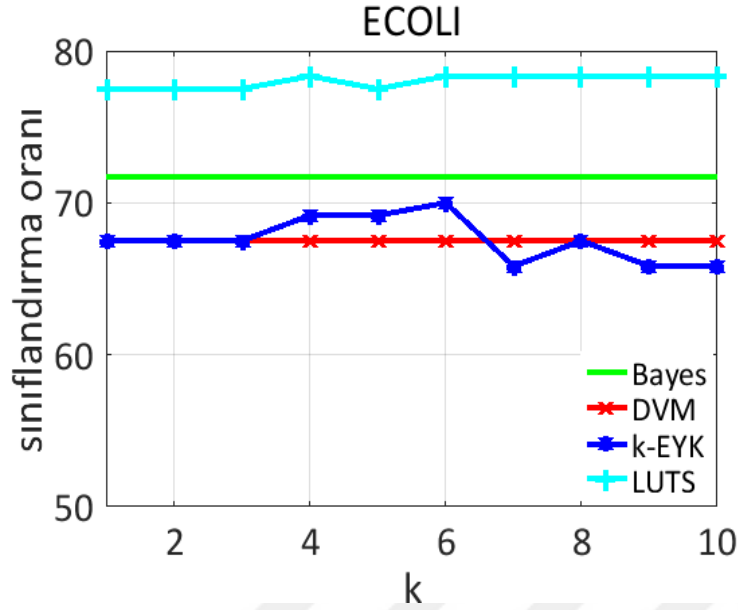
Bu tez çalışmasının 3.3.1. bölümünde önerilen Lorentz Uzaklığı Temelli Sınıflandırma (LUTS) algoritmasının sınıflandırma problemine uygulanabilirliğini ve sınıflandırma başarısını görmek amacıyla ECOLI, ILPD, SEEDS, VERTEBRAL veri kümeleri ile denemeler yapılmıştır. Seçilen veri kümelerinde öznitelik sayısı 6 ile 13 arasında değişmektedir (Çizelge 2.1. bkz). Denemelerde geliştirilen LUTS yöntemi için bu öznitelikler rasgele iki öznitelik kombinasyonları seçilerek kullanılmıştır. Uygulamada kullanılan veri kümelerindeki eğitim ve test örneklerinin dağılımı, ayrıca denemelerde kullanılan ikili öznitelik grupları ile ilgili bazı bilgiler Çizelge 4.1’ de verilmiştir.

Çizelge 4.1. LUTS yönteminde kullanılan veri kümeleri ile ilgili istatistiksel bilgiler

Veri kümeleri	Eğitim örnek sayısı	Test örnek sayısı	Kullanılan öznitelikler
ECOLI	80	140	1 ve 2
ILPD	140	223	2 ve 10
SEEDS	46	94	3 ve 7
VERTEBRAL	110	200	2 ve 3

Yapılan deneysel çalışmada geliştirilen LUTS yönteminin ürettiği sonuçlar diğer klasik Bayes, DVM ve k-EYK sınıflandırma algoritmalarının sonuçları ile karşılaştırılmıştır. Denemelerde ECOLI veri kümesinden seçilen (1 ve 2) öznitelik grubu için Bayes yöntemi %71.67 oranında sınıflandırma başarısı verirken, aynı durum için DVM ve k-EYK yöntemlerinin sınıflandırma başarısı sırasıyla %67.50 ve %70.00 olarak elde edilmiştir. Önerilen LUTS yöntemi ECOLI veri kümesi için diğer yöntemlerden daha iyi olmak üzere %78.33 oranında en iyi sınıflandırma başarısını üretmektedir. Bu sonuçlardan görüldüğü gibi ECOLI veri kümesi için önerilen LUTS yöntemi klasik sınıflandırıcılardan en yüksek başarıyı üreten Bayes yöntemine göre %7 oranında daha iyi sonuç vermiştir. Ayrıca, önerilen LUTS yöntemi ECOLI için en iyi sonucu $k=4$ için verirken, Öklid uzayında en yakın komşu tekniği ile çalışan k-EYK yöntemi en iyi sınıflandırma sonucunu $k=6$ iken üretmektedir. Bu durum önerilen LUTS yönteminin Lorentz metriği ile bazı veri kümeleri için daha az sayıdaki komşular ile en yüksek sonucu üretebileceğini göstermektedir. ECOLI veri kümesi

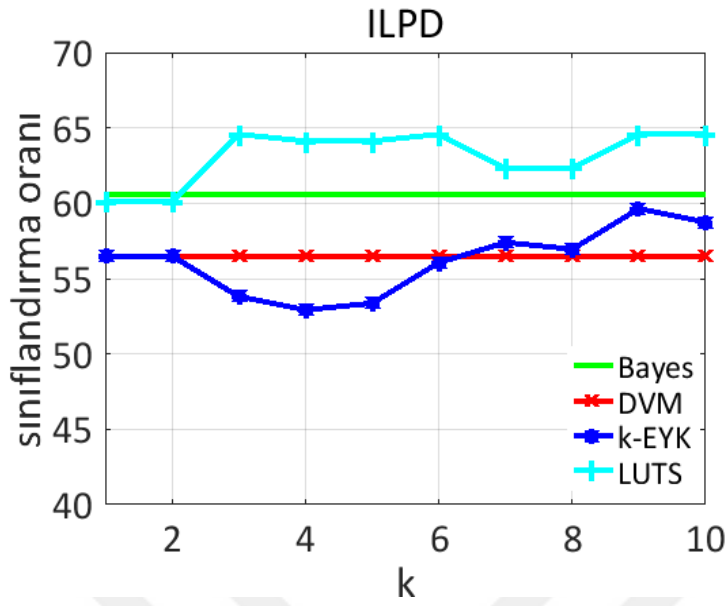
için k değerinin değişimindeki LUTS yönteminin çıktılarının diğer klasik yöntemlerin sonuçları ile görsel karşılaştırması Şekil 4.1’de verilmiştir. Bayes ve DVM yöntemleri k parametresinden bağımsız olduğu için düz çizgi ile gösterilmiştir.



Şekil 4.1. k parametresinin değişimine göre ECOLI veri kümesi için LUTS ve sınıflandırıcıların başarı grafiği

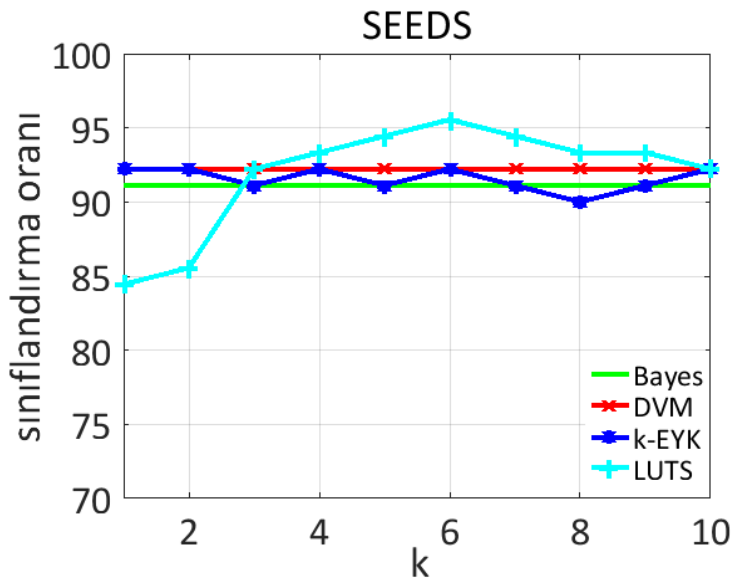
ILPD veri kümesinden seçilen (2 ve 10) öznitelik grubu için Bayes sınıflandırıcısı %60.54 oranında başarı üretmektedir. Bu sonuç aynı zamanda ILPD için LUTS yöntemi ile %64.57 oranında üretilen en yüksek sonuca en yakın sonuçtur. Aynı durum için DVM ve k-EYK yöntemleri sırasıyla %56.50 ve %59.64 oranında başarı göstermiştir. ILPD veri kümesi için LUTS yöntemi $k=3$ iken en iyi sınıflandırma başarısını vererek aynı durum için k-EYK yönteminin $k=9$ için ürettiği sınıflandırma sonucunu %5 oranında geçmiştir. ILPD veri kümesi için denemelerde elde edilen sayısal sonuçların görsel karşılaştırması Şekil 4.2’de verilmiştir.

SEEDS veri kümesindeki örneklerin dağılımına bağlı olarak klasik Bayes, DVM ve k-EYK sınıflandırıcıları bu veri kümesinden seçilen (3 ve 7) öznitelik ikilisi için oldukça yüksek başarı üretmektedirler. Bayes sınıflandırıcısı SEEDS için %91.11 oranında başarı gösterirken, DVM ve k-EYK sınıflandırıcıları %92.22 oranında sonuç üretmiştir. Denemelerde SEEDS veri kümesi için LUTS yönteminin %95.56 oranında çıktı üreterek en iyi sınıflandırma başarısını verdiği görülmüştür.



Şekil 4.2. k parametresinin değişimine göre ILPD veri kümesi için LUTS ve diğer sınıflandırıcıların başarı grafiği

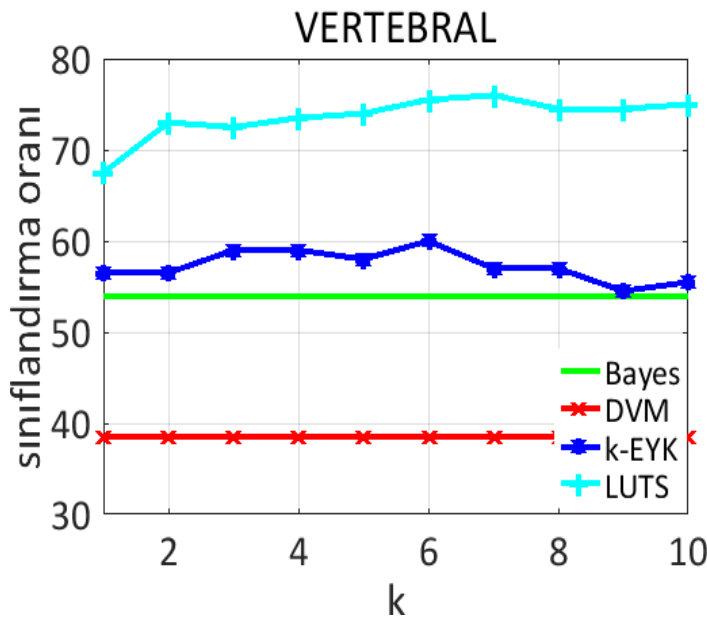
Bu sınıflandırma başarısı Lorentz uzaklık ölçütü temelinde çalışan LUTS yöntemi ile $k=6$ iken elde edilmiştir. SEEDS için k-EYK yönteminin en iyi sınıflandırma sonucu $k=1$ iken elde edilmiştir. SEEDS veri kümesi için k değerinin değişiminin sınıflandırma başarılarına etkisi görsel olarak Şekil 4.3'te verilmiştir.



Şekil 4.3. k parametresinin değişimine göre SEEDS veri kümesi için LUTS ve diğer sınıflandırıcıların başarı grafiği

İki boyutlu Lorentz uzayında sınıflandırma yapan LUTS yönteminin denenmesinde kullanılan son veri kümesi VERTEBRAL için en düşük sonuç DVM yöntemi ile %38.50

oranında elde edilmiştir. Aynı durum için Bayes sınıflandırıcısı %54 oranında sınıflandırma başarısı üretmiştir. VERTEBRAL veri kümesi için k-EYK yönteminin en yüksek sınıflandırma çıktısı k=6 iken %60 oranında elde edilmiştir. LUTS yöntemi bu sonucu %16 oranında arttırarak VERTEBRAL için en iyi sınıflandırma başarısı k=7 iken %76 olarak elde edilmiştir. Denemelerde VERTEBRAL veri kümesinin (2 ve 3) ikili öznitelik grupları kullanılmıştır ve elde edilen sayısal sonuçlar k değerinin değişimine göre Şekil 4.4'te verilmiştir.



Şekil 4.4. k parametresinin değişimine göre VERTEBRAL veri kümesi için LUTS ve diğer sınıflandırıcıların başarı grafiği

Denemelerde kullanılan ECOLI, ILPD, SEEDS, VERTEBRAL veri kümeleri için klasik Bayes, DVM ve k-EYK sınıflandırıcıları ve önerilen LUTS yöntemi ile elde edilen en iyi sınıflandırma başarıları sayısal olarak Çizelge 4.2'de verilmiştir.

Çizelge 4.2. LUTS yönteminin en iyi sınıflandırma sonuçlarının karşılaştırması

Veri kümeleri	Sınıflandırıcılar			
	Bayes	DVM	k-EYK	LUTS
ECOLI	71.67	67.50	70.00	78.33
ILPD	60.54	56.50	59.64	64.57
SEEDS	91.11	92.22	92.22	95.56
VERTEBRAL	54.00	38.50	60.00	76.00

LUTS ve k-EYK yöntemlerinin k komşu sayısına göre ürettikleri en iyi sınıflandırma başarılarının sayısal karşılaştırması Çizelge 4.3'te verilmiştir.

Çizelge 4.3. LUTS yönteminin k değerine göre sınıflandırma başarılarının karşılaştırması

Veri kümesi	Sınıflandırıcı			
	k	k-EYK	k	LUTS
ECOLI	6	70.00	4	78.33
ILPD	9	59.64	3	64.57
SEEDS	1	92.22	6	95.56
VERTEBRAL	6	60.00	7	76.00

Genel olarak iki boyutlu Lorentz uzayında Lorentz metriği ile sınıflandırma yapan LUTS yöntemi ile ECOLI, ILPD, SEEDS ve VERTEBRAL veri kümeleri üzerinde yapılan denemelerde elde edilen sonuçlara göre önerilen yöntemin diğer klasik k-EYK, Bayes ve DVM yöntemlerine göre daha başarılı sınıflandırma sonucu ürettiği görülmüştür.

Bir karar çizgisine göre ayarlanmış Lorentz sınıflandırıcısı

Bu tez çalışmasının 3.3.3. bölümünde anlatılan ve Algoritma-3 ile tanımlanan yöntemde önışlem aşamasında en iyi döndürme açısını belirleyen karar doğrusu Destek Vektör Makinesi veya Doğrusal Diskriminant Analizi yöntemleri ile hesaplanmaktadır. Daha sonra, hesaplanan karar çizgilerinin yardımı ile veriler önce Lorentz metriği açısından en uygun pozisyona taşınmakta ve sonra Lorentz uzayında sınıflandırma yapılmaktadır. Bu adımların bütünü Karar Çizgisine Göre Ayarlanmış Lorentz Sınıflandırıcısını (KÇ-GALS) oluşturmaktadır. Karar doğrularını hesaplama ve verilerin en iyi pozisyona taşınması Algoritma-3'te önışlem olarak tanımlanmıştır. Bu bölümde, KÇ-GALS yönteminin sınıflandırma başarısını test etmek amacıyla RELAX, SEEDS, TELESCOPE, VERTEBRAL ve WISCONSIN veri kümeleri üzerinde denemeler yapılmıştır. Denemelerde önerilen KÇ-GALS yöntemi ile elde edilen sayısal sonuçlar ve literatürdeki klasik k-EYK, Bayes ve DVM sınıflandırıcılarının sonuçları karşılaştırılarak değerlendirilmiştir.

Denemelerde kullanılan RELAX, SEEDS, TELESCOPE, VERTEBRAL ve WISCONSIN veri kümelerindeki özniteliklerin sayısı 6-33 aralığında değişmektedir.

Denemelerde kullanılan veri kümeleri ile ilgili bazı istatistiksel bilgiler Çizelge 4.4'te verilmiştir.

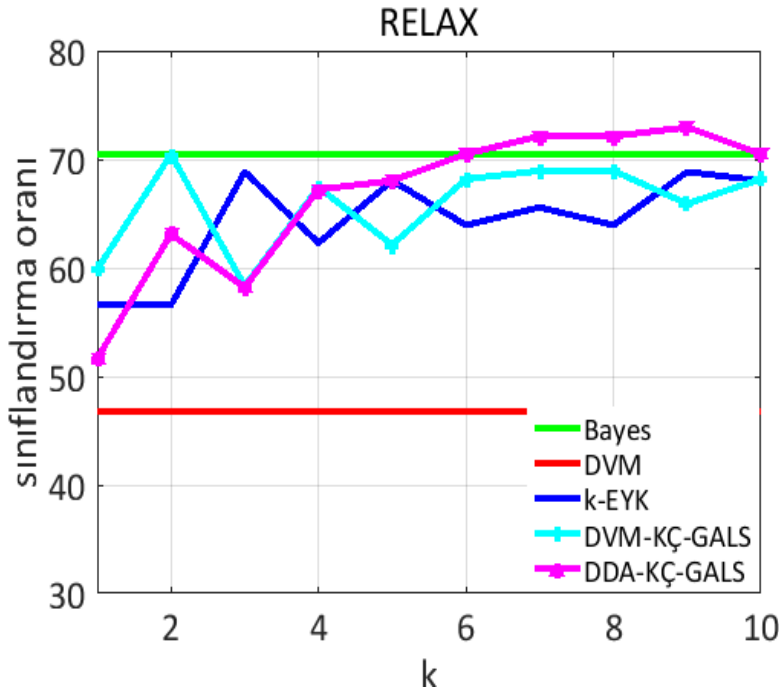
Çizelge 4.4. KÇ-GALS yönteminde kullanılan veri kümeleri ile ilgili istatistiksel bilgiler

	Eğitim örnek sayısı	Test örnek sayısı	Kullanılan öznitelikler
RELAX	60	122	2 ve 5
SEEDS	46	94	4 ve 7
TELESCOPE	134	266	2 ve 7
VERTEBRAL	110	200	2 ve 4
WISCONSIN	66	132	2 ve 7

KÇ-GALS yöntemi farklı sınıflara ait örnekleri ayırt etmede k-En Yakın Komşu yöntemine benzer teknikle çalışmaktadır, yani en yakın komşu sayısı ve aralarındaki Lorentz uzaklığına göre sınıflandırma yapmaktadır. Bu bağlamda denemelerde k değerinin 1 ile 10 arasında değişimindeki KÇ-GALS yönteminin hem DVM karar çizgisi, hem de DDA karar çizgisinin yardımı ile elde edilen sınıflandırma sonuçları analiz edilmiştir. Ayrıca, önerilen DVM-KÇ-LS ve DDA-KÇ-GALS yöntemleri ile elde edilen sınıflandırma sonuçları k-EYK, Bayes ve DVM gibi klasik sınıflandırıcıların çıktıları ile karşılaştırılmıştır.

Önerilen KÇ-GALS yönteminde DVM veya DDA yöntemlerinin yardımı ile hesaplanan karar çizgisini kullanarak örnekleri null eksenine göre en iyi pozisyona taşımak için döndürme açısının hesaplanması Önişlem-2'de anlatılmaktadır. Elde edilen en iyi döndürme açısı ile KÇ-GALS yöntemi en iyi sınıflandırma sonucunu vermektedir. Ancak, veri kümelerinin dağılımına göre DVM ve DDA yöntemleri ile hesaplanan karar çizgisinin pozisyonu ve bunlar ile elde edilen en iyi döndürme açıları farklılık göstermektedir. Bu farkın görsel karşılaştırması Şekil 3.8'de verilmiştir. Örneğin, RELAX veri kümesinden seçilen (2 ve 5) öznitelik ikilisi için en iyi döndürme açısı DVM ve DDA ile hesaplanan karar çizgileri ile sırasıyla 33.19° ve 25.87° olarak bulunmuştur. Böylece, DVM ile hesaplanan karar çizgisinin gösterdiği 33.19° açısıyla RELAX veri kümesini döndürdüğümüzde DVM-KÇ-GALS yöntemi $k=2$ iken en iyi sınıflandırma sonucunu %70.45 oranında üretmektedir. DDA ile hesaplanan karar çizgisinin gösterdiği 25.87° değerindeki en iyi döndürme açısıyla DDA-KÇ-GALS yöntemi $k=9$ iken RELAX veri

kümesi için %73.00 değerinde çıktı üreterek en yüksek sınıflandırma başarısını vermiştir. RELAX veri kümesi için klasik Bayes ve DVM yöntemleri sırasıyla %70.49 ve %46.72 oranında başarı üretmişlerdir. Aynı durum için k-EYK yöntemi $k=3$ iken %68.85 oranında sınıflandırma sonucu üretmiştir. Bu tez çalışmasında geliştirilen DVM-KÇ-GALS ve DDA-KÇ-GALS yöntemlerinin k değerinin değişiminde ürettiği sınıflandırma başarısının görsel karşılaştırması Şekil 4.5'te verilmiştir.



Şekil 4.5. RELAX veri kümesi için k değerlerine göre sınıflandırma oranları

Şekil 4.5'ten görüldüğü gibi k değerinin değişmesi durumunda önerilen DVM-KÇ-GALS ve DDA-KÇ-GALS yöntemlerinin de sınıflandırma başarısı değişmektedir. Elde edilen sayısal sonuçlardan görüldüğü gibi iki boyutlu Lorentz uzayında örneklerin sınıflandırılması amacıyla geliştirilen Önişlem-2'de DVM karar çizgisinin kullanılması durumunda RELAX veri kümesi için DVM-KÇ-GALS sınıflandırma başarısı Bayes yönteminin çıktısı ile yakın oranda kaydedilmiştir. Ancak, DDA-KÇ-GALS yöntemi RELAX veri kümesi için klasik yöntemlerden en yüksek çıktıyı üreten Bayes metodunun çıktısını %3 oranında arttırmıştır. Bu da geliştirilen Önişlem-2 adımlarının doğruluğunu ve DVM ile DDA yöntemlerinin örneklerinin dağılımından farklı derecede etkilendiğini göstermektedir.

Denemelerde DVM-KÇ-GALS, DDA-KÇ-GALS ve k-EYK yöntemlerinin RELAX veri

kümesi için en yüksek sınıflandırma sonuçları farklı sayıdaki komşu sayılarına göre kaydedildiği tespit edilmiştir. Deneylerde k değerinin değişimindeki sınıflandırma sonuçları ve KÇ-GALS yöntemi ile en yüksek sınıflandırma sonucunu üreten k değeri sayısal olarak Çizelge 4.5'te verilmiştir.

Çizelge 4.5. RELAX için k komşu sayısına göre sınıflandırma sonuçları

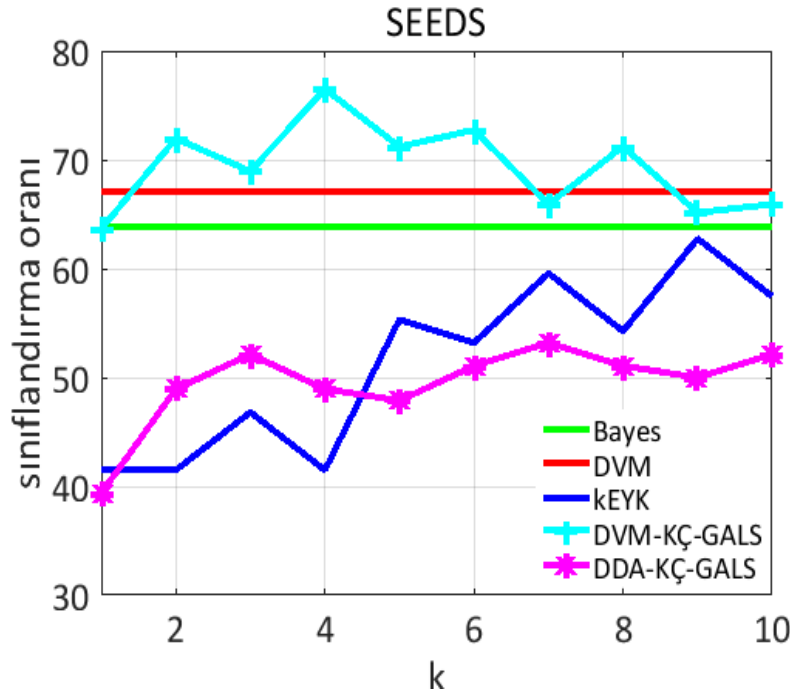
	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
Bayes	70.49	70.49	70.49	70.49	70.49	70.49	70.49	70.49	70.49	70.49
DVM	46.72	46.72	46.72	46.72	46.72	46.72	46.72	46.72	46.72	46.72
k-EYK	56.55	56.56	68.85	62.30	68.03	63.93	65.57	63.93	68.85	68.03
DVM-KÇ-GALS	59.85	70.45	58.33	67.42	62.12	68.18	68.94	68.94	65.91	68.18
DDA-KÇ-GALS	51.64	63.11	58.20	67.21	68.03	70.49	72.13	72.13	73.00	70.49

DVM-KÇ-GALS ve DDA-KÇ-GALS yöntemlerinin performansını değerlendirmek amacıyla kullanılan SEEDS veri kümesinden seçilen (4 ve 7) öznitelikler ikilisi için Bayes ve DVM sınıflandırıcıları sırasıyla %63.83 ve %67.02 oranında başarı kaydetmiştir. Aynı durum için k-EYK yöntemi k=9 iken %62.76 oranında sınıflandırma sonucunu üretmiştir. Bu çalışmada geliştirilen DDA-KÇ-GALS yöntemi SEEDS veri kümesini hesapladığı en iyi 60.52° açığa döndürerek en iyi sınıflandırma başarısını k=7 iken %53.19 oranında kaydetmiştir. Bu sonuç diğer klasik sınıflandırıcıların ürettiği sonuçlara göre oldukça düşüktür. Ancak, SEEDS kümesini 39.51° açığa döndüren DVM-KÇ-GALS yöntemi k=4 iken diğer yöntemlerden en yüksek sonucu üreten DVM çıktısını %9 oranında arttırarak ve %76.51 oranında çıktı üreterek en iyi sınıflandırma sonucunu kaydetmiştir. Bu durum bu araştırmada önerilen Önişlem-2'nin geçerliliğini ve doğruluğunu göstermektedir. SEEDS veri kümesi için elde edilen sınıflandırma sonuçlarının görselleştirilmesi Şekil 4.6'da verilmiştir.

Denemelerde SEEDS için KÇ-GALS yöntemi ile elde edilen sınıflandırma sonuçlarının k değerinin değişimindeki değerleri ve KÇ-GALS ile en yüksek sınıflandırma sonucunu üreten k değeri sayısal olarak Çizelge 4.6'te verilmiştir.

Çizelge 4.6. SEEDS için k komşu sayısına göre sınıflandırma sonuçları

	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
Bayes	63.83	63.83	63.83	63.83	63.83	63.83	63.83	63.83	63.83	63.83
DVM	67.02	67.02	67.02	67.02	67.02	67.02	67.02	67.02	67.02	67.02
k-EYK	41.48	41.48	46.80	41.48	55.31	53.19	59.54	54.25	62.76	57.44
DVM-KÇ-GALS	63.63	71.96	68.93	76.51	71.21	72.72	65.90	71.21	65.15	65.90
DDA-KÇ-GALS	39.36	48.93	52.12	48.93	47.87	51.06	53.19	51.06	50.00	52.12



Şekil 4.6. SEEDS veri kümesi için k değerlerine göre sınıflandırma oranları

Denemelerde TELESCOPE veri kümesinden seçilen (2 ve 7) öznitelik ikilisi için Bayes yöntemi %53.38 oranında sınıflandırma başarısı kaydetmiştir. Bayes yönteminin çıktısına çok yakın sonuç DVM sınıflandırıcısı tarafından %53.76 oranında kaydedilmiştir. Aynı durum için k-EYK algoritması k=1 iken %62.78 oranında sınıflandırma başarısı sergilerken, Algoritma 3. ile tanımlanan DDA-KÇ-GALS ve DVM-KÇ-GALS yöntemleri sırasıyla %69.67 ve %68.42 oranlarında sınıflandırma sonuçlarını üretmiştir. Bu sonuçlardan görüldüğü gibi, TELESCOPE veri kümesinde SEEDS ve RELAX durumlarından farklı olarak bu çalışmada önerilen DDA-KÇ-GALS ve DVM-KÇ-GALS yöntemlerinin her ikisi de klasik sınıflandırıcılara göre daha başarılı sonuç üretmektedirler. Ayrıca, TELESCOPE

veri kümesinin örneklerini DDA-KÇ-GALS yöntemi 13.62° açığa döndürerek $k=6$ iken en iyi sonucu üretmiştir. DVM-KÇ-GALS yönteminde ise en iyi sınıflandırma sonucunu elde edebilmek için örnekler 43.17° açısıyla döndürülmüştür. Bununla birlikte DVM-KÇ-GALS yöntemi tek komşuya ($k=1$) bakarak klasik yöntemlere göre daha başarılı sınıflandırma sonucu üretmektedir.

TELESCOPE için denemelerde önerilen DDA-KÇ-GALS ve DVM-KÇ-GALS yöntemleri ile elde edilen sayısal sonuçların karşılaştırılması Çizelge 4.7’de verilmiştir.

Çizelge 4.7. TELESCOPE için k komşu sayısına göre sınıflandırma sonuçları

	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
Bayes	53.38	53.38	53.38	53.38	53.38	53.38	53.38	53.38	53.38	53.38
DVM	53.76	53.76	53.76	53.76	53.76	53.76	53.76	53.76	53.76	53.76
k-EYK	62.78	62.78	54.13	59.77	50.75	53.75	49.62	53.38	49.24	54.13
DVM-KÇ-GALS	57.58	65.91	57.58	68.94	66.67	69.67	61.36	65.91	65.91	66.67
DDA-KÇ-GALS	68.42	58.65	57.14	55.26	52.63	49.25	52.63	55.26	52.26	55.64

VERTEBRAL veri kümesi için Bayes ve DVM yöntemleri sırasıyla %43.48 ve %37.20 oranlarında sınıflandırma başarısı üretmişlerdir. Aynı durum için k-EYK yöntemi $k=4$ iken %58.45 oranında sınıflandırma başarısı kaydetmiştir. TELESCOPE veri kümesinden elde edilen (2 ve 4) öznitelikleri için DDA karar çizgisi temelinde Lorentz uzayında sınıflandırma yapan DDA-KÇ-GALS yöntemi örnekleri 80.26° açığa döndürerek %62.80 oranında sonuç üretmiştir. Aynı durum için örnekleri en iyi pozisyona taşımak için DVM karar çizgisinin yardımı ile hesaplanan 44.33° açığa döndüren DVM-KÇ-GALS yöntemi $k=6$ iken en iyi sınıflandırma başarısını %69.67 oranında üretmiştir. Böylece, önerilen DVM-KÇ-GALS yöntemi k-EYK ve DDA-KÇ-GALS yöntemlerinin çıktılarını sırasıyla %11 ve %7 oranda arttırmıştır.

VERTEBRAL veri kümesi için k parametresinin farklı değerlerinde elde edilen sınıflandırma sonuçlarının sayısal karşılaştırılması Çizelge 4.8’de verilmiştir.

Çizelge 4.8. VERTEBRAL için k komşu sayısına göre sınıflandırma sonuçları

	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
Bayes	43.48	43.48	43.48	43.48	43.48	43.48	43.48	43.48	43.48	43.48
DVM	37.20	37.20	37.20	37.20	37.20	37.20	37.20	37.20	37.20	37.20
k-EYK	55.07	55.07	57.00	58.45	54.10	56.52	54.59	53.14	54.58	55.07
DVM-KÇ-GALS	59.85	68.18	61.36	68.18	62.12	69.69	59.09	69.69	62.12	63.63
DDA-KÇ-GALS	52.17	62.80	57.00	59.42	55.07	58.93	53.14	57.48	49.27	56.52

DVM ve DDA yöntemlerinin yardımı ile hesaplanan karar çizgileri ile örnekleri en iyi pozisyona döndürme yoluyla taşıyarak en iyi sınıflandırma sonucunu üreten DVM-KÇ-GALS ve DDA-KÇ-GALS yöntemlerinin denemesinde kullanılan son veri kümesi WISNCONSIN için elde edilen sonuçları Çizelge 4.9’da verilmiştir.

Bu tez araştırmasında önerilen DVM-KÇ-GALS ve DDA-KÇ-GALS yöntemleri WISCONSIN veri kümesinden elde edilen (2 ve 7) öznitelik ikilisini sırasıyla 40.76° ve 44.20° açılara döndürmektedirler ve sırasıyla k=4 iken %69.70 ve k=2 iken %78.03 oranlarında sınıflandırma sonuçlarını kaydetmiştir. Aynı durum için Bayes ve DVM yöntemleri sırasıyla %55.30 ve %44.70 oranında sonuç üretmiştir. WISCONSIN için k-EYK yöntemi k=8 iken %68.94 oranında çıktı üretirken diğer klasik sınıflandırıcılara göre oldukça yüksek sonuç kaydetmiştir. Ancak, k-EYK yönteminin sonucunu yaklaşık %9 oranında arttıran DDA-KÇ-GALS algoritması WISCONSIN veri kümesi için en yüksek sınıflandırma sonucunu üretmiştir.

Çizelge 4.9. WISCONSIN için k komşu sayısına göre sınıflandırma sonuçları

	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
Bayes	55.30	55.30	55.30	55.30	55.30	55.30	55.30	55.30	55.30	55.30
DVM	44.70	44.70	44.70	44.70	44.70	44.70	44.70	44.70	44.70	44.70
k-EYK	64.39	64.39	63.63	64.39	65.15	68.18	68.18	68.94	67.42	67.42
DVM-KÇ-GALS	55.30	68.94	59.09	69.70	63.64	66.67	63.64	68.18	65.15	65.15
DDA-KÇ-GALS	65.91	78.03	64.39	71.21	65.90	69.69	62.89	69.70	66.67	73.48

Sonuç olarak denemelerde elde edilen sayısal sonuçlardan DVM ve DDA ile karar doğruları ile en iyi döndürme açılarını hesaplayarak ve ona göre en iyi pozisyona taşımayla verileri önişlemeden geçiren ve Lorentz uzayında sınıflandıran KÇ-GALS yönteminin diğer klasik yöntemlere göre daha başarılı sonuç ürettiği görülmüştür.

İki boyutlu Lorentz sınıflandırıcısı

Bu bölümde bu tez çalışmasında Bölüm 3.3.4'te sırasıyla Algoritma-4, Algoritma-5 ve Algoritma-6 ile tanımlanan yöntemler ile elde edilen sonuçlar verilmiştir. Yani, sıkıştırma tabanlı İki Boyutlu Lorentz Sınıflandırıcısı yöntemlerinin sayısal sonuçları analiz edilmiştir. Denemelerde önerilen Sıkıştırma ile Lorentz Sınıflandırıcısı (S-LS), Optimal Parametrelili Lorentz Sınıflandırıcısı (OP-LS) ve Döndürme ile Lorentz Sınıflandırıcısı (D-LS) yöntemlerinin geçerliliğini ve doğruluğunu test etmek amacıyla serbest kullanıma açık olan CLIMATE, GESTURE, PARKINSON, RELAX, VERTEBRAL, WINE [20] veri kümeleri kullanılmıştır. Ayrıca, elde edilen sınıflandırma sonuçları Bayes, DVM ve k-EYK yöntemlerinin sonuçları ile karşılaştırılarak analiz edilmiştir.

Denemelerde iki boyutlu Lorentz uzayı incelendiği için tüm veri kümeleri ikili özniteliklere gruplanarak kullanılmıştır. Örneğin, RELAX veri kümesinden seçilen (3 ve 5) öznitelik ikilisi için S-LS yöntemi %73.50 oranında sınıflandırma başarısını üretmiştir. Aynı durum için k-EYK yöntemi %62.39 oranında başarı göstermiştir. Bu sonuçlardan görüldüğü gibi Öklid uzaklık ölçütü yerine Lorentz uzaklığının kullanımı halinde sınıflandırma başarısı yaklaşık %10 oranında artmıştır. Literatürden bilindiği gibi, iki sınıflı örnekler için başarılı sonuç üreten DVM algoritması RELAX veri kümesi için %47.86 oranında sonuç vermiştir. Bununla birlikte, Bayes yöntemi %72.65 oranında sınıflandırma başarısını göstermiştir ve bu sonuç S-LS sonucuna en yakın sonuçtur. D-LS yöntemi Bayes yöntemi ile aynı oranda sınıflandırma başarısını göstermiştir. Ancak, RELAX veri kümesi için en iyi sınıflandırma sonucu en iyi sıkıştırma matrisi temelindeki OP-LS yöntemi tarafından %82.05 oranında verilmiştir.

Buna benzer durum, yani D-LS yönteminin S-LS yöntemine göre daha düşük sınıflandırma sonucunu ürettiği durum CLIMATE ve GESTURE veri kümelerinde görülmüştür. Ancak, her iki durumunda da D-LS yönteminin çıktıları diğer klasik yöntemlere göre daha başarılıdır. CLIMATE ve GESTURE kümeleri için OP-LS yöntemi sırasıyla %93.99 ve

%91.38 oranından en yüksek sınıflandırma sonuçlarını vermiştir. Bununla birlikte, OP-LS yöntemi GESTURE için ortalama %20 oranında artış göstermektedir. GESTURE kümesi için S-LS yönteminin çıktısına en yakın sınıflandırma başarısı k-EYK yöntemi tarafından %69.83 oranında verilmiştir. Buna benzer durum CLIMATE için de geçerlidir ve S-LS yönteminin sonucuna en yakın çıktı k-EYK yöntemi ile %90.14 oranında elde edilmiştir.

Denemelerde WINE kümesinden seçilen (2 ve 3) öznitelik ikilisi için önerilen S-LS, OP-LS ve D-LS yöntemlerinin üçü de %96.20 oranında aynı sınıflandırma başarısını vermişlerdir. Aynı öznitelik ikilisinde bu sonuca en yakın sınıflandırma başarısı k-EYK yöntemi tarafından %93.67 oranında elde edilmiştir. WINE için Bayes ve DVM yöntemleri sırasıyla %84.81 ve %88.61 oranında sınıflandırma başarısını üretmiştir.

PARKINSON için S-LS ve D-LS yöntemleri sırasıyla %73.30 ve %73.33 oranında sınıflandırma sonuçlarını vermiştir. DVM yönteminin çok yüksek sınıflandırma sonucu üretmesine rağmen S-LS ve D-LS yöntemleri %2 oranında daha başarılı sonuç göstermiştir. Ayrıca, OP-LS yöntemi PARKINSON için %80 oranında en iyi çıktı üreterek kendisine en yakın sınıflandırma sonucunu önemli ölçüde arttırmıştır.

VERTEBRAL için en iyi sınıflandırma sonucu D-LS tarafından %76.50 oranında verilmiştir. Aynı durum için S-LS ve OP-LS yöntemleri sırasıyla %70.00 ve %72.00 oranında sınıflandırma başarısı üretmiştir. VERTEBRAL için S-LS yönteminin çıktısına en yakın sonuç k-EYK yöntemi tarafından %60.00 olarak verilmiştir ve en kötü sonuç DVM tarafından %38.50 oranında elde edilmiştir. OP-LS yönteminin D-LS yöntemine göre düşük sonuç ürettiği tek durum VERTEBRAL veri kümesi için geçerlidir ve diğer veri kümelerinin tamamında OP-LS yönteminin üstünlüğü sabittir.

Genel olarak, denemelerde elde edilen sayısal sonuçlara göre önerilen S-LS ve D-LS yöntemleri diğer klasik Bayes, DVM ve k-EYK sınıflandırıcılarına göre daha başarılı olduğu görülmüştür. Ancak, iki boyutlu Lorentz uzayında OP-LS yönteminin sınıflandırma açısından en iyi yöntem olduğu tespit edilmiştir. Bununla birlikte, sonuçlardan görüldüğü gibi S-LS yöntemi çok iyi dağılmış olan veri kümelerinde diğer klasik sınıflandırıcıların sonuçlarına göre hissedilir ölçüde fark yaratan başarı verememektedir. Ancak, tam tersi durumda, yani klasik sınıflandırıcıların düşük sonuç ürettiği durumlarda S-LS yöntemi çok başarılı sınıflandırma başarısı göstermektedir. Bu durum WINE ve CLIMATE veri kümeleri

durumlarında net bir şekilde görülmektedir. Ayrıca, S-LS yöntemi Bayes ve DVM yöntemlerinin en düşük sınıflandırma sonuçlarını ürettiği VERTEBRAL VE GESTURE durumlarında %20 oranında başarıyı arttırmıştır. Bununla birlikte, aynı durum için sınıflandırma başarıları arasındaki fark her iki D-LS ve OP-LS yöntemi için %30 oranında elde edilmiştir.

İki boyutlu Lorentz uzayında sıkıştırma tabanlı sınıflandırma yapan S-LS, D-LS ve OP-LS yöntemleri ile elde edilen sınıflandırma sonuçlarının sayısal karşılaştırması Çizelge 4.10'da verilmiştir.

Çizelge 4.10. S-LS, D-LS ve OP-LS yöntemlerinin sınıflandırma sonuçlarının kıyaslanması (Öznitelik ikilileri parantez içinde verilmiştir)

	CLIMATE (2/3)	GESTURE (3/9)	PARKINSON (2/6)	RELAX (3/5)	VERTEBRAL (2/4)	WINE (2/3)
Bayes	88.70	64.66	66.67	72.65	54.00	84.81
DVM	80.77	57.33	71.67	47.86	38.50	88.61
k-EYK	90.14	69.83	60.00	62.39	60.00	93.67
S-LS	93.26	84.48	73.30	73.50	70.00	96.20
D-LS	93.03	78.45	73.33	72.65	76.50	96.20
OP-LS	93.99	91.38	80.00	82.05	72.00	96.20

Genel olarak, iki boyutlu Lorentz uzayında sınıflandırma yapmak amacıyla geliştirilen sıkıştırma tabanlı Lorentz Sınıflandırıcısı ile üretilen sınıflandırma sonuçları tüm öznitelik ikililerinde Bayes, DVM ve k-EYK yöntemlerine göre daha iyi değildir. Ancak, çoğunluğunda daha başarılı sonuç üretmektedir. Bu durum veri kümesinin yapısından ve örneklerin dağılımından kaynaklandığı düşünülmektedir.

Algoritma-5 ile tanımlanan OP-LS yönteminin sıkıştırma matrisinin denemelerde kullanılan w ve q en iyi parametrelerinin değerleri Çizelge 4.11'de verilmiştir. Bu parametreler veri kümelerinden elde edilen tüm öznitelik ikilileri için geçerlidir.

Çizelge 4.11. OP-LS denemelerinde kullanılan w ve q değerleri

Veri kümesi	w	q
CLIMATE	1.7	0.01
GESTURE	2	0.36
PARKINSON	1.8	0.06
RELAX	2	0.09
VERTEBRAL	1.9	0.29
WINE	1.1	0.02

Denemelerde CLIMATE veri kümesi için en iyi parametreleri hesaplamada S-LS yönteminin en kötü sınıflandırma sonucunu üreten öznitelik ikilisi kullanılmıştır. CLIMATE veri kümesinden elde edilen (2 ve 5) öznitelik ikilisi için S-LS yöntemi %87.50 oranında başarı sergilemektedir ve bu en düşük orandır. Daha sonra 3.3.4. bölümünde belirtilen aralıkta en yüksek sınıflandırma sonuçlarını üreten w ve q parametrelerinin değerleri en iyi parametre olarak alınmıştır. CLIMATE kümesi için en iyi parametre değerleri $w=1.7$ ve $q=0.01$ olarak elde edilmiştir. Bu parametreler ile CLIMATE veri kümesinin tüm ikili öznitelik gruplarından elde edilen sınıflandırma sonuçları değerlendirilmiştir. CLIMATE veri kümesinden elde edilen sekiz öznitelik grubu için Bayes, DVM, k-EYK, S-LS, OP-LS ve D-LS yöntemleri ile denemelerde elde edilen sayısal sonuçların karşılaştırılması Çizelge 4.12’de verilmiştir. Sonuçlara göre CLIMATE kümesi için en düşük sonuç üreten öznitelik ikilisi için %87.50 oranında verilen S-LS sonucu OP-LS tarafından %4 oranında arttırılarak %91.35 olarak verilmiştir. Ayrıca, en iyi parametre kullanımı ile CLIMATE veri kümesinden elde edilen tüm öznitelik ikililerinde sınıflandırma sonuçlarının önemli ölçüde arttığı görülmüştür.

Denemelerde kullanılan PARKINSON veri kümesi için sayısal sonuçların karşılaştırılması Çizelge 4.13’te verilmiştir. Çizelge 4.13’te verilen sınıflandırma sonuçlarına göre S-LS yöntemi PARKINSON için bazı öznitelik grupları için k-EYK ve diğer yöntemlere göre daha düşük çıktı üretmektedir. Bu durum tamamen bu öznitelik ikililerinin dağılımından kaynaklanmaktadır. Örneğin, (2, 9) ve (2, 12) ikili öznitelik grupları için S-LS yöntemi k-EYK yönteminin çıktılarını %5 oranında arttırarak daha başarılı sınıflandırma sonuçlarını üretmektedir. Ancak, aynı durumlar için DVM tarafından verilen sonuçlara göre az farkla olsa da düşüktür. Bu bağlamda, D-LS yönteminde belirtildiği gibi S-LS yönteminin önışlem

adımı 45° açığa döndürme işlemi ile genişletildiğinde, PARKINSON için D-LS yöntemi tarafından üretilen sınıflandırma sonuçları hissedilir seviyede artmaktadır. Örneğin, PARKINSON veri kümesinden elde edilen (2, 9) ve (2, 12) ikili öznelik gruplarında S-LS yöntemi %65 oranında sınıflandırma başarısı gösterirken, döndürme sonrasında D-LS yönteminin çıktıları aynı durumlar için sırasıyla %73.33 ve %71.66 olarak verilmiştir. PARKINSON veri kümesinin diğer durumlarında, örneğin (2, 10) ikili öznelik grubunda S-LS yönteminin sınıflandırma sonucuna göre D-LS yöntemi ile %8 oranında daha yüksek sonuç üretmiştir.

Genel olarak, PARKINSON veri kümesi için D-LS yöntemi k-EYK ve diğer sınıflandırma yöntemlerine göre daha başarılı sınıflandırma sonuçları vermektedir. PARKINSON veri kümesi için elde edilen sınıflandırma sonuçlarının yer aldığı Çizelge 4.13'ten görüldüğü gibi, önceden belirlenen açıyla örneklerin eğitim öncesi döndürülmesi D-LS yönteminin çıktılarını olumlu etkilediği görülmüştür.

Denemelerde kullanılan RELAX veri kümesi için sıkıştırma matrisinin en iyi parametrelerin kullanımı ile OP-LS yöntemi tarafından elde edilen sınıflandırma sonuçları ikili öznelik gruplarının çoğunluğunda S-LS, D-LS ve diğer klasik sınıflandırıcılara göre daha başarılı olduğu görülmüştür. Bu durumun RELAX veri kümesinin yapısından kaynaklandığı görülmüştür.

İki boyutlu Lorentz uzayındaki S-LS, D-LS ve OP-LS yöntemlerinin CLIMATE, PARKINSON ve RELAX veri kümelerinin ikili grupları için ürettiği sınıflandırma sonuçlarının diğer klasik sınıflandırıcıların sonuçları ile karşılaştırması sırasıyla Çizelge 4.12, Çizelge 4.13 ve Çizelge 4.14 ile verilmiştir.

Çizelge 4.12. CLIMATE veri kümesinden elde edilen 8 ikili öznitelik grubu için sınıflandırma sonuçları

	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)	(1, 7)	(1, 8)	(1, 9)
k-EYK	90.14	88.94	92.06	89.18	90.86	91.10	89.66	91.10
Bayes	88.70	84.85	85.33	84.61	86.29	85.57	83.89	85.09
DVM	80.76	69.71	69.95	69.71	68.26	69.95	67.54	69.71
S-LS	93.26	91.58	90.62	90.14	92.30	91.34	92.06	93.02
D-LS	92.06	89.18	90.38	89.42	89.18	90.62	88.46	88.94
OP-LS	93.99	92.78	93.50	92.54	93.75	93.50	92.30	93.75

Çizelge 4.13. PARKINSON veri kümesinden elde edilen 8 ikili öznitelik grubu için sınıflandırma sonuçları

	(2, 6)	(2, 7)	(2, 8)	(2, 9)	(2, 10)	(2, 11)	(2, 12)	(2, 13)
k-EYK	58.33	50.00	60.00	60.00	51.66	61.66	60.00	45.00
Bayes	58.33	65.00	66.66	68.33	61.66	48.33	68.33	41.66
DVM	56.66	73.33	71.66	66.66	70.00	58.33	66.66	28.33
S-LS	66.66	70.00	73.33	65.00	68.33	66.66	65.00	55.00
D-LS	61.66	75.00	73.33	73.33	75.00	66.66	71.66	33.33
OP-LS	66.66	70.00	80.00	70.00	66.66	63.33	70.00	53.33

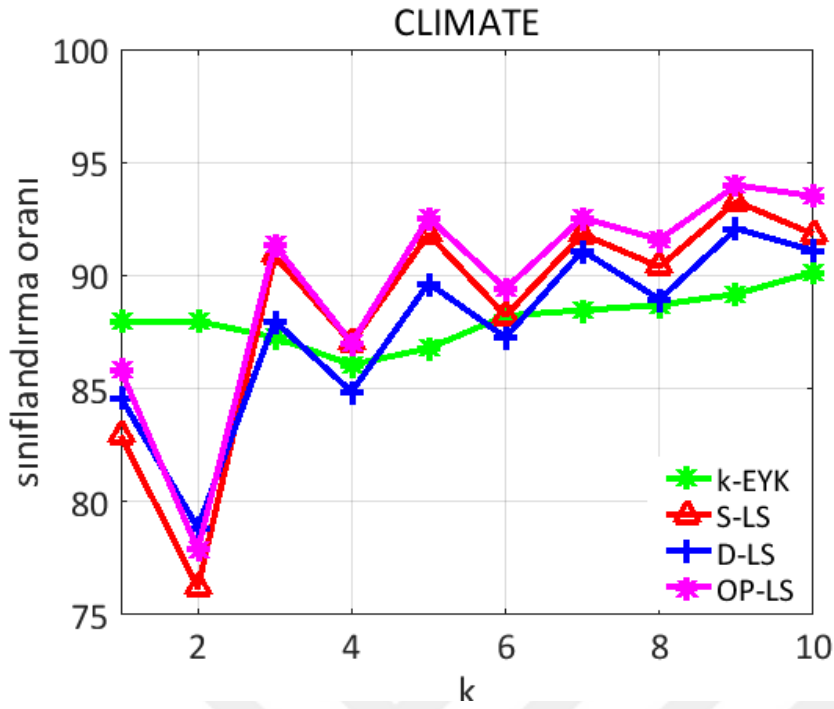
Çizelge 4.14. RELAX veri kümesinden elde edilen 8 ikili öznitelik grubu için sınıflandırma sonuçları

	(2, 3)	(2, 4)	(2, 5)	(2, 6)	(2, 7)	(2, 8)	(2, 9)	(2, 10)
k-EYK	64.10	62.39	64.95	72.64	63.24	72.64	69.23	69.23
Bayes	71.79	72.65	75.21	76.06	58.97	74.35	72.64	72.64
DVM	48.71	47.86	40.17	50.42	43.58	47.86	40.17	42.73
S-LS	77.77	73.50	78.63	76.06	75.21	76.06	75.21	76.06
D-LS	76.92	76.92	76.92	76.92	76.92	76.06	73.50	76.92
OP-LS	78.63	82.05	76.92	76.06	76.92	76.06	79.48	76.06

OP-LS yönteminde sadece en kötü sonucu üreten ikili öznitelik grubuna göre en iyi parametreleri hesaplayıp, onları veri kümesinin tamamına kullanmak OP-LS yönteminin işlem karmaşıklığını önemli ölçüde azaltmaktadır. Ayrıca, veri kümesinden elde edilen her ikili öznitelik grubu için ayrı ayrı en iyi parametreler de hesaplanabilir ve sınıflandırma sonuçları da ona göre arttırılabilir. Ancak, bu yöntemin çalışma zamanını da aynı oranda arttırmaktadır. Genel olarak denemelerde elde edilen sınıflandırma sonuçlarına bakıldığında OP-LS yöntemi k-EYK, S-LS ve D-LS yöntemlerine göre ikili özniteliklerin çoğunluğunda daha başarılı sınıflandırma sonucu üretmektedir.

S-LS, D-LS ve OP-LS yöntemleri iki boyutlu Lorentz uzayında en yakın komşulara göre sınıflandırma yapmaktadır ve benzer teknik Öklid uzayında k-EYK yönteminde kullanılmaktadır. Bu bağlamda, Lorentz metriği temelindeki S-LS yönteminin k komşu sayısının değişimdeki sınıflandırma sonuçları incelenmiştir. Uygulamada elde edilen sonuçlara göre önerilen S-LS yöntemi k değerinin artımında klasik k-EYK yöntemi kadar yumuşak geçiş yapmamaktadır. Yani, S-LS yönteminin sınıflandırma başarısının grafiği bazen sert düşüş ve yüksek çıkış yapmaktadır. Örneğin, CLIMATE veri kümesi için k değerinin küçük değerlerinde S-LS, D-LS ve OP-LS yönteminin sınıflandırma başarısı %77 ve %87 oranları arasında değişmektedir. Ancak, k=5 ve daha üst değerlerinde S-LS, D-LS ve OP-LS yöntemlerinin sınıflandırma başarıları %92 ve %94 arasında değişmektedir. Aynı durumlar için k-EYK yöntemi daha düşük oranda değişim göstermektedir. Sonuç olarak k değerinin küçük değerlerinde önerilen S-LS, D-LS ve OP-LS yöntemler düşük sınıflandırma sonucu üretmektedirler ve k değerinin artımında sınıflandırma sonuçları hızla artmaktadırlar. Bu duruma rağmen S-LS, D-LS ve OP-LS yöntemleri k-EYK ve diğer klasik yöntemlere göre daha başarılı sonuç üretmektedirler. Bu durumun CLIMATE veri kümesi için görselleştirilmesi Şekil 4.7’de verilmiştir.

S-LS, D-LS ve OP-LS yöntemleri ile CLIMATE, GESTURE ve VERTEBRAL veri kümelerinden elde edilen sırasıyla (1, 8), (1, 7) ve (2, 8) ikili öznitelik grupları için k değerinin değişimdeki sınıflandırma sonuçları Çizelge 4.15, Çizelge 4.16 ve Çizelge 4.17’de verilmiştir.



Şekil 4.7. CLIMATE veri kümesi için k değerinin değişimindeki sınıflandırma sonucu

Çizelge 4.15. CLIMATE veri kümesi için k-EYK, S-LS, D-LS ve OP-LS yöntemlerinin k değerinin değişimindeki sınıflandırma sonuçları

	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
k-EYK	87.98	87.98	87.25	86.05	86.77	88.22	88.46	88.70	89.18	90.14
Bayes	88.70	88.70	88.70	88.70	88.70	88.70	88.70	88.70	88.70	88.70
DVM	80.77	80.77	80.77	80.77	80.77	80.77	80.77	80.77	80.77	80.77
S-LS	82.93	76.20	90.86	87.01	91.82	88.22	91.82	90.38	93.27	91.83
D-LS	84.61	78.84	87.98	84.85	89.66	87.25	91.10	88.94	92.07	91.11
OP-LS	85.81	77.88	91.34	87.01	92.54	89.42	92.54	91.58	93.99	93.50

Çizelge 4.16. GESTURE veri kümesi için k-EYK, S-LS, D-LS ve OP-LS yöntemlerinin k değerinin değişimindeki sınıflandırma sonuçları

	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
k-EYK	63.36	63.36	69.83	58.62	58.62	58.62	69.39	68.10	69.82	68.96
Bayes	64.66	64.66	64.66	64.66	64.66	64.66	64.66	64.66	64.66	64.66
DVM	57.33	57.33	57.33	57.33	57.33	57.33	57.33	57.33	57.33	57.33
S-LS	71.12	79.74	84.48	78.01	53.44	60.34	60.77	66.81	66.81	77.58
D-LS	78.44	75.86	78.45	75.86	72.84	73.27	73.27	71.55	71.55	69.82
OP-LS	89.65	90.94	91.38	88.36	75.86	86.20	86.20	86.20	86.20	86.63

Çizelge 4.17. VERTEBRAL veri kümesi için k-EYK, S-LS, D-LS ve OP-LS yöntemlerinin k değerinin değişimindeki sınıflandırma sonuçları

	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
k-EYK	52.50	52.50	50.50	51.00	53.50	52.50	54.00	55.00	54.50	53.50
Bayes	48.50	48.50	48.50	48.50	48.50	48.50	48.50	48.50	48.50	48.50
DVM	47.00	47.00	47.00	47.00	47.00	47.00	47.00	47.00	47.00	47.00
S-LS	50.50	52.50	50.00	53.50	51.00	54.50	54.00	55.50	55.00	59.50
D-LS	66.50	68.50	70.00	72.00	72.00	72.50	73.00	74.50	77.50	74.00
OP-LS	64.50	70.50	69.50	71.50	71.00	72.00	71.50	68.50	67.00	70.50

4.2. Çok Boyutlu Lorentz Sınıflandırıcısı Sonuçları

Çok Boyutlu Lorentz Sınıflandırıcıları bu tez araştırmasının 3.4.3 bölümünde tanımlanmıştır ve sırasıyla Algoritma-7, Algoritma-8 ve Algoritma-9 ile verilmiştir. Bu bölümde bu algoritmaların ikiden çok özniteliğe sahip veri kümeleri üzerinde denemeleri yapılmıştır ve deneysel sonuçları analiz edilmiştir. Çok boyutlu verilerin Lorentz uzayında sınıflandırılması için önerilen ÇBLS, OP-ÇBLS ve D-ÇBLS yöntemlerinin sınıflandırma performansını test etmek amacıyla genel kullanıma açık olan GESTURE, LVST, MADELON, RELAX, SONAR, SEEDS, TELESCOPE, WINE ve WISCONSIN veri kümeleri kullanılmıştır.

Denemelerde kullanılan veri kümeleri Çizelge 2.1.'den de görüldüğü gibi farklı sayıda öznitelik ve örnek sayısından oluşmaktadırlar. Bu bağlamda, Eş. 3.17.'e göre bu veri kümelerinden elde edilen ikili öznitelik gruplarının sayısı da değişmektedir. Örneğin, GESTURE veri kümesi toplamda 18 öznitelik içermektedir ve Eş. 3.17.'e göre toplamda 153 ikili öznitelik kombinasyonu elde edilmiştir. Bu durum sırasıyla 12, 10 ve 13 öznitelik içeren RELAX, TELESCOPE ve WINE veri kümeleri için 66, 45 ve 78 olarak elde edilmiştir. Kullanılan veri kümelerinden WISCONSIN veri kümesi için toplamda 528 ikili öznitelik kombinasyonu elde edilmiştir. SEEDS veri kümesi en az öznitelik sayısına sahip olanıdır ve toplamda 21 ikili öznitelik grubu elde edilmiştir. LSVT ve SONAR veri kümeleri için sahip oldukları 310 ve 60 özniteliklerden sırasıyla 47895 ve 1770 ikili öznitelik grubu üretilmiştir. Kullanılan veri kümelerinden en çok özniteliğe sahip olan MADELON veri kümesinden 500 öznitelikten 124750 ikili grup elde edilmiştir.

Veri kümelerinden elde edilen kombinasyonların sayısına bakıldığında öznitelik sayısının artımına orantılı olarak ikili kombinasyonların da sayısı kat kat artmaktadır. Dolayısıyla, veri kümesini en iyi şekilde temsil edecek olan ve istenilen sınıflandırma başarısını üretecek en iyi ikili kombinasyon sayısı önemlidir. Çünkü bu değer ÇBLS algoritmasının çalışma zamanını ve üreteceği sınıflandırma başarısını doğrudan etkilemektedir. Bu doğrultuda, denemelerde veri kümelerinden elde edilen ikili öznitelik kombinasyonlarından LJ ayırt etme ölçütü temelinde çalışan LÖS öznitelik seçme algoritması ile en iyi ikili öznitelik gruplarını seçerken g grup sayısı ÇBLS, OP-ÇBLS ve D-ÇBLS algoritmaları için 20 olarak belirlenmiştir. Seçilen 20 öznitelik ikili gruplar halinde olmasından dolayı büyük boyuttaki veri kümeleri için toplamda 40 farklı öznitelik anlamına gelmektedir.

Veri kümelerindeki öznitelik sayısının farklı olmasından dolayı üretilen ikili kombinasyonların da sayısı farklıdır. Ancak, GESTURE, LVST, MADELON, RELAX, SONAR, SEEDS, TELESCOPE, WINE ve WISCONSIN veri kümeleri için seçilen 20 ikili öznitelik kombinasyonları tüm veri kümelerinde, geliştirilen yeni ÇBLS algoritması ile en başarılı sınıflandırma sonucunu üretmeye yeterli olmuştur. Bununla birlikte, denemelerde kullanılan veri kümelerinin çoğunluğunda en başarılı sınıflandırma sonucunu üreten en iyi $g_{opt} = (1, g)$ değerinin belirlenen g değerinden az olabileceğini göstermiştir. Bu durum yeni ÇBLS algoritmasının çalışma zamanı ve etkililiği açısından verimliliğini arttırmaktadır. Denemelerde kullanılan veri kümelerindeki öznitelik sayısı ve onlardan elde edilen ikili kombinasyonlar ile ilgili sayısal bilgiler Çizelge 4.18'de verilmiştir. Ayrıca, tabloda veri

kümeleri için seçilen g ile yeni ÇBLS yönteminde en iyi sınıflandırma sonucunu üreten en iyi g_{opt} değerleri de kıyaslanmıştır.

Çizelge 4.18. Veri kümeleri istatistiği (g : veri kümesinden seçilen ikili öznitelik grup sayısı)

	öznitelik sayısı	ikili öznitelik grubu	g	g_{opt}
GESTURE	18	153	20	20
LVST	310	47895	20	2
MADELON	500	124750	20	6
RELAX	12	66	20	14
SEEDS	7	21	20	19
SONAR	60	1770	20	15
TELESCOPE	10	45	20	8
WINE	13	78	20	10
WISCONSIN	33	528	20	4

Bu çalışmada ÇBLS, OP-ÇBLS ve D-ÇBLS yöntemleri ile elde edilen deneysel sonuçlar klasik k-EYK, Bayes ve DVM sınıflandırıcılarının sonuçları ile karşılaştırılmıştır. Yapılan karşılaştırmanın anlamlı olmasını ve sınıflandırıcılara uygulanan adımların benzerliğini sağlamak adına ÇBLS yöntemine has olan sıkıştırma adımı dışındaki diğer adımlar klasik sınıflandırıcılar için de geçerlidir. Yani, çok boyutlu veri kümeleri klasik sınıflandırıcılar için ÇBLS yönteminde olduğu gibi, Eş. 3.17 ile ikili gruplara ayrılarak denemelerde kullanılmıştır. Klasik sınıflandırıcılar durumunda Eş. 3.17 ile elde edilen ikili gruplardan 20 en iyi ikili grupları seçmek için Lorentz Öznitelik Seçme yönteminin Öklid analoğu kullanılmıştır. Bununla birlikte, klasik k-EYK, Bayes ve DVM yöntemlerinin veri kümelerinin yalın halinin kullanımı durumundaki, yani önışlem uygulamadan tüm öznitelikleri kullanarak elde edilen sonuçları ÇBLS yönteminin çıktıları ile karşılaştırılmıştır. Bu işlem bu araştırmada önerilen önışlem adımlarının geçerliliğini ve doğrulunu test etmek amacıyla yapılmıştır. Denemelerde kullanılan veri kümelerinin orijinal

hali için klasik k-EYK, Bayes ve DVM yöntemlerinin çıktıları ile ÇBLS yönteminin sınıflandırma sonuçlarının sayısal karşılaştırılması Çizelge 4.19’da verilmiştir.

Çizelge 4.19. ÇBLS sonucu ile Bayes, DVM, k-EYK yöntemlerinin orijinal veri kümelerinin kullanımında elde edilen sonuçların karşılaştırması

	Bayes	DVM	k-EYK	ÇBLS
GESTURE	84.56	53.69	80.20	94.63
LVST	44.71	69.41	58.82	67.06
MADELON	57.46	55.97	53.73	61.94
RELAX	49.18	49.18	60.66	74.59
SEEDS	90.43	91.49	93.62	96.81
SONAR	46.76	52.52	50.36	62.59
TELESCOPE	56.39	53.38	61.65	68.42
WINE	94.19	95.35	96.51	96.51
WISCONSIN	50.76	61.36	64.39	79.55

Çizelge 4.19’den görüldüğü gibi çok boyutlu verilerin Lorentz uzayında sınıflandırma için geliştirilen ÇBLS yöntemi bir durum hariç, denemelerde kullanılan 9 veri kümelerinin tamamında klasik sınıflandırıcılara göre daha başarılı sınıflandırma sonucu vermektedir. Denemelerde LVST veri kümesi için Bayes ve k-EYK yöntemlerinin ürettiği sonuçlar sırasıyla %44.71 ve %58.82 oranında verilirken, aynı durum için DVM ve önerilen ÇBLS yöntemleri sırasıyla %69.41 ve %67.06 oranında sınıflandırma başarısı göstermiştir. ÇBLS yöntemi Bayes ve k-EYK yöntemlerinin sınıflandırma başarılarına göre sırasıyla yaklaşık %22 ve %9 oranında daha yüksek sonuç verirken, DVM göre daha düşük sonuç üretmiştir. Bu durumun LVST veri kümesinin yapısından kaynaklandığı düşünülmektedir.

Bu tez araştırmasının 3.3.4. bölümünde belirtildiği gibi Lorentz metriği ile sınıflandırma başarısı açısından λ sıkıştırma matrisindeki w, q parametrelerinin aldığı değerler önemlidir. Bu bağlamda OP-ÇBLS yönteminin testinde kullanılan GESTURE, LVST, MADELON,

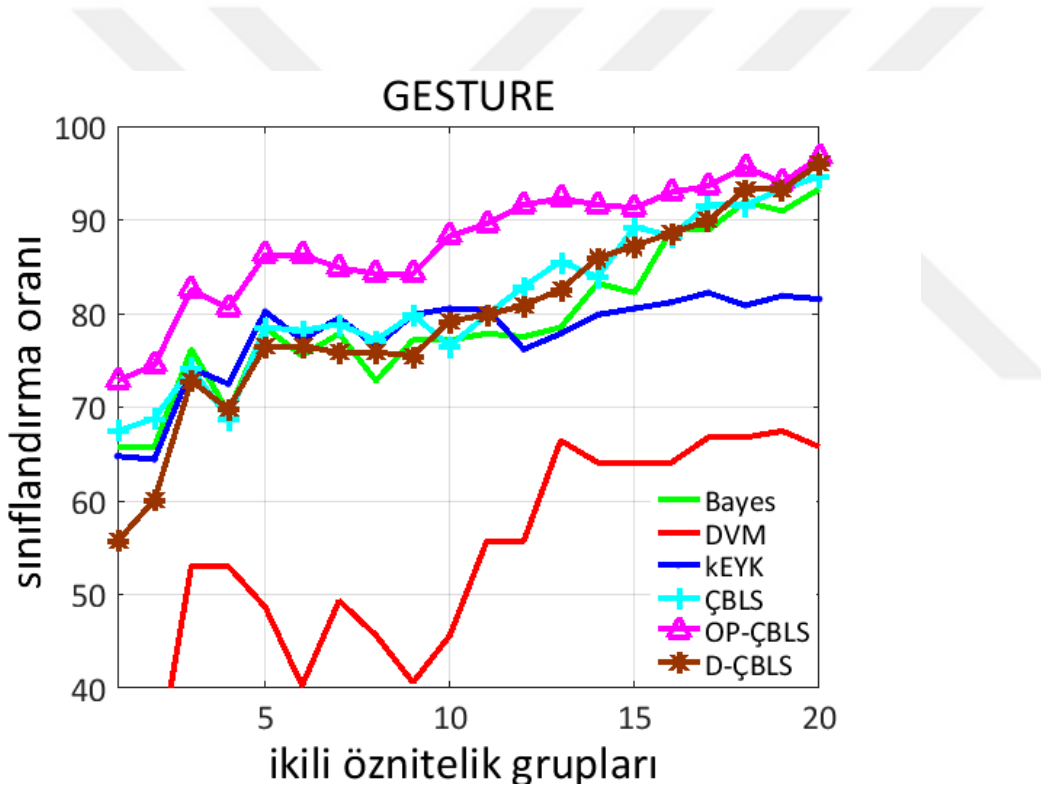
RELAX, SONAR, SEEDS, TELESCOPE, WINE ve WISCONSIN veri kümeleri için w , q parametrelerinin en iyi değerleri hesaplanmıştır. Bu verilerin Lorentz uzayında sınıflandırılmasında en iyi sınıflandırma başarısını veren sıkıştırma matrisinin en iyi değerleri Çizelge 4.20’de verilmiştir.

Çizelge 4.20. Veri kümeleri için sıkıştırma matrisi parametrelerinin en iyi değerleri

	w_{opt}	q_{opt}
GESTURE	0.9	1.8
LSVT	1	1.5
MADDELON	1	1.6
RELAX	0.1	1.9
SEEDS	2	1.4
SONAR	0.25	2
TELESCOPE	1	1.1
WINE	2	1.8
WISCONSIN	0.9	1.9

Önerilen ÇBLS yöntemi ile yapılan denemelerde kullanılan GESTURE veri kümesi için DVM yönteminin en iyi sınıflandırma başarısı %67.45 oranında verilmiştir. Aynı durum için diğer klasik sınıflandırıcıların gösterdikleri en iyi başarı oranı k-EYK için %85.23 ve Bayes için %93.29 olarak elde edilmiştir. ÇBLS ve D-ÇBLS yöntemlerinin GESTURE veri kümesi için ürettikleri en iyi sınıflandırma sonuçları ise %94.63 ve %95.97 oranında olmuştur. GESTURE veri kümesi için en iyi sınıflandırma sonucu OP-ÇBLS yöntemi tarafından %96.64 oranında verilmiştir. OP-ÇBLS yöntemi klasik sınıflandırıcılardan kendisine en yakın %93.29 oranında sınıflandırma başarısını üreten Bayes yönteminin sonucunu %3 oranında arttırmıştır. Daha önce belirtildiği gibi $g = 20$ olarak alınmıştır ve denemelerde GESTURE veri kümesi için $g_{opt} = 20$ olarak tespit edilmiştir. Bunun anlamı da, ÇBLS algoritması GESTURE veri kümesinden elde edilen 153 ikili öznitelik kombinasyonlarından LÖS öznitelik seçme yöntemi ile 20 ikili grubu seçerek en iyi sınıflandırma başarısını vermiştir. Yani, GESTURE veri kümesi için yeni ÇBLS yöntemi ile en iyi sınıflandırma sonucunu elde etmek için toplam 153 ikili öznitelik grubunun %12 kısmının yeterli olması, önerilen yöntemin etkinliğini ve geçerliliğini kanıtlamaktadır.

Ayrıca, $g = 1$ iken, yani sadece iki öznitelik ile yeni ÇBLS yöntemi %67.45 oranında başarı üreterek klasik DVM yönteminin en iyi sonucunu geçmiştir. Sonra, sırasıyla $g = 12$ iken k-EYK ve $g = 13$ iken de Bayes sınıflandırıcılarına göre hissedilir oranda daha yüksek sınıflandırma sonucunu göstererek üstünlüğünü ortaya koymuştur. D-ÇBLS yöntemi ile ÇBLS yönteminin sınıflandırma grafiği birbirine benzer ve yakın aralıklarda değişmektedir. Ancak, OP-ÇBLS yöntemi GESTURE veri kümesinden seçilen 20 ikili öznitelik gruplarının tamamında ÇBLS ve D-ÇBLS yöntemlerine göre daha başarılı sınıflandırma sonuçları vermektedir. GESTURE veri kümesi için üstte beyan edilenler ve yeni ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemlerinin g değerinin artımında verdiği sınıflandırma sonuçlarının klasik sınıflandırıcıların çıktıları ile görsel karşılaştırması Şekil 4.8’de verilmiştir.



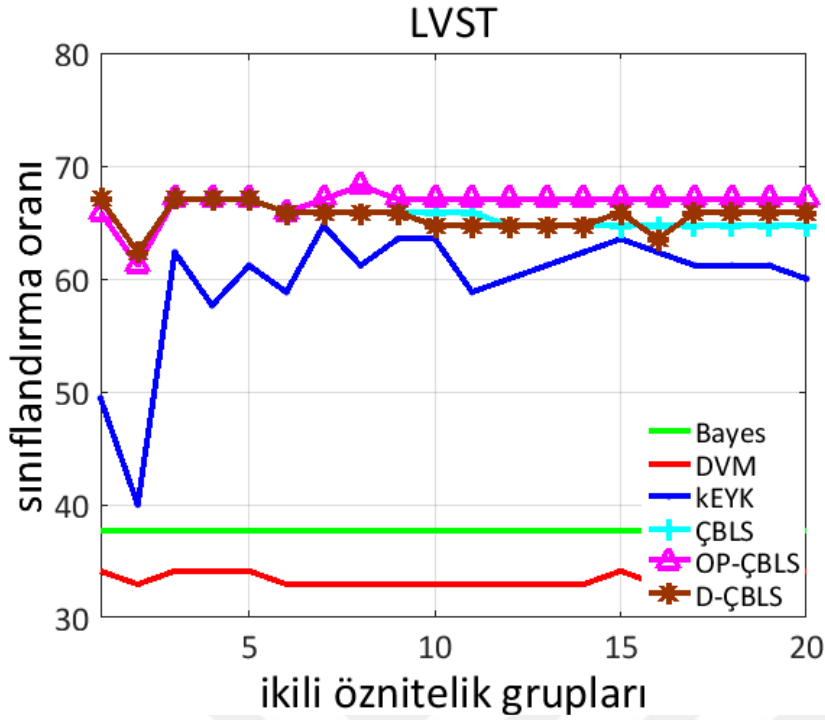
Şekil 4.8. ÇBLS yönteminin GESTURE veri kümesi için sınıflandırma sonuçları

LVST veri kümesi denemelerde kullanılan ve 310 öznitelikle en çok özniteliğe sahip ikinci büyük veri kümesidir. Çok özniteliğe sahip olması gerçek dünya problemlerine yakın olması bakımından ÇBLS yönteminin geçerliliğini arttırmaktadır. LVST veri kümesi için Bayes yönteminin en iyi sınıflandırma sonucu %37.65 olarak verilmiştir. Aynı durum için DVM algoritmasının en iyi sınıflandırma başarısı %34.12 olarak elde edilmiştir. LVST veri kümesi için Bayes ve DVM yöntemlerinin bu kadar düşük sonuç üretmeleri öznitelik sayısından ters etkilene olarak yorumlanmaktadır. LVST veri kümesi ile yapılan denemelerde k-EYK

yönteminin en iyi sınıflandırma sonucu %64.71 oranında elde edilmiştir ve bu sonuç Bayes ve DVM yöntemlerinin en iyi sınıflandırma sonuçlarına göre oldukça yüksektir. k-EYK yönteminin en iyi sınıflandırma sonucu aynı zamanda LVST veri kümesi için ÇBLS yönteminin %67.06 orandaki en iyi sınıflandırma başarısına en yakın sonuçtur. Elde edilen sayısal sınıflandırma sonuçlarından görüldüğü gibi ÇBLS yöntemi diğer klasik algoritmaların sınıflandırma sonuçlarına göre yaklaşık %3 oranında daha yüksek sonuç vererek üstünlüğünü ortaya koymaktadır. LVST kümesinden seçilen 20 ikili öznitelik grup için D-ÇBLS yönteminin en iyi sınıflandırma sonucu ÇBLS yöntem sonucu ile aynıdır. Buradan LVST kümesi için ÇBLS yöntemine döndürme işleminin eklenmesinin sınıflandırma açısından herhangi bir etki yaratmadığı görülmüştür. Halbuki, ÇBLS yöntemindeki sıkıştırma matrisinin en iyi parametreler ile donatılması az da olsa sınıflandırma sonucunu olumlu etkilemektedir. Yani, LVST kümesi için en iyi sınıflandırma başarısı OP-ÇBLS yöntemi tarafından %68.24 olarak elde edilmiştir.

Ayrıca, LVST veri kümesi için elde edilen toplam 47895 ikili öznitelik grubundan önerilen LÖS yöntemi ile sadece en iyi 20 tanesi seçilerek deneme yapılmıştır ve ÇBLS yönteminin en iyi sınıflandırma sonucu $g_{opt}=1$ iken verilmiştir. Bununla birlikte LVST kümesinden seçilen 20 ikili öznitelik gruplarının tamamında ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemleri klasik sınıflandırıcılara göre daha başarılı sonuç üretmektedir. Seçilen g sayıdaki ikili öznitelik gruplarının değişik değerlerindeki ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemlerinin sınıflandırma sonuçlarının görsel karşılaştırması Şekil 4.9'da verilmiştir.

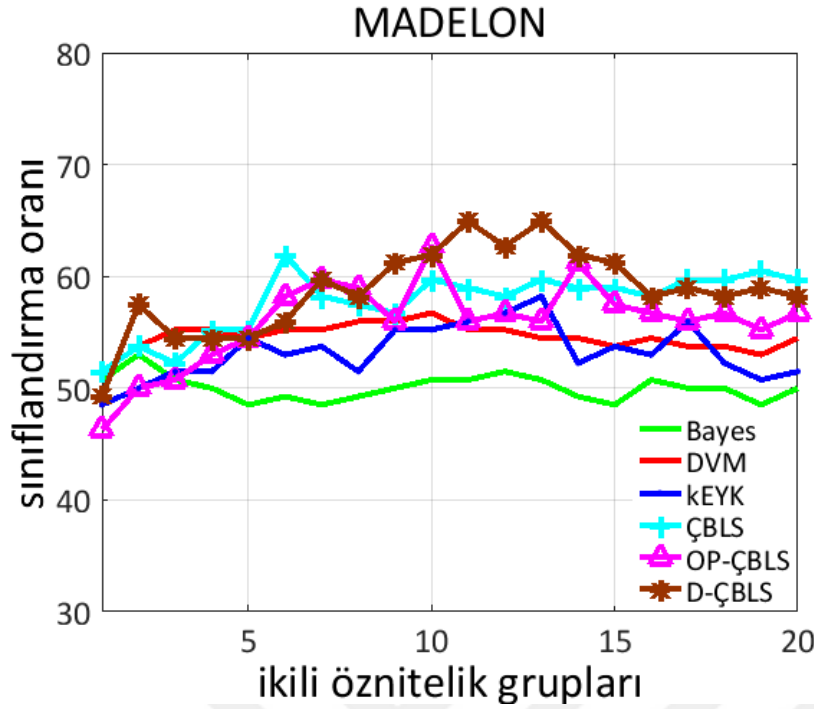
Denemelerde kullanılan 500 özniteliğe sahip MADELON veri kümesi için Eş. 3.17. ile toplamda 124750 öznitelik ikilisi elde edilmiştir. Bu veri kümesi için Bayes yönteminin en iyi sınıflandırma sonucu %52.99 oranında verilirken DVM algoritması seçilen 20 ikili öznitelik grupları için %56.72 oranında en iyi sınıflandırma sonucunu göstermiştir. MADELON veri kümesi için DVM yönteminin en iyi sonucunu yaklaşık %2 oranında arttıran k-EYK yönteminin en iyi sonucu %58.21 olarak elde edilmiştir. Bu çalışmada önerilen ÇBLS algoritması aynı durum için %61.94 oranında sınıflandırma başarısını üreterek diğer klasik yöntemlere göre daha başarılı sonuç üretmektedir. D-ÇBLS ve OP-ÇBLS yöntemleri MADELON veri kümesinden seçilen 20 ikili öznitelik grupları için sırasıyla %64.93 ve %62.69 oranında en iyi sınıflandırma sonuçlarını üretmiştir.



Şekil 4.9. ÇBLS yönteminin LVST veri kümesi için sınıflandırma sonuçları

Denemelerde elde edilen sınıflandırma sonuçlarından görüldüğü gibi MADELON veri kümesi için GESTURE ve LVST kümelerinden farklı olarak en iyi sınıflandırma sonucu döndürme tabanlı D-ÇBLS yöntemi ile verilmiştir. ÇBLS algoritması MADELON kümesi için g değerinin ilk birkaçı dışındaki diğer tüm seçilen ikili öznitelik gruplarında klasik yöntemlere göre daha başarılı sınıflandırma sonucu vermektedir. Bu duruma D-ÇBLS yöntemi için de geçerlidir. Bu durum LÖS öznitelik seçme yönteminin ve ÇBLS yönteminin sadece belirli öznitelikler için değil, MADELON veri kümesinin tüm öznitelikleri için geçerli ve kullanılabilir olduğunu göstermektedir. MADELON veri kümesi için yeni LÖS yöntemi ile seçilen 20 ikili öznitelik grupları için ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemleri ile elde edilen sınıflandırma sonuçlarının görselleştirilmesi Şekil 4.10'da verilmiştir.

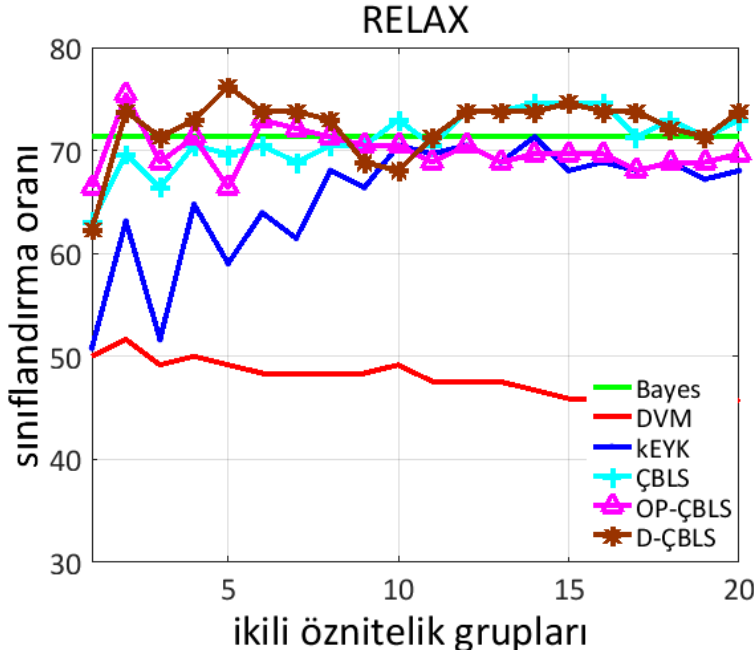
RELAX veri kümesinden 12 öznitelikten toplamda 66 ikili öznitelik kombinasyonu elde edilmiştir. Denemelerde RELAX için DVM yönteminin en iyi sınıflandırma başarısı %51.64 oranında elde edilmiştir. Klasik Bayes ve k-EYK yöntemlerinin RELAX veri kümesinden seçilen 20 öznitelik ikilisi için en iyi sınıflandırma sonuçları aynı oranda sırasıyla %71.31 ve %71.31 olarak verilmiştir. Aynı durum için önerilen yeni ÇBLS sınıflandırma algoritması $g_{opt} = 14$ iken kendisine en yakın k-EYK yönteminin sınıflandırma sonucunu %3 arttırarak %74.59 oranında sınıflandırma başarısı göstermiştir.



Şekil 4.10. ÇBLS yönteminin MADELON veri kümesi için sınıflandırma sonuçları

Bu çalışmada önerilen ve en iyi parametreler ile verileri sıkıştırma yaparak Lorentz uzayında sınıflandırma yapan OP-ÇBLS yönteminin RELAX veri kümesi için en iyi sınıflandırma sonucu %75.41 olarak elde edilmiştir. RELAX kümesinden seçilen ikili grupları döndürerek Lorentz uzayında sınıflandırma yapan D-ÇBLS algoritması bu veri kümesi için en iyi sınıflandırma sonucunu %76.23 oranında üretmiştir. RELAX veri kümesi için LÖS öznelilik seçme yöntemi ile toplam 66 ikili öznelilik grupların yaklaşık %30'unu oluşturan en iyi ikili öznelilik grubunu seçerek sınıflandırma yapan ÇBLS yöntemi kendisine en yakın olan k-EYK metodunun sonucunu %3 oranında arttırmıştır. RELAX veri kümesinde ÇBLS yöntemi $g = 1$ iken %63.11 oranında başarı üreterek klasik yöntemlerden k-EYK ve DVM sonuçlarına göre %13 oranında daha yüksek başarı göstermiştir. Ayrıca, önerilen yöntem g değeri 1 ile 14 aralığında değişirken yaklaşık %10 oranında sınıflandırma başarısını arttırmıştır. Bununla birlikte, önerilen ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemleri RELAX kümesi için en iyi sınıflandırma sonuçlarını sırasıyla $g_{opt}=14$, $g_{opt}=5$ ve $g_{opt}=2$ iken vermişlerdir. Bu durum ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemlerinin Lorentz uzayında farklı sayıdaki öznelilik sayıları ile en iyi sınıflandırma başarısını ürettiklerini göstermektedir. Ayrıca, ÇBLS yöntemi $g=10$ sonrasında diğer klasik yöntemlere göre daha başarılı sonuç üretirken, OP-ÇBLS yöntemi aynı durum için düşüş göstermektedir. Ancak, D-ÇBLS yöntemi seçilen ikili öznelilik gruplarının tamamına yakınında diğer klasik

sınıflandırıcılara göre daha başarılı sonuç üretmektedir. Bu döndürme işleminin RELAX veri kümesinin Lorentz uzayında sınıflandırılmasında olumlu etki yarattığını ve bu işlemin önışlem adımına eklenmesinin doğru adım olduğunu göstermektedir. RELAX veri kümesi için klasik sınıflandırıcı ile yeni ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemlerinin çıktılarının görsel karşılaştırması Şekil 4.11’de verilmiştir.

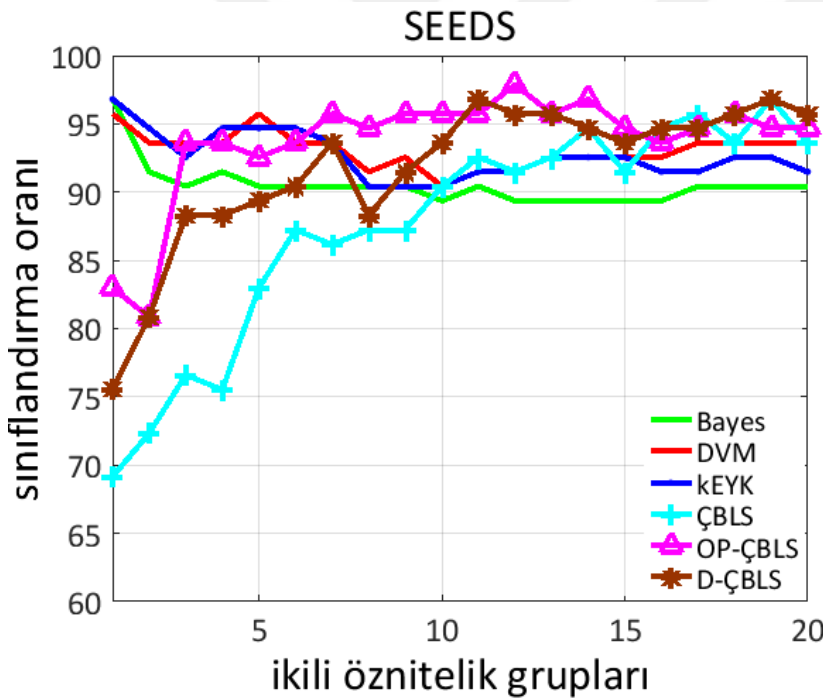


Şekil 4.11. ÇBLS yönteminin RELAX veri kümesi için sınıflandırma sonuçları

Bu tez çalışması kapsamında geliştirilen ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemlerini test etmek amacıyla yapılan denemelerde SEEDS veri kümesi de kullanılmıştır. Toplamda 21 ikili öznelik grubu Eş. 3.17 ile elde edilmiştir. Önerilen LÖS yöntemi ile SEEDS kümesinden toplamda 20 ikili grup seçilmiştir ve bu veri kümesi için en iyi sınıflandırma başarısı %97.87 oranında yeni OP-ÇBLS yöntemi ile elde edilmiştir. Aynı, durum için k-EYK algoritmasının ürettiği en iyi sonuç %95.74 olarak verilmiştir. Klasik Bayes ve DVM algoritmalarının SEEDS veri kümesinde ürettiği en iyi sonuçlar aynıdır ve %96.81 oranındadır. Bu duruma benzer sonuç, yani aynı orandaki sınıflandırma başarısı ÇBLS ve D-ÇBLS yöntemleri tarafından da SEEDS için en iyi sınıflandırma sonucu olarak verilmiştir.

Denemede elde edilen sayısal sonuçlardan görüldüğü gibi OP-ÇBLS yöntemi SEEDS veri kümesi için en iyi sınıflandırma sonucunu $g_{opt}=12$ iken vermektedir. ÇBLS ve D-ÇBLS yöntemleri SEEDS için en iyi sınıflandırma sonuçlarını sırasıyla $g_{opt}=19$ ve $g_{opt}=11$ iken

vermişlerdir. Bununla birlikte ÇBLS ve D-ÇBLS yöntemleri seçilen 20 ikili öznitelik gruplarının ilk 10 tanesinde diğer klasik yöntemlere göre düşük sonuçlar üretmiştir. Ancak, aynı durum için OP-ÇBLS yöntemi seçilen öznitelik gruplarının tamamına yakınında diğer sınıflandırıcılara göre daha başarılı sonuç üretmektedir. Bu durum ÇBLS ve D-ÇBLS yöntemleri için LVST ve MADELON veri kümelerinde yapılan denemelerde gözlemlenmiştir ve bu aynı zamanda önerilen yöntemlerin belirli bir öznitelik grubu için değil, tüm öznitelikler için geçerli olduğunu ortaya koymaktadır. SEEDS veri kümesi için klasik sınıflandırıcıların ürettiği en iyi sınıflandırma sonuçlarının yüksek oranda olmasına rağmen önerilen ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemler daha iyi sonuç üretebilmektedir. SEEDS kümesinden LÖS yöntemi ile seçilen 20 öznitelik grubu için ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemleri ile elde edilen sayısal sonuçların diğer klasik yöntemlerle karşılaştırması görsel olarak Şekil 4.12’de verilmiştir.

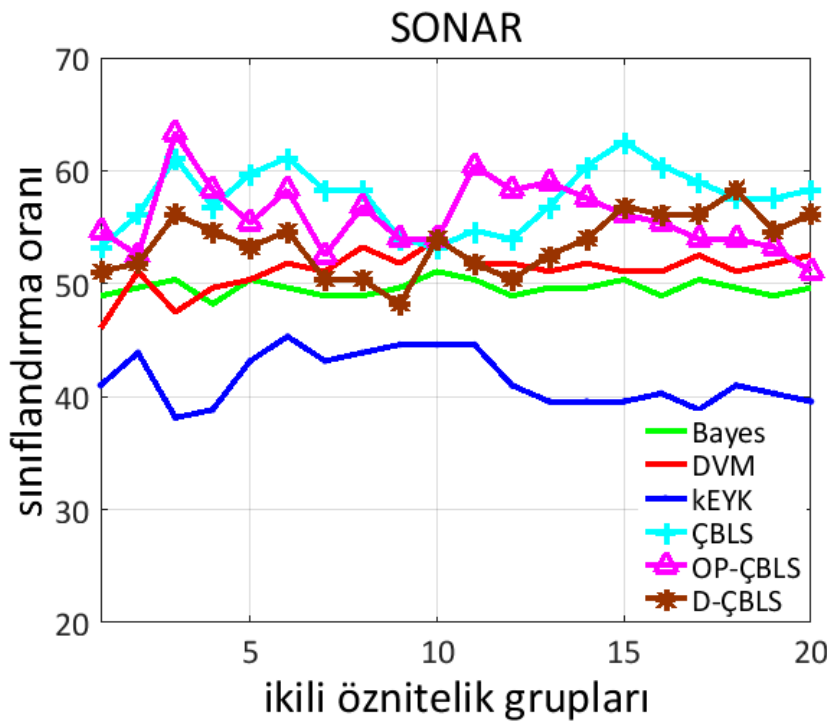


Şekil 4.12. ÇBLS yönteminin SEEDS veri kümesi için sınıflandırma sonuçları

Bu araştırmada geliştirilen ÇBLS, D-ÇBLS ve OP-ÇBLS algoritmalarının geçerliliğini ve doğruluğunu test etmek amacıyla kullanılan diğer bir veri kümesi SONAR veri kümesidir. Bu veri kümesi için Bayes sınıflandırıcısının en iyi sınıflandırma sonucu %51.08 oranında verilmiştir. Önceki veri kümelerindeki durumlardan farklı olarak DVM yöntemi SONAR veri kümesi için diğer klasik sınıflandırıcılara göre daha başarılı sonuç üretmiştir ve en iyi

sınıflandırma sonucunu %53.96 olarak vermiştir. Aynı durum için k-EYK algoritmasının en iyi sonucu %45.32 oranında elde edilmiştir. ÇBLS algoritmasının SONAR veri kümesi için verdiği en iyi sınıflandırma sonucu %62.59 oranında $g_{opt} = 15$ iken ile elde edilmiştir. ÇBLS yöntemine döndürme işleminin eklenmesi ile elde edilen D-ÇBLS yönteminin en iyi sınıflandırma başarısı %58.27 oranında olmuştur. Sonuçlardan SONAR kümesi için döndürme işleminin ters etki yarattığı görülmüştür. ÇBLS yönteminde en iyi parametrelili sıkıştırma matrisinin kullanımı ile elde edilen OP-ÇBLS yönteminin en iyi sınıflandırma sonucu ve aynı zamanda SEEDS için en iyi sınıflandırma başarısı %63.31 olarak verilmiştir.

ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemleri yeni gelen test örneğinin sınıf etiketini belirlemede en yakın komşu tekniğini ve çoğunluk mantığını kullanarak belirlemektedir. Benzer teknik Öklid uzayında k-EYK algoritması için de geçerlidir. Ancak, SONAR veri kümesi durumunda k-EYK algoritması seçilen tüm ikili öznelik gruplarında en düşük sonuçlar üretmektedir. ÇBLS ve OP-ÇBLS algoritmaları seçilen 20 ikili öznelik gruplarının tamamında, D-ÇBLS yöntemi ise çoğunda diğer yöntemlere göre daha başarılı sınıflandırma başarısı üretmektedir. Bu Lorentz uzaklık ölçütünün önemli bir ayırt edici ölçüt olduğunu ortaya koymaktadır. SONAR veri kümesi için denemelerde elde edilen sayısal sonuçların görselleştirilmesi Şekil 4.13'te verilmiştir.

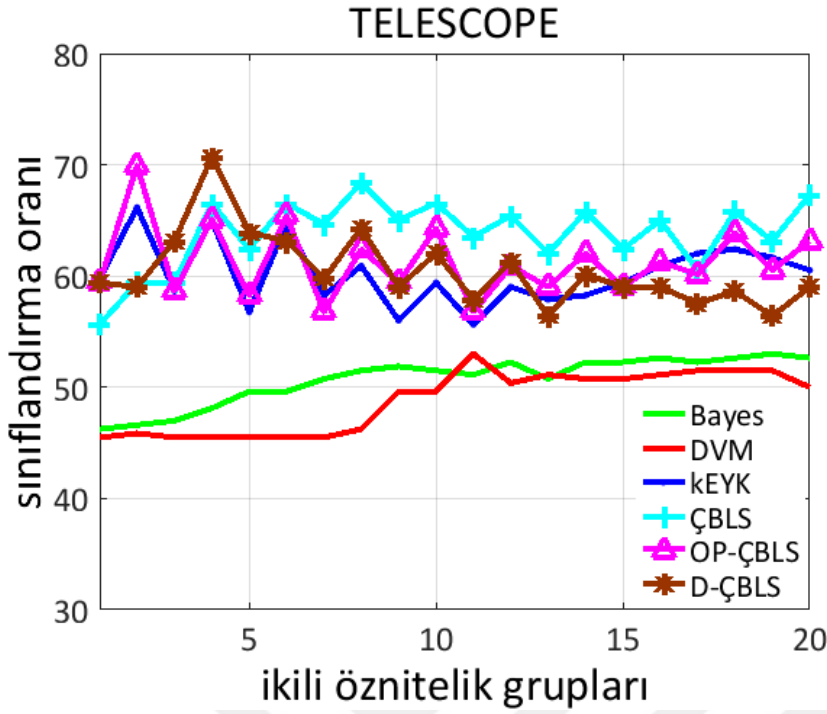


Şekil 4.13. ÇBLS yönteminin SONAR veri kümesi için sınıflandırma sonuçları

Bu araştırma denemelerinde kullanılan TELESCOPE veri kümesi için sahip olduğu 10 öznitelikten Eş. 3.17 ile toplamda 45 ikili öznitelik kombinasyonu oluşturulmuştur. Bayes ve DVM algoritmalarının TELESCOPE veri kümesinden seçilen toplam 20 ikili öznitelik grubu için ürettikleri en iyi sınıflandırma sonuçları aynı oranda ve %53.01 olarak verilmiştir. Bu sonuç aynı zamanda TELESCOPE veri kümesi için sınıflandırma algoritmaları tarafından elde edilen en iyi sınıflandırma sonuçları arasındaki en düşük sonuçtur.

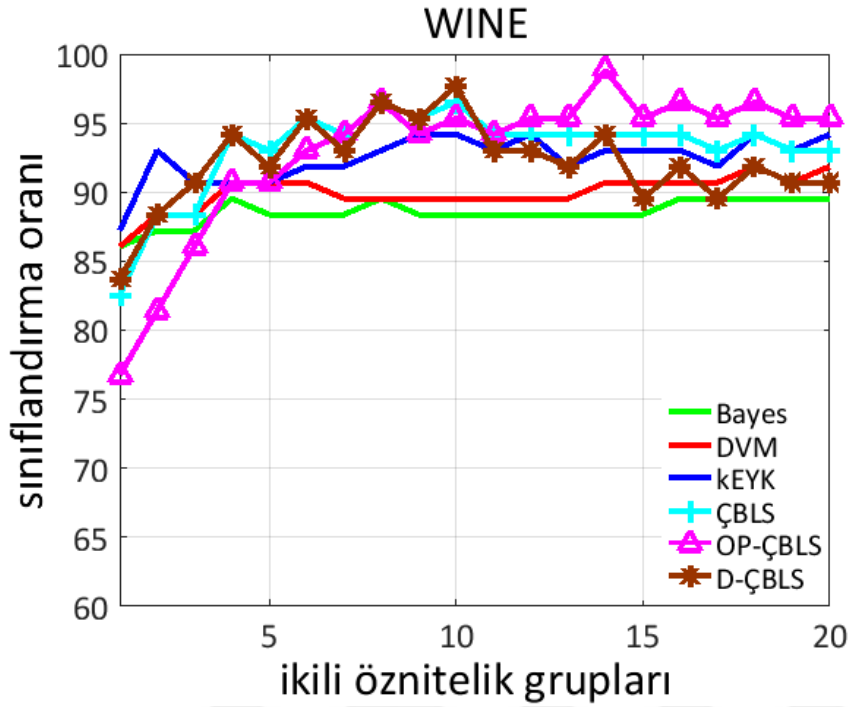
Denemelerde TELESCOPE için en iyi sınıflandırma başarısı önerilen yeni D-ÇBLS yöntemi ile toplam 20 iterasyonun dördüncü adımında ($g_{opt}=4$ iken) %70.68 oranında elde edilmiştir. Aynı durum için ÇBLS ve OP-ÇBLS yöntemlerinin en iyi sınıflandırma sonuçları sırasıyla %68.42 ve %69.92 olarak verilmiştir. Bu sonuçlara en yakın sınıflandırma sonucu %66.17 oranında k-EYK algoritması ile verilmiştir ve bu aynı zamanda k-EYK yönteminin TELESCOPE kümesi için ürettiği en iyi sınıflandırma başarısıdır. Sınıflandırma sonuçlarına göre TELESCOPE veri kümesi için k-EYK yönteminin verdiği en iyi sınıflandırma başarısı en iyi sıkıştırma parametrelerinin kullanımı ile %3 ve 45° açıyla döndürme sonrasında %4 oranında sırasıyla OP-ÇBLS ve D-ÇBLS ile artırılmıştır. TELESCOPE için ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemleri ile elde edilen sayısal sınıflandırma sonuçlarının görsel olarak karşılaştırması Şekil 4.14'te verilmiştir.

Sonuçlara göre ÇBLS algoritmasının seçilen 20 ikili öznitelik gruplarının ilk ikisi dışında diğer sınıflandırıcılara göre daha başarılı sonuç üretmektedir. Aynı durum için D-ÇBLS ve OP-ÇBLS yöntemleri ilk iterasyon adımlarında, yani g değeri 1 ile 5 arasında değişirken yüksek orandaki sınıflandırma sonuçlarını üretmektedirler ve iterasyon artımında sınıflandırma sonuçlarında azalma izlenmektedir. Ayrıca, TELESCOPE için ÇBLS, D-ÇBLS ve OP-ÇBLS sınıflandırma yöntemleri ilk iterasyonda, yani sadece iki öznitelikle ($g=1$) DVM ve Bayes yöntemlerinin tüm ikili gruplarındaki en iyi sınıflandırma başarısından daha yüksek sonuç üretmektedir.



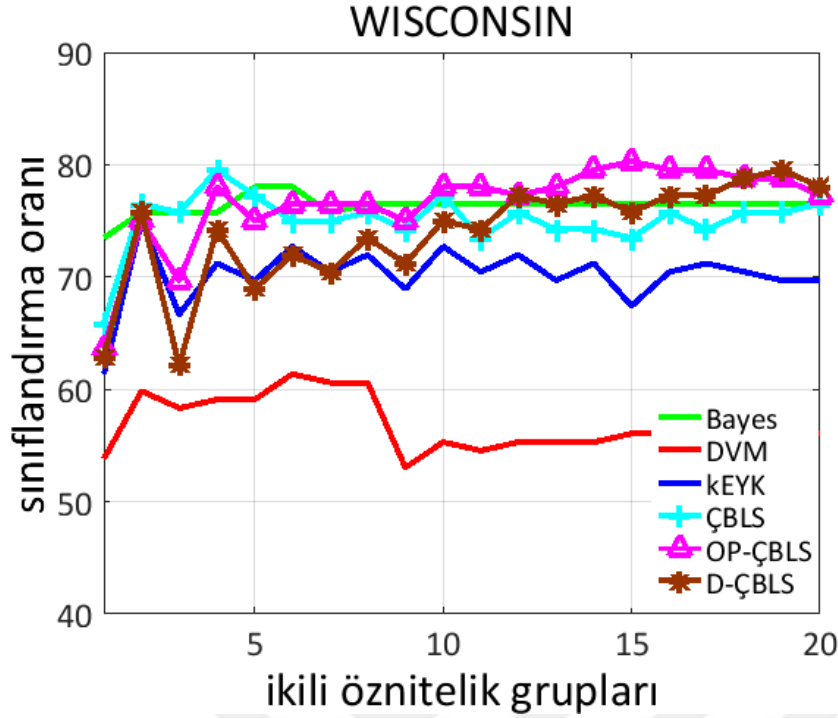
Şekil 4.14. ÇBLS yönteminin TELESCOPE veri kümesi için sınıflandırma sonuçları

WINE veri kümesi için ÇBLS ve D-ÇBLS yöntemlerinin en iyi sınıflandırma başarıları sırasıyla %96.51 ve %97.67 olarak verilmiştir. Aynı durum için Bayes ve DVM algoritmalarının ürettiği en iyi sınıflandırma başarıları sırasıyla %89.53 ve %91.86 olarak elde edilmiştir. Denemelerde k-EYK algoritmasının LÖS ile WINE kümesinden seçilen 20 ikili öznitelikler için ürettiği sonuçlardan en iyisi %94.19 olarak elde edilmiştir ve bu sonuç aynı zamanda WINE veri kümesi için klasik sınıflandırma algoritmalarının ürettiği en iyi sonuçtur. Bununla birlikte, WINE için klasik yöntemler tarafından üretilen sınıflandırma sonuçlarının yüksek olmasına rağmen ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemleri k-EYK yönteminin çıktısını sırasıyla %2, %3 ve %4 oranında arttırarak daha başarılı sınıflandırma sonucu vermişlerdir. Bunlara ek olarak, Şekil 4.15'ten açıkça görsel olarak görüldüğü gibi önerilen OP-ÇBLS algoritması seçilen ilk üç ikili öznitelik grubu dışındaki diğer iterasyonlarda klasik sınıflandırıcılarla göre daha başarılı sonuç üretmektedir. WINE veri kümesi için OP-ÇBLS yöntemi en iyi sınıflandırma başarısını $g_{opt}=14$ iken vermiştir. D-ÇBLS ve ÇBLS yöntemlerinin her ikisi de WINE için en yüksek sınıflandırma sonuçlarını $g_{opt}=10$ iken vermişlerdir. Denemede kullanılan klasik sınıflandırıcı sonuçları ile önerilen ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemlerin sınıflandırma sonuçlarının görsel karşılaştırması Şekil 4.15'te verilmiştir.



Şekil 4.15. ÇBLS yönteminin WINE veri kümesi için sınıflandırma sonuçları

Lorentz uzayında çok boyutlu verilerin sınıflandırılması amacıyla bu çalışmada önerilen ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemlerinin performansını değerlendirmek için son olarak WISCONSIN veri kümesi kullanılmıştır. WISCONSIN veri kümesinden LÖS ile seçilen 20 ikili öznitelik grupları için DVM tarafından üretilen sonuçların en iyisi %61.36 oranında verilmiştir. Aynı durum için önerilen ÇBLS sınıflandırma algoritması %79.55 oranında sınıflandırma başarısı üretirken, bu oranla aynı sınıflandırma başarısı döndürme tabanlı D-ÇBLS yönteminin WISCONSIN için ürettiği en iyi çıktı olarak elde edilmiştir. WISCONSIN veri kümesi için Eş. 3.17 ile toplamda 528 ikili öznitelik grubu elde edilmiştir ve LÖS yöntemi ile 20 ikili seçilmiştir ve ÇBLS yöntemi için sadece ilk 4'ü ($g_{opt}=4$) en iyi sınıflandırma başarısını elde etmeye yeterli olmuştur. Bu durum ÇBLS yönteminin çalışma zamanını önemli ölçüde azaltmaktadır. D-ÇBLS ve OP-ÇBLS yöntemlerinin en iyi sınıflandırma sonuçları sırasıyla $g_{opt}=19$ ve $g_{opt}=15$ iken elde edilmiştir. Bayes ve k-EYK algoritmalarının WISCONSIN veri kümesi için ürettiği en iyi sınıflandırma sonuçları sırasıyla %77.27 ve %61.36 olarak verilmiştir. Aynı durum için en iyi sınıflandırma başarısı OP-ÇBLS yöntemi ile %80.30 olarak verilmiştir. WISCONSIN veri kümesi için önerilen ÇBLS, D-ÇBLS ve OP-ÇBLS algoritmaları ve klasik sınıflandırıcı algoritmalarının sonuçlarının görsel karşılaştırması Şekil 4.16'da verilmiştir.



Şekil 4.16. ÇBLS yönteminin WISCONSIN veri kümesi için sınıflandırma sonuçları

Sonuç olarak, bu çalışmada kullanıma açık çok boyutlu verilerin Lorentz uzayında sınıflandırılması için önerilen ÇBLS, OP-ÇBLS ve D-ÇBLS yöntemlerinin sınıflandırma performansını test etmek amacıyla denemelerde GESTURE, LVST, MADELON, RELAX, SONAR, SEEDS, TELESCOPE, WINE ve WISCONSIN veri kümeleri kullanılmıştır. Denemelerde elde edilen sınıflandırma sonuçlarına göre ÇBLS, OP-ÇBLS ve D-ÇBLS yöntemlerinin performansı analiz edilerek klasik sınıflandırıcılara göre daha başarılı oldukları tespit edilmiştir. Böylece, Lorentz uzayında ikiden çok boyutlu verilerle çalışan yeni ÇBLS, OP-ÇBLS ve D-ÇBLS sınıflandırma yöntemleri literatüre kazandırılmıştır. Denemelerde kullanılan tüm veri kümeleri için yeni ÇBLS, OP-ÇBLS ve D-ÇBLS yöntemlerinin ürettiği en iyi sınıflandırma sonuçları diğer yöntemlerin çıktıları ile kıyasla sayısal olarak aşağıdaki Çizelge 4.21’de verilmiştir.

Çizelge 4.21’den görüldüğü gibi çok boyutlu verilen Lorentz uzayında sınıflandırılması için geliştirilen ÇBLS, OP-ÇBLS ve D-ÇBLS yöntemleri diğer klasik sınıflandırıcılara göre daha başarılı sonuçlar üretmektedirler. Ancak, denemelerde kullanılan 9 veri kümesinin altısında OP-ÇBLS yöntemi ve diğer üçünde D-ÇBLS yöntemi en iyi sınıflandırma başarısını üretmektedir. Bu nedenle uygulamada çok boyutlu verilen Lorentz uzayında sınıflandırılmasında OP-ÇBLS yönteminin daha etkili olduğu görülmüştür.

Çizelge 4.21. ÇBLS, D-ÇBLS ve OP-ÇBLS yöntemlerinin en iyi sınıflandırma sonuçlarının karşılaştırılması

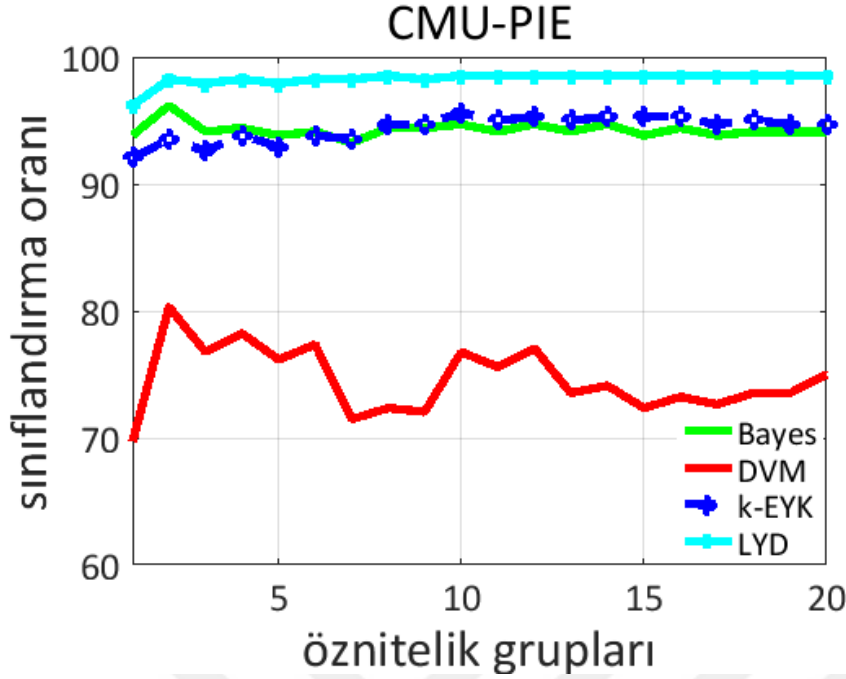
	Bayes	DVM	K-EYK	ÇBLS	OP-ÇBLS	D-ÇBLS
GESTURE	93.29	67.45	85.23	94.63	96.64	95.97
LVST	37.65	34.12	64.71	67.06	68.24	67.06
MADELON	52.99	56.72	58.21	61.94	62.69	64.93
RELAX	71.31	51.64	71.31	74.59	75.41	76.23
SEEDS	96.81	95.74	96.81	96.81	97.87	96.81
SONAR	51.08	53.96	45.32	62.59	63.31	58.27
TELESCOPE	53.01	53.01	66.17	68.42	69.92	70.68
WINE	89.53	91.86	94.19	96.51	98.84	97.67
WISCONSIN	78.03	61.36	75.00	79.55	80.30	79.95

4.3. Yüz Doğrulama Sonuçları

Bu bölümde Algoritma-10 ile tanımlanan Lorentz Yüz Doğrulama yönteminin doğruluğunu ve çalışma performansını değerlendirmek amacıyla yapılan uygulama sonuçları verilmiştir. Önerilen LYD yöntemi Lorentz uzayında çok boyutlu yüz veri kümelerini kullanarak yüz doğrulama yapmaktadır. Bu doğrultuda denemelerde genel kullanıma açık olan CMU PIE, ORL ve Yale B yüz veri kümeleri kullanılmıştır. Denemelerde kullanılan yüz veri kümeleri farklı sayıda sınıflara ait olan yüz resimlerini içermektedir. Dolayısıyla önerilen yeni LYD algoritmasında CMU PIE, ORL ve Yale B veri kümelerindeki U_j sınıfa ait yüzler doğru ve diğerleri yanlış resimler olarak eğitim yapılmaktadır. Bu araştırmada ayrıca, LYD yönteminin ürettiği sonuçlar k-En Yakın Komşu (k-EYK), Destek Vektör Makinesi (DVM) ve Bayes gibi klasik yöntemlerin sonuçları ile kıyaslanarak önerilen yeni yöntemin geçerliliği test edilmiştir. Bu bağlamda, gerçekleştirilen yöntem kıyaslama adımının anlamlı olmasını sağlamak adına klasik algoritmalar için uygulanan işlemlerin en iyi ikili öznitelik alt gruplarını seçme adımı haricindeki tümü işlemler LYD algoritması ile aynıdır. Klasik algoritmalar için denemelerde LÖS yönteminin Öklid analoğu kullanılmıştır.

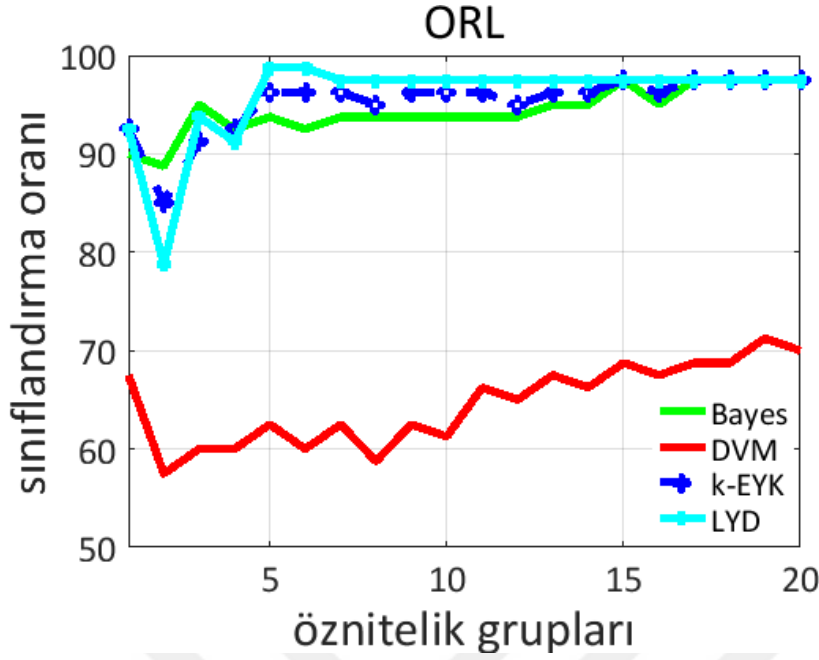
Araştırmada kullanılan CMU PIE, ORL ve Yale B yüz veri kümelerinin her biri toplamda 484 öznitelik içermektedir. Bu araştırmada önerilen LYD yönteminin önışlem adımında yüz veri kümelerinden Eş. 3.17 ile toplamda 116886 tane ikili öznitelik grupları elde edilmektedir. Denemelerde $g=20$ olarak alınarak ve Lorentz Öznitelik Seçme yöntemi ile 20 ikili grup seçilmiştir. Bu araştırmada önerilen LYD yönteminin CMU-PIE yüz veri tabanından seçilen 20 ikili öznitelik grupları için ürettiği çıktılar ile aynı durum için diğer klasik yöntemlerin ürettiği doğrulama sonuçlarının görsel karşılaştırılması Şekil 4.17’de verilmiştir.

Uygulamada CMU-PIE yüz veri kümesi için Bayes yönteminin seçilen 20 ikili öznitelik grupları için ürettiği en iyi doğrulama oranı %96.18 olarak verilmiştir. Aynı durum için DVM yönteminin en iyi doğrulama sonucu %80.29 oranında elde edilmiştir. CMU-PIE için DVM yönteminin en iyi doğrulama başarısını %15 arttıran k-EYK yöntemi %95.59 oranında en iyi doğrulama başarısını yakalamıştır. Ancak, LÖS ile CMU-PIE kümesinden seçilen 20 ikili öznitelik grupları için en iyi yüz doğrulama başarısı önerilen LYD yöntemi ile %98.53 olarak verilmiştir. Ayrıca, Bayes ve k-EYK yöntemlerinin doğrulama oranlarının %95 üzerinde olmasına rağmen onların CMU-PIE için ürettikleri en iyi sonuçlarını önerilen LYD yöntemi sırasıyla %2 ve %3 oranında arttırmıştır. Bu tez çalışmasının 3.5. bölümünde belirtildiği gibi LYD yönteminde çok sınıflı yüz veri kümeleri doğru ve yanlış yüz resimleri olarak, yani iki sınıflı yüz doğrulama örnekleri olarak ele alınmıştır. Bu bağlamda iki sınıflı veri kümelerinde çok iyi başarı gösteren DVM yöntemi CMU-PIE yüz doğrulama veri kümesinden seçilen 20 ikili öznitelik gruplarının tamamında diğer yöntemlere göre oldukça düşük sonuç üretmektedir. Denemelerde elde edilen doğrulama sonuçlarından görüldüğü gibi CMU-PIE için en iyi yüz doğrulama başarısını LYD yöntemi %98.53 oranında üretmektedir ve bu oranı yakalamak için yeni LYD için LÖS yöntemi ile seçilen toplam 20 en iyi ikili öznitelik grubunun 8 tanesi, yani 16 en iyi öznitelik yeterli olmuştur. Bu durum önerilen yöntemin çalışma zamanını önemli ölçüde azaltarak, etkililiğini arttırmaktadır. Ayrıca, önerilen yöntem seçilen ikili öznitelik gruplarının tamamında diğer yöntemlere göre daha başarılı doğrulama başarısını üretmektedir.



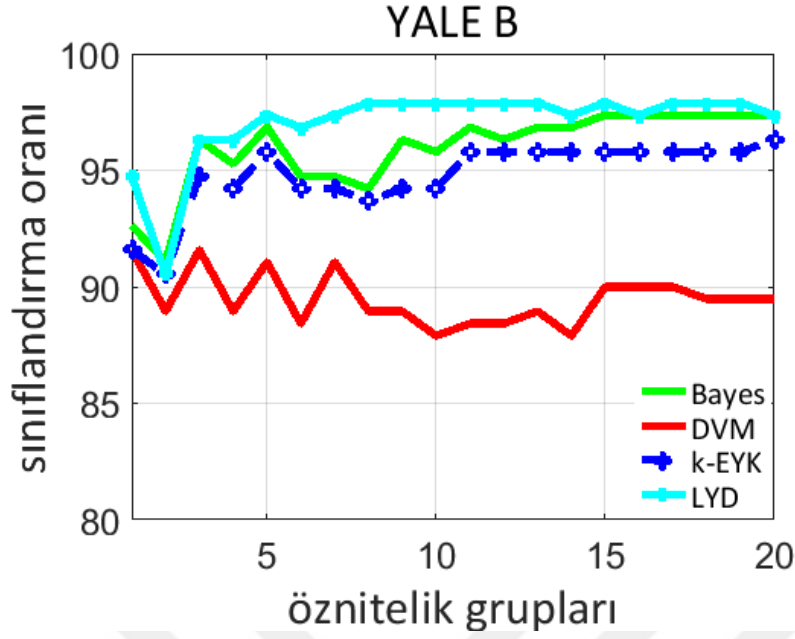
Şekil 4.17. CMU-PIE için yüz doğrulama oranları

Lorentz uzayında çok boyutlu yüz veri kümeleri ile yüz doğrulama amaçlı geliştirilen LYD yöntemini test etmek amacıyla denemelerde kullanılan diğer bir yüz veri kümesi ORL için k-EYK yönteminin en iyi doğrulama başarısı %97.50 oranında verilmiştir ve bu oran oldukça yüksek sonuçtur. Aynı durum için Bayes yönteminin en iyi doğrulama oranı k-EYK yönteminin en iyi doğrulama oranına eşit olarak elde edilmiştir. ORL yüz veri kümesi için DVM yönteminin en iyi doğrulama başarısı %71.25 oranında verilmiştir. DVM yöntemi ORL kümesinden LÖS ile seçilen 20 öznitelik ikili gruplarının tamamında diğer yöntemlere göre daha düşük doğrulama başarısını üretmektedir. LYD yöntemi ORL için seçilen $g=20$ en iyi ikili öznitelik gruplarında en iyi doğrulama başarısını %98.75 oranında vermiştir. LYD yöntemi bu sonucu $g=5$ iken verirken Bayes ve k-EYK yöntemleri ORL yüz veri kümesi için ürettiği en iyi doğrulama başarısını $g=15$ iken üretmektedirler. Ayrıca, önerilen LYD yöntemi ORL veri kümesi için ilk beş ikili öznitelik grupları dışında diğer yöntemlere göre daha başarılı sonuç üretmektedir. Bu durum aşağıdaki Şekil 4.18'den görsel olarak daha açık görülmektedir.



Şekil 4.18. ORL için yüz doğrulama oranları

Denemelerde kullanılan Yale B yüz veri kümesi için k-EYK yönteminin ürettiği en iyi doğrulama başarısı %96.32 olarak verilmiştir. Aynı durum için Bayes yöntemi %97.37 oranında en iyi başarı üretirken, DVM yöntemi Yale B kümesinden seçilen 20 ikili öznitelik grupları için en iyi doğrulama başarısını %91.58 oranında vermiştir. Önerilen yeni LYD yöntemi Yale B veri kümesi için en iyi doğrulama başarısını %97.89 oranında üretmiştir. Bununla birlikte ORL kümesinden seçilen 20 ikili öznitelik gruplarının tamamına yakınında önerilen LYD yöntemi diğer yöntemlere göre daha başarılı sonuç üretmektedir. Ayrıca, önerilen LYD yöntemi en iyi doğrulama başarısını $g = 8$ iken, yani en iyi 8 ikili öznitelik grup veya 16 öznitelik ile yakalamıştır. Yani, önerilen LYD yöntemi Yale B yüz veri kümesindeki toplam 484 öznitelikten en iyi 16 özniteliği Lorentz metriğine göre seçerek %97.89 oranında oldukça yüksek ve diğer yöntemlere göre daha başarılı sonuç üretmektedir. Bu durum Lorentz metriği temelindeki LÖS yönteminin etkili bir öznitelik seçme ve boyut indirgeme yöntemi olduğunu göstermektedir. Ayrıca, LYD yöntemin yüz doğrulama probleminde etkili bir araç olduğunu kanıtlamaktadır. Yale B yüz veri kümesinden seçilen 20 ikili öznitelik grupları için önerilen LYD yönteminin g değerinin artımındaki sonuçlarının görsel olarak karşılaştırması Şekil 4.19’da verilmiştir.



Şekil 4.19. Yale B için yüz doğrulama oranları

Denemelerde CMU PIE, ORL ve Yale B yüz veri tabanları için LÖS yöntemi ile seçilen en iyi 10 ikili öznitelik grupları için k-EYK, Bayes, DVM ve Lorentz Yüz Doğrulama yöntemlerinin doğrulama başarılarının sayısal karşılaştırması sırasıyla Çizelge 4.22, Çizelge 4.23 ve Çizelge 4.24'te verilmiştir. Elde edilen sonuçlara göre yüz veri tabanları için Lorentz metriğine göre sınıflandırma yapan LYD algoritması diğer yöntemlere göre daha başarılı doğrulama oranı verdiği görülmüştür.

Çizelge 4.22. CMU-PIE yüz veri tabanından seçilen 10 ikili öznitelik grubu için sınıflandırma sonuçları

	1	2	3	4	5	6	7	8	9	10
Bayes	93.82	96.18	94.12	94.41	93.82	94.12	93.24	94.41	94.41	94.71
DVM	69.71	80.29	76.76	78.24	76.18	77.35	71.47	72.35	72.06	76.76
k-EYK	92.06	93.53	92.65	93.82	92.94	93.82	93.53	94.71	94.71	95.59
LYD	96.18	98.24	97.94	98.24	97.94	98.24	98.24	98.53	98.24	98.53

Çizelge 4.23. ORL yüz veri tabanından seçilen 10 ikili öznitelik grubu için sınıflandırma sonuçları

	1	2	3	4	5	6	7	8	9	10
Bayes	90.00	88.75	95.00	92.50	93.75	92.50	93.75	93.75	93.75	93.75
DVM	67.50	57.50	60.00	60.00	62.50	60.00	62.50	58.75	62.50	61.25
k-EYK	92.50	85.00	91.25	92.50	96.25	96.25	96.25	95.00	96.25	96.25
LYD	92.50	78.75	93.75	91.25	98.75	98.75	97.50	97.50	97.50	97.50

Çizelge 4.24. Yale B yüz veri tabanından seçilen 10 ikili öznitelik grubu için sınıflandırma sonuçları

	1	2	3	4	5	6	7	8	9	10
Bayes	92.63	91.05	96.32	95.26	96.84	94.74	94.74	94.21	96.32	95.79
DVM	91.58	88.95	91.58	88.95	91.05	88.42	91.05	88.95	88.95	87.89
k-EYK	91.58	90.53	94.74	94.21	95.79	94.21	94.21	93.68	94.21	94.21
LYD	94.74	90.53	96.32	96.32	97.37	96.84	97.37	97.89	97.89	97.89

Çok boyutlu yüz resimlerini kullanarak yüz doğrulama problemini Lorentz uzayında uygulamak amacıyla geliştirilen Lorentz Yüz Doğrulama yönteminin denemelerde elde edilen sonuçlara göre klasik yöntemlere göre daha başarılı olduğu görülmüştür. Geliştirilen LYD yönteminin CMU PIE, ORL ve Yale B yüz veri kümeleri üzerinde yapılan denemelerde elde edilen en iyi yüz doğrulama başarılarının klasik yöntemlerin çıktıları ile karşılaştırması aşağıdaki Çizelge 4.25'te verilmiştir.

Çizelge. 4.25. Yüz veri kümeleri için LYD yönteminin en iyi yüz doğrulama oranlarının karşılaştırması

	DVM	Bayes	k-EYK	LYD
CMU-PIE	80.29	96.18	95.59	98.53
ORL	71.25	97.50	97.50	98.75
Yale B	91.58	97.37	96.32	97.89

4.4. Lorentz Metriği Temelli Destek Vektör Makinesinin Sonuçları

Bu bölümde Lorentz uzayında sınıflandırma yapan yeni Lorentz Destek Vektör Makinesi yönteminin LVST, SONAR, TELESCOPE ve WISCONSIN veri kümeleri ile denemeleri yapılmıştır. Önerilen LDVM yöntemi ayırt etme ölçütü olarak Lorentz metriğini kullanmaktadır ve klasik DVM yönteminin çalışma mantığı ile modellenmiştir. LDVM yöntemi bu tez çalışmasının 2.2. bölümünde tanımlanan Lorentz uzayındaki komşuluk kavramı temelinde oluşturulan hiperdüzlem ile destek vektörleri arasındaki uzaklığı maksimize etmeye çalışmaktadır ve böylece farklı sınıfa ait örnekleri birbirinden ayırt etmektedir. Yani bu alanı en maksimum seviyede ayırt edecek olan en iyi hiperdüzlemi Lorentz uzayının özel yapısı ve metriğine göre belirlemektedir. Bunun için de Lagrange çarpanları sınıf etiket bilgileri ve eğitim örnekleri ile çözülen amaç fonksiyonunu, yani lineer kernel fonksiyon içermektedir. Bu kernel fonksiyon iç çarpım işlemini içermektedir ve iç çarpım işlemi Lorentz uzayına göre yapılandırılmıştır. Yani örnekler arasındaki ayırt edici ölçüt olarak kullanılan destek vektörler ile hiperdüzlem arasındaki uzaklığı yeni Lorentz metriğine göre değiştirilmiş kernel fonksiyonu ile gerçekleştirerek optimizasyon problemini çözmektedir. Böylece, Lagrange çarpanlarını bularak Lorentz metriğine göre sınıflandırma gerçekleştirmektedir.

Bu tez araştırmasında önerilen yeni LDVM yöntemi ile elde edilen sonuçlar k-En Yakın Komşu (k-EYK), Destek Vektör Makinesi (DVM) ve Bayes gibi klasik yöntemlerin sonuçları ile karşılaştırılmıştır. İlk olarak, söz konusu klasik yöntemlerin yalın hali ile orijinal veri kümelerinden elde edilen sonuçlar değerlendirilmiştir. Örneğin, SONAR veri kümesi için Bayes, k-EYK ve DVM yöntemleri sırasıyla %46.76, %50.36 ve %52.52

oranında sınıflandırma başarısını üretmektedir. Aynı veri kümesi için önerilen LDVM yöntemi %67.63 oranında sınıflandırma başarısı göstermiştir. Bu çalışmada önerilen LDVM yönteminin LVST, SONAR, TELESCOPE ve WISCONSIN veri kümelerindeki sınıflandırma sonuçları ile bu kümelerin yalın halindeki klasik sınıflandırıcıların sınıflandırma sonuçlarının sayısal karşılaştırması Çizelge 4.26'da verilmiştir. Çizelge 4.26'dan görüldüğü gibi Lorentz uzayında sınıflandırma yapan LDVM yöntemi veri kümelerinden rasgele seçilen 15 ikili öznitelik grupları için ürettiği sınıflandırma sonuçları aynı veri kümelerinde klasik sınıflandırıcıların özniteliklerin tamamını kullandığı durum için ürettiği sonuçlara göre daha başarılıdır.

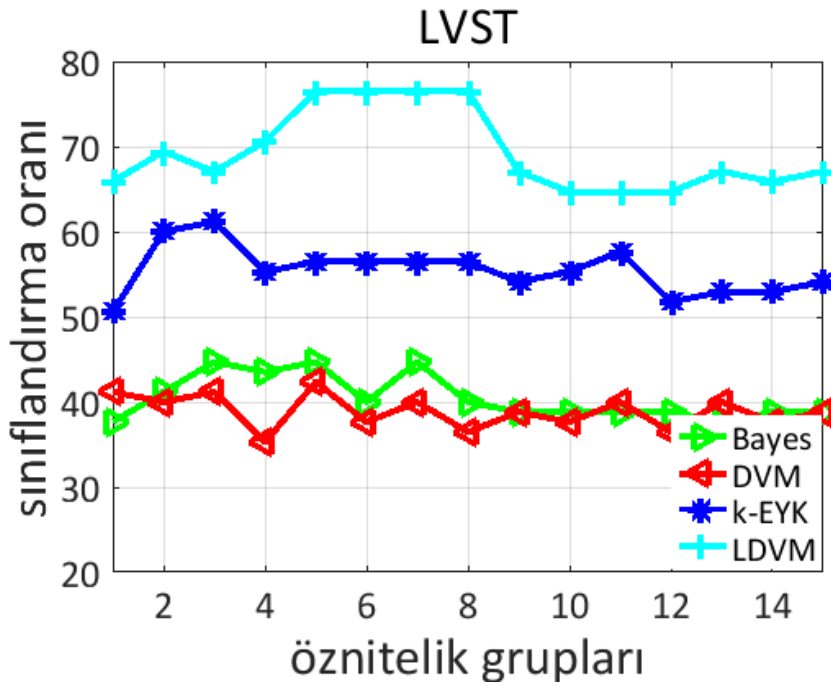
Çizelge 4.26. Klasik sınıflandırıcıların veri kümelerinin yalın halindeki sınıflandırma sonuçları

	Bayes	DVM	k-EYK	LDVM
LVST	44.71	69.41	58.82	76.47
SONAR	46.76	52.52	50.36	67.63
TELESCOPE	56.39	53.38	61.65	64.66
WISCONSIN	50.76	61.36	64.39	65.91

Ancak, önerilen LDVM algoritması önişlem içermektedir. Bu nedenle, bundan sonraki denemelerde karşılaştırmanın anlamlı olması adına klasik yöntemler için de önişlem, ikili öznitelik grupları ve buna bağlı karar mekanizması kullanılmıştır. Bu makalede kullanılan veri kümeleri değişik sayıdaki öznitelige sahiptirler ve Eş. 3.17 ile farklı sayıdaki ikili öznitelik grupları elde edilmektedir. Ancak, denemelerde toplam ikili grup sayısından sadece $g=15$ tanesi rasgele seçilmektedir.

LVST veri kümelerinden elde edilen $g=15$ ikili öznitelik grupları için önerilen LDVM ve diğer klasik yöntemlerin ürettiği sınıflandırma sonuçlarının görselleştirilmesi Şekil 4.20'de verilmiştir. Şekil 4.20'den görüldüğü gibi LVST veri kümesi için DVM ve Bayes yöntemleri için en iyi sınıflandırma başarıları sırasıyla %42.35 ve %44.71 oranında verilirken, k-EYK için bu oran %61.18 olarak elde edilmiştir. LVST için en iyi sınıflandırma sonucu önerilen LDVM algoritması tarafından $g = 5$ iken %76.47 oranında verilmiştir.

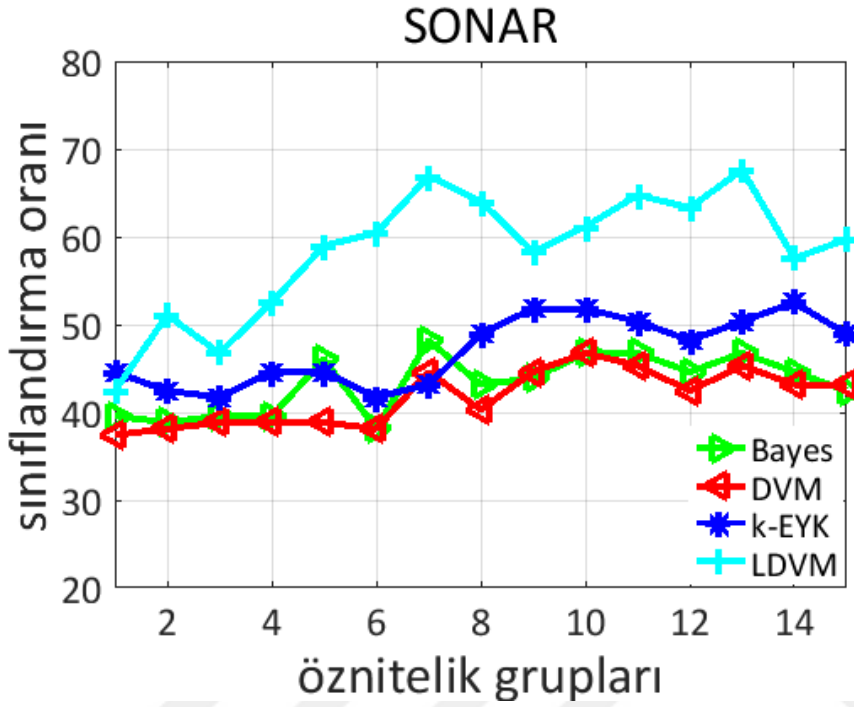
Denemelerde LVST veri kümesinde elde edilen ikili öznitelik gruplarından rasgele seçilen 15 öznitelik ikilisinin tamamında önerilen LDVM yönteminin diğer yöntemlere göre daha başarılı sonuç ürettiği görülmüştür. Bununla birlikte LVST veri kümesi için önerilen LDVM algoritması Öklid uzayındaki analoğu olan DVM yönteminin en iyi sınıflandırma sonucunu %34 oranında arttırmıştır. Bu oran Bayes ve k-EYK yöntemleri için sırasıyla %32 ve %16 olarak kaydedilmiştir. Bu durum Lorentz metriği temelinde geliştirilen LDVM yöntemini sınıflandırma probleminin çözümünde önemli bir araç konumuna getirmektedir.



Şekil 4.20. LVST veri kümesi için LDVM yönteminin sınıflandırma sonuçlarının karşılaştırması

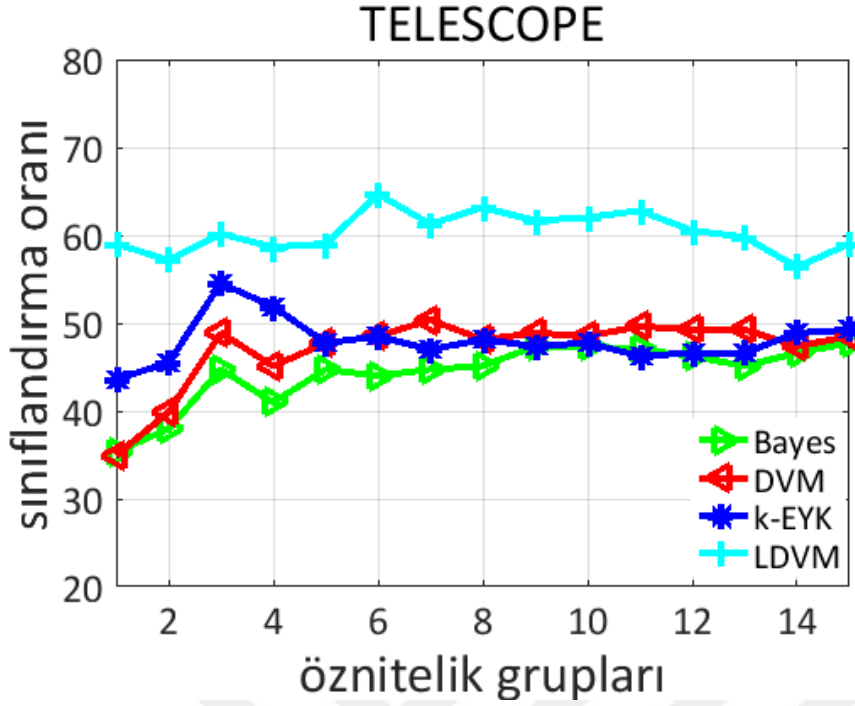
SONAR veri kümesi için DVM ve Bayes yöntemlerinin ürettiği en iyi sınıflandırma sonuçları yaklaşık %2 fark ile sırasıyla %46.74 ve %48.20 oranında verilmiştir. Aynı durum için k-EYK yönteminin en iyi başarısı %52.52 oranında elde edilmiştir ve bu sonuca göre %15 oranında daha yüksek sınıflandırma başarısı üreten LDVM yöntemi SONAR için en iyi sınıflandırma başarısını %67.63 oranında vermiştir. SONAR veri kümesinden elde edilen 15 ikili öznitelik grupları için önerilen LDVM yöntemimin sınıflandırma sonuçları ile klasik sınıflandırıcıların sonuçlarının görsel karşılaştırması Şekil 4.21’de verilmiştir.

Elde edilen sonuçlardan görüldüğü gibi SONAR kümesi için önerilen LDVM yöntemi ilk ikili öznitelik haricinde diğer öznitelik gruplarının tamamında Bayes, DVM ve k-EYK yöntemlere göre daha başarılı sınıflandırma sonuçlarını üretmiştir.



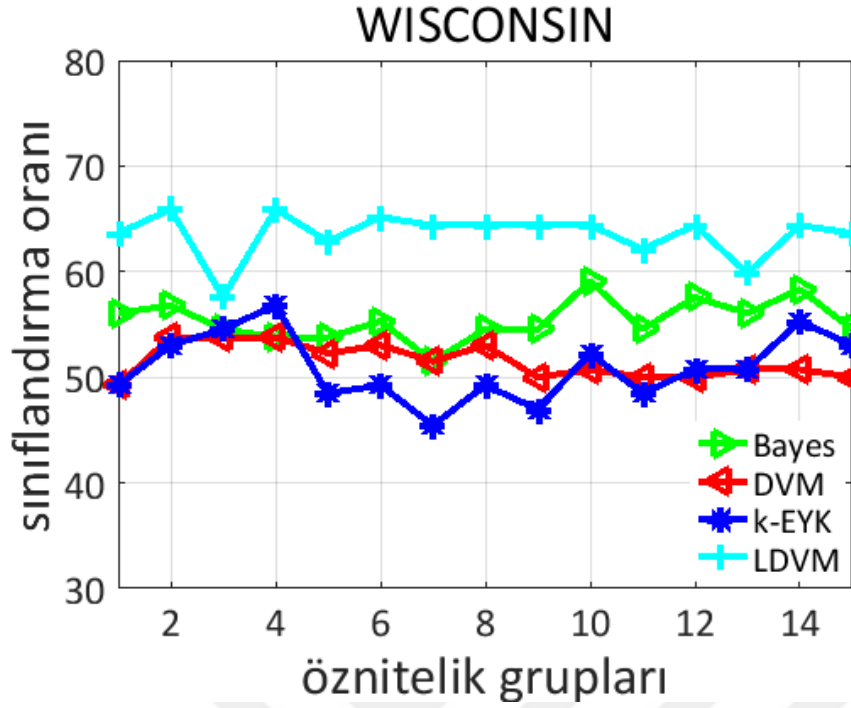
Şekil 4.21. SONAR veri kümesi için LDVM yönteminin sınıflandırma sonuçlarının karşılaştırması

Denemelerde kullanılan TELESCOPE veri kümesi için Bayes, DVM ve k-EYK yöntemlerinin ürettiği en iyi sınıflandırma sonuçları sırasıyla %47.74, %50.38 ve %54.51 olarak verilmiştir. Aynı şartlar için %64.66 oranında başarı üreten yeni LDVM yöntemi kendisine en yakın sonucu veren k-EYK yönteminin sınıflandırma başarısını %10 oranında arttırarak kesin üstünlüğünü ortaya koymaktadır. TELESCOPE veri kümesi için elde edilen sınıflandırma sonuçlarından görüldüğü gibi LDVM yöntemi klasik sınıflandırıcıların başarılarına göre ortalama %12 oranında daha yüksek sonuç vermiştir. Bununla birlikte rasgele seçilen ikili öznitelik gruplarının sadece ilk altısı ile TELESCOPE veri kümesi için önerilen LDVM yöntemi en iyi sınıflandırma başarısını verebilmektedir. Bu durum önerilen LDVM yönteminin çalışma zamanı açısından ve az sayıdaki öznitelikleri ile en iyi sınıflandırma başarısını elde etme açısından da etkili bir yöntem olduğunu ortaya koymaktadır. Lorentz uzayında verilerin sınıflandırılması amacıyla geliştirilen LDVM yönteminin TELESCOPE veri kümesi için ürettiği sınıflandırma sonuçlarının diğer yöntemlerin çıktıları ile karşılaştırması Şekil 4.22’de verilmiştir.



Şekil 4.22. TELESCOPE veri kümesi için LDVM yönteminin sınıflandırma sonuçlarının karşılaştırması

LDVM yönteminin test edilmesinde kullanılan son veri kümesi WISCONSIN için DVM ve Bayes yöntemlerinin ürettiği en iyi sınıflandırma sonuçları sırasıyla %53.79 ve %56.82 olarak verilmiştir. Aynı durum için elde edilen iyi sınıflandırma başarısı %65.91 oranında yeni LDVM tarafından elde edilmiştir. Bu sonuca en yakın sınıflandırma başarısı ve aynı zamanda Bayes algoritmasının en iyi sonucu %59.09 olarak elde edilmiştir. Deneysel sonuçlardan görüldüğü gibi LDVM yöntemi WISCONSIN veri kümesi için Bayes, DVM ve k-EYK yöntemlerinin ürettiği en iyi sınıflandırma başarılarını sırasıyla %6, %12 ve %9 oranında ve ortalama %9 arttırmıştır. Bu sınıflandırma problemi için çok önemli bir artış olarak değerlendirilebilir. Ayrıca, önerilen LDVM yöntemi WISCONSIN veri kümesinde elde edilen 15 öznitelik ikilisinin tamamında diğer klasik sınıflandırıcılara göre daha başarılı sonuç üretmektedir. Benzer durum TELESCOPE, LVST veri kümelerinde de gözlemlenmiştir. WISCONSIN veri kümesi için klasik sınıflandırıcıların ürettiği sonuçların önerilen LDVM yönteminin sonuçları ile görsel karşılaştırması Şekil 4.23'te verilmiştir.



Şekil 4.23. WISCONSIN veri kümesi için LDVM yönteminin sınıflandırma sonuçlarının karşılaştırması

Sonuç olarak denemelerde elde edilen sınıflandırma sonuçlarına göre LDVM yöntemi diğer sınıflandırıcılara göre daha iyi sınıflandırma başarısını vermektedir. Üstte beyan edilen görsel sonuçlara ek olarak, LDVM yönteminin sınıflandırma sonuçları ile klasik Bayes, DVM ve k-EYK yöntemlerinin LVST, SONAR, TELESCOPE ve WISCONSIN veri kümelerinden seçilen 10 ikili öznitelik grupları için ürettiği çıktıların sayısal karşılaştırması sırasıyla Çizelge 4.27, Çizelge 4.28, Çizelge 4.29 ve Çizelge 4.30'da verilmiştir.

Çizelge 4.27. LVST kümesinden seçilen 10 ikili öznitelik grubu için LDVM yönteminin sınıflandırma sonuçları

	1	2	3	4	5	6	7	8	9	10
Bayes	37.65	41.18	44.71	43.53	44.71	40.00	44.71	40.00	38.82	38.82
DVM	41.18	40.00	41.18	35.29	42.35	37.65	40.00	36.47	38.82	37.65
k-EYK	50.59	60.00	61.18	55.29	56.47	56.47	56.47	56.47	54.12	55.29
LDVM	65.88	69.41	67.06	70.59	76.47	76.47	76.47	76.47	67.06	64.71

Çizelge 4.28. SONAR kümesinden seçilen 10 ikili öznitelik grubu için LDVM yönteminin sınıflandırma sonuçları

	1	2	3	4	5	6	7	8	9	10
Bayes	39.57	38.85	39.57	39.57	46.04	38.13	48.20	43.17	43.88	46.76
DVM	37.41	38.13	38.85	38.85	38.85	38.13	44.60	40.29	44.60	46.76
k-EYK	44.60	42.45	41.73	44.60	44.60	41.73	43.17	48.92	51.80	51.80
LDVM	42.45	51.08	46.76	52.52	58.99	60.43	66.91	64.03	58.27	61.15

Çizelge 4.29. TELESCOPE kümesinden seçilen 10 ikili öznitelik grubu için LDVM yönteminin sınıflandırma sonuçları

	1	2	3	4	5	6	7	8	9	10
Bayes	35.34	37.97	44.74	40.98	44.74	43.98	44.74	45.11	47.37	47.37
DVM	34.96	39.85	48.87	45.11	47.74	48.50	50.38	48.12	48.87	48.50
k-EYK	43.61	45.49	54.51	51.88	47.74	48.50	46.99	48.12	47.37	47.74
LDVM	59.02	57.14	60.15	58.65	59.02	64.66	61.28	63.16	61.65	62.03

Çizelge 4.30. WISCONSIN kümesinden seçilen 10 ikili öznitelik grubu için LDVM yönteminin sınıflandırma sonuçları

	1	2	3	4	5	6	7	8	9	10
Bayes	56.06	56.82	54.55	53.79	53.79	55.30	51.52	54.55	54.55	59.09
DVM	49.24	53.79	53.79	53.79	52.27	53.03	51.52	53.03	50.00	50.76
k-EYK	49.24	53.03	54.55	56.82	48.48	49.24	45.45	49.24	46.97	52.27
LDVM	63.64	65.91	57.58	65.91	62.88	65.15	64.39	64.39	64.39	64.39

Son olarak bu tez çalışmasında önerilen LDVM yönteminin LVST, SONAR, TELESCOPE ve WISCONSIN veri kümeleri için ürettiği en iyi sınıflandırma sonuçlarının klasik k-EYK

Bayes ve DVM yöntemlerinin en iyi sınıflandırma sonuçları ile karşılaştırması Çizelge 4.31’de verilmiştir.

Çizelge 4.31. LDVM yönteminin en iyi sınıflandırma sonuçları

	Bayes	DVM	k-EYK	LDVM
LVST	44.71	42.35	61.18	76.47
SONAR	48.20	46.76	52.52	67.63
TELESCOPE	47.74	50.38	54.51	64.66
WISCONSIN	59.09	53.79	56.82	65.91

6. SONUÇLAR VE ÖNERİLER

Bu tez araştırmasında Lorentz uzayının sınıflandırma probleminde kullanımı çok yönlü olarak araştırılmıştır. Bu doğrultuda Lorentz uzayının metriği ve özellikleri sınıflandırma açısından incelenerek analiz edilmiştir. Lorentz metriğinin Öklid analogundan farklı olması sebebiyle Lorentz uzayında noktaların dağılımı ve kendi aralarındaki ilişki Lorentz uzaklığı açısından irdelenmiştir. İlk olarak Lorentz uzayındaki noktalar arası komşuluk gösterilmiştir. Lorentz uzayındaki noktalar arası komşuluk kavramına göre noktaların dağılımı veya pozisyonunun sınıflandırma açısından önemli olduğu tespit edilmiştir. Örnekleri Öklid uzayında belirli bir açıya döndürerek Lorentz metriği açısından en iyi sınıflandırma başarısını üretecek en iyi pozisyona taşıma yöntemi geliştirilmiştir. Ayrıca, Destek Vektör Makinesi (DVM) veya Doğrusal Diskriminant Analizi (DDA) yöntemleri ile önce örnekleri ayırt eden karar doğrusu hesaplanmaktadır ve bu doğrunun yardımı ile örneklerin döndürülmesi gereken en iyi açı hesaplanmaktadır. Uygulamada DVM ve DDA ile karar doğruları ile en iyi döndürme açılarını hesaplayarak ve ona göre verileri en iyi pozisyona taşıyarak önişlemden geçiren ve Lorentz uzayında sınıflandırma yapan KÇ-GALS yönteminin diğer klasik yöntemlerle kıyasla daha başarılı sonuç ürettiği görülmüştür.

Lorentz uzaklığının sınıflandırma probleminde ayırt edici ölçüt olarak kullanım tekniği ortaya koyulmuştur. Bu doğrultuda matris çarpımı (sıkıştırma), döndürme ve öteleme gibi temel matematik işlemlerden yararlanarak Önişlem-1 önerilmiştir. Önişlem-1 adımı örneklerin iki boyutlu Lorentz uzayında Lorentz Uzaklığı Temelli Sınıflandırman (LUTS) yöntemi ile sınıflandırılması ve sınıflandırma başarısını arttırmak amacıyla geliştirilen ek yöntemidir. Önerilen LUTS yönteminde eğitim örnekleri sınıf ortalamalarına göre ayrı ayrı sıkıştırılarak 45° açıya döndürülmektedir. Yeni gelen test örneği sırasıyla her sınıfın ortalamasına göre sıkıştırılarak döndürülmektedir. Test örneklerinin sınıf etiketini belirleme Lorentz uzaklığı temelli ek karar verme mekanizması ile yapılmaktadır. Bu çalışmada ECOLI, ILPD, SEEDS, VERTEBRAL veri kümeleri ile yapılan denemelerde önerilen LUTS yönteminin Bayes, DVM ve k-EYK klasik sınıflandırıcılarına göre daha başarılı sınıflandırma sonucu ürettiği görülmüştür.

Bu tez araştırmasında Lorentz uzaklığı temelindeki Sıkıştırma ile Lorentz Sınıflandırıcısı (S-LS) yeni olarak geliştirilmiştir. Geliştirilen S-LS yönteminin genel kullanıma açık birkaç veri kümeleri ile geçerliliği ve doğruluğu test edilmiştir. Elde edilen sonuçlara göre S-LS

yönteminin sınıflandırma başarısı örneklerin dağılımına göre değişik oranlarda artmaktadır. Yani, iyi dağılmış ve diğer yöntemlerin yüksek sonuç ürettiği durumlarda az farkla başarıyı arttırmaktadır. Ancak, diğer yöntemlerin düşük sonuç ürettiği durumlarda, önerilen S-LS yönteminin başarısı önemli ölçüde artmaktadır. Bu oran bazen %20 oranında kaydedilmektedir.

Bu çalışmada önerilen S-LS yönteminin önişlem adımı döndürme ve sıkıştırma matrisinin en iyi parametrelerini hesaplama adımları ile genişletilerek sırasıyla D-LS ve OP-LS yöntemleri geliştirilmiştir. Yeni S-LS, D-LS ve OP-LS yöntemleri VERTEBRAL, RELAX, CLIMATE, PARKINSON, WINE ve GESTURE veri kümeleri ile test edilmiştir. Geliştirilen D-LS yönteminde döndürme açısı sabit 45° olarak alınmıştır. OP-LS yöntemindeki sıkıştırma matrisinin en iyi parametrelerinin hesaplanmasında en düşük sonuç üreten öznitelik ikilisi baz alınmıştır. Genel olarak önerilen yeni S-LS yöntemi ve genişletilmiş versiyonu olan D-LS ve OP-LS yöntemlerinin denemelerde diğer klasik sınıflandırıcılara göre daha başarılı sonuçlar ürettiği görülmüştür. Böylece, bu araştırmada Lorentz uzaklığının sınıflandırma probleminde kullanımı yöntem olarak ortaya koyulmuş ve bu yöntemin doğruluğu ispatlanmıştır.

Bu tez araştırmasında ikiden çok öznitelige sahip veri kümelerinin Lorentz uzayında sınıflandırılması amacıyla Lorentz metriği temelli yeni Lorentz Öznitelik Seçme (LÖS) yöntemi geliştirilmiştir. Önerilen yeni öznitelik seçme yöntemi sınıf içi ve sınıflar arası örneklerin dağılım matrislerinden oluşan ayırt edici ölçütünün Lorentz metriğine göre yapılandırılmış LJ versiyonu temelinde veri kümelerinden en iyi ikili öznitelik gruplarını seçmektedir ve sınıflandırma başarısını önemli ölçüde arttırmaktadır. Ayrıca, denemelerde çok boyutlu verilerde önerilen LÖS yönteminin işlem karmaşıklığını hissedilir seviyede azaltmaktadır.

Bu araştırmada önerilen LÖS ile seçilen en iyi ikili öznitelik gruplarının Lorentz uzayında sınıflandırılması için Çok Boyutlu Lorentz Sınıflandırıcısı (ÇBLS) geliştirilmiştir. Geliştirilen ÇBLS yöntemi seçilen ikili öznitelik gruplarının tümünü kullanarak sınıflandırma yapmaktadır ve bunun için çoğunluk tabanlı ve Lorentz uzaklığı temelinde ek karar verme mekanizması geliştirilmiştir. Bununla birlikte ÇBLS yönteminin önişlem adımı döndürme ve en iyi parametre hesaplama işlemleri ile geliştirilerek sırasıyla çok boyutlu veriler için D-ÇBLS ve OP-ÇBLS yöntemleri geliştirilmiştir. Geliştirilen yöntemlerde

döndürme ve en iyi parametre hesaplama işlemleri D-LS ve OP-LS yöntemlerinin adımları ile aynıdır.

Genel olarak, bu çalışmada kullanıma açık Çok boyutlu verilerin Lorentz uzayında sınıflandırılması için önerilen ÇBLS, OP-ÇBLS ve D-ÇBLS yöntemlerinin sınıflandırma performansını test etmek amacıyla denemelerde GESTURE, LVST, MADELON, RELAX, SONAR, SEEDS, TELESCOPE, WINE ve WISCONSIN veri kümeleri kullanılmıştır. Çok boyutlu verilen Lorentz uzayında sınıflandırılması için geliştirilen ÇBLS, OP-ÇBLS ve D-ÇBLS yöntemleri diğer klasik sınıflandırıcılara göre daha başarılı sonuçlar üretmektedirler. Denemelerde kullanılan 9 veri kümesinin altısında OP-ÇBLS yöntemi ve diğer üçünde D-ÇBLS yöntemi en iyi sınıflandırma başarısını üretmektedir. Bu nedenle uygulamada çok boyutlu verilen Lorentz uzayında sınıflandırılmasında OP-ÇBLS yönteminin daha etkili olduğu görülmüştür. Böylece, Lorentz uzayında ikiden çok boyutlu verilerle çalışan yeni ÇBLS, OP-ÇBLS ve D-ÇBLS sınıflandırma algoritmaları literatüre kazandırılmıştır.

Bu tez araştırmasında çok boyutlu yüz resimleri için yeni Lorentz Yüz Doğrulama (LYD) yöntemi geliştirilmiştir. Geliştirilen yeni LYD algoritması Lorentz uzayının uzaklık ölçütü ile doğru resim ile yanlış olanlarını benzerlik değerine göre ayırt etmektedir. Önerilen algoritma ayrıca Lorentz uzayının metriği temelinde geliştirilen Lorentz Öznitelik Seçme yöntemini içermektedir. Geliştirilen LÖS yöntemi orijinal yüz veri kümesini en iyi şekilde temsil edecek olan k en iyi ikili öznitelik kombinasyonlarını seçerek yüz doğrulama başarısını önemli ölçüde arttırmaktadır. Bununla birlikte önerilen yeni LYD algoritmasının geçerliliğini ve etkililiğini test etmek amacıyla CMU PIE, ORL ve Yale B yüz veri kümeleri üzerinde performans değerlendirilmiştir ve denemelerde elde edilen sonuçlara göre geliştirilen LYD yönteminin mevcut klasik yöntemlerle kıyasla daha başarılı olduğu görülmektedir.

Ayrıca, Lorentz uzayının sınıflandırma probleminde kullanımı konusunun kapsamı genişletilerek Lorentz metriği temelinde sınıflandırma yapan yeni Lorentz Destek Vektör Makinesi yöntemi geliştirilmiştir. Klasik DVM yönteminin yapısal özelliklerini Lorentz uzayına göre yapılandırarak geliştirilen yeni yöntem Lorentz metriğine göre farklı sınıflara ait örnekleri ayıran hiperdüzlemi belirlemektedir. Hiperdüzlemden en yakın noktalara kadar olan marjinin maksimize edilmesinde Lorentz uzaklığı kullanılmaktadır. Ayrıca, sınıflandırma başarısını ve çalışma zamanını azaltmak adına bir önışlem geliştirilmiştir.

Önerilen LDVM yönteminin kullanılabilirliği ve performansını test etmek amacıyla k-EYK, DVM ve Bayes gibi klasik sınıflandırıcılar ile karşılaştırılmıştır. Genel kullanıma açık olan LSVT, SONAR, TELESCOPE ve WISCONSIN veri kümeleri üzerinde yapılan denemelerde geliştirilen LDVM yöntemi mevcut klasik yöntemlere göre ortalama %12 oranında başarıyı arttırmaktadır.

Sonuç olarak, bu tez çalışmasında Lorentz uzayının sınıflandırma probleminde kullanımı ile ilgili toplamda 11 sınıflandırma algoritması, 2 önışlem tekniğı geliştirilmiş ve bu algoritmaların geçerliliğini test etmek amacıyla denemelerde 17 veri kümesi kullanılmıştır.

Sonraki çalışmalarda çok sınıflı örnekler ve çok boyutlu veri kümelerinin Lorentz uzayında doğrudan sınıflandırılmasının araştırılması hedeflenmektedir. Ayrıca, Lorentz metriğinin ayırt edici ölçüt olarak diğere problemlerde de uygulanabilirliğinin irdelenmesi ve Lorentz uzayında geliştirilen yöntemlerin çeşitli gerçek uygulamalarda test edilmesi hedeflenmektedir.

KAYNAKLAR

1. Theodoridis, S., Koutroumbas, K. (2009). *Pattern Recognition*, (4th edition). San Diego: Elsevier.
2. Bkassiny, M., Li, Y., Jayaweera, S. K. (2013). A survey on machine-learning techniques in cognitive radios. *IEEE Communications Surveys & Tutorials*, 15(3), 1136-1159.
3. Louridas P., Ebert C. (2016). Machine Learning, in *IEEE Software*, 33 (5), 110-115.
4. Smith, S. J., Bourgoin, M. O., Sims, K., Voorhees, H. L. (1994). Handwritten character classification using nearest neighbor in large databases. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9), 915-919.
5. Xiang, S., Nie, F., Zhang, C. (2008). Learning a Mahalanobis distance metric for data clustering and classification. *Pattern Recognition*, 41(12), 3600-3612.
6. Siegler, M. A., Jain, U., Raj, B., Stern, R. M. (1997). *Automatic segmentation, classification and clustering of broadcast news audio*. DARPA Speech Recognition Workshop, 97-99.
7. Wang, Z., Hu, Y., Chia, L. T. (2010). Image-to-class distance metric learning for image classification. *Computer Vision–ECCV*, 706-719
8. Beem, J. K., Ehrlich, P., Easley, K. (1996). *Global Lorentz Geometry* (Second Edition). New York: CRC Press, 62.
9. Drumm, T. A. (2012). *Geometry, Topology and Dynamics of Character Varieties: Lorentz Geometry*. Singapore: World Scientific, 23, 247-281.
10. Sternberg, S. (2003). Semi-Riemann Geometry and General Relativity, URL: <http://www.webcitation.org/query?id=1499054522632043>, Son Erişim Tarihi: 29.06.2017.
11. Müller, O., Sanchez, M. (2013). An Invitation to Lorentz Geometry, *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 1-31.
12. O'Neill, B. (1983). *Semi-Riemannian geometry with applications to relativity*. New York: Academic Press, 459.
13. Chen, B. Y. (2011). *Pseudo-Riemannian geometry, [delta]-invariants and applications*. Singapore: World Scientific, 473.
14. Gromoll, D., Klingenberg, W., Meyer, W. (1968). *Riemannsche Geometrie Im Groben*. Berlin: Springer-Verlag, 338.
15. Liu, R., Su, Z., Lin, Z., Hou, X., (2009). *Lorentz Discriminant Projection and Its Applications*. Paper presented at the 9th Asian Conference on Computer Vision (ACCV'09), 3, 311-320.

16. Liu, R., Lin, Z., Su, Z., Tang, K. (2010). Feature extraction by learning Lorentzian metric tensor and its extensions. *Pattern Recognition*, 43(10), 3298-3306.
17. Tang, K., Yu, C., Liu, R., Su, Z., (2011). A Semi-Supervised Dimensionality Reduction Framework for Face Recognition Based on Sparse Lorentz Metric Tensors. *Journal of Information & Computational Science*, 8(4), 601-608.
18. Wang, L., Liu, R., Su, Z. (2012). Robust Visual Tracking via Incremental Lorentz Discriminant Projection, *Journal of Information & Computational Science*, 9(7), 1749-1759.
19. Lin, T., Zha, H. (2008). Riemannian manifold learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5), 796-809.
20. Zhao, D., Lin, Z., Tang, X. (2008). *Classification via semi-Riemannian spaces*. IEEE Conference on Computer Vision and Pattern Recognition, (CVPR 2008). 1-8.
21. Zhang, W., Lin, Z., Tang, X. (2011). Learning semi-Riemannian metrics for semisupervised feature extraction. *IEEE Transactions on Knowledge and Data Engineering*, 23(4), 600-611.
22. Li, Y., Wong, K. M., De Bruin, H. (2012). Electroencephalogram signals classification for sleepstate decision-a Riemannian geometry approach. *IET Signal Processing*, 6(4), 288-299.
23. Barachant, A., Bonnet, S., Congedo, M., Jutten, C. (2012). Multiclass brain-computer interface classification by Riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4), 920-928.
24. Deza, M. M., Deza, E. (2006). *Dictionary of distances*. Amsterdam: Elsevier, 378.
25. Gündoğan, H., Keçilioğlu, O. (2006) Lorentz Matrix Multiplication And The Motions On Lorentz Plane, *Glasnik Matematicki*, 41, 329-334.
26. Marcus, M., Minc, H. (1992). *A survey of matrix theory and matrix inequalities*. New York: Dover, 208.
27. Largeteau-Skapin, G., Andres, E. (2009). Discrete-euclidean operations. *Discrete Applied Mathematics*, 157(3), 510-523.
28. İnternet: Lichman, M. (2013). UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science. URL: <http://www.webcitation.org/query?id=1499054913129384>, Son Erişim Tarihi: 29.06.2017.
29. Lauer, F., Bloch, G. (2008). Incorporating prior knowledge in support vector machines for classification: A review. *Neurocomputing*, 71(7), 1578-1594.
30. Mountrakis, G., Im, J., Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247-259.

31. Barakat, N., Bradley, A. P. (2010). Rule extraction from support vector machines: a review. *Neurocomputing*, 74(1), 178-190.
32. Qi, Z., Tian, Y., Shi, Y. (2013). Robust twin support vector machine for pattern classification. *Pattern Recognition*, 46(1), 305-316.
33. Houari, R., Bounceur, A., Kechadi, M. T., Tari, A. K., Euler, R. (2016). Dimensionality reduction in data mining: A Copula approach. *Expert Systems with Applications*, 64, 247-260.
34. Sun, X., Liu, Y., Xu, M., Chen, H., Han, J., Wang, K. (2013). Feature selection using dynamic weights for classification. *Knowledge-Based Systems*, 37, 541-549.
35. Spolaôr, N., Monard, M. C., Tsoumakas, G., Lee, H. D. (2016). A systematic review of multi-label feature selection and a new method based on label construction. *Neurocomputing*, 180, 3-15.
36. Brualdi, R., (2010). *Introductory Combinatorics* (5th edition), New Jersey: Pearson Prentice Hall, 598.
37. Abate, A. F., Nappi, M., Riccio, D., Sabatino, G. (2007). 2D and 3D face recognition: A survey. *Pattern recognition letters*, 28(14), 1885-1906.
38. Alizadeh, M., Abolfazli, S., Zamani, M., Baharun, S., Sakurai, K. (2016). Authentication in mobile cloud computing: A survey. *Journal of Network and Computer Applications*, 61, 59-80.
39. Tan, X., Chen, S., Zhou, Z. H., Zhang, F. (2006). Face recognition from a single image per person: A survey. *Pattern Recognition*, 39(9), 1725-1745.
40. Jafri, R., & Arabnia, H. R. (2009). A survey of face recognition techniques. *Jips*, 5(2), 41-68.
41. Platt, J. C. (1999). 12 fast training of support vector machines using sequential minimal optimization. *Advances in Kernel Methods*, 185-208.
42. Avidan, S. (2004). Support vector tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8), 1064-1072.
43. Hu, M., Chen, Y., Kwok, J. T. Y. (2009). Building sparse multiple-kernel SVM classifiers. *IEEE Transactions on Neural Networks*, 20(5), 827-839.
44. Pal, M., & Foody, G. M. (2010). Feature selection for classification of hyperspectral data by SVM. *IEEE Transactions on Geoscience and Remote Sensing*, 48(5), 2297-2307.
45. Kerimbekov, Y., Bilge, H. Ş., Uğurlu, H. H. (2016). The use of Lorentzian distance metric in classification problems. *Pattern Recognition Letters*, 84, 170-176.
46. Bilge, H. S., Kerimbekov, Y., Uğurlu, H. H. (2013). *Dimensionality reduction based on Lorentzian Manifold for face recognition*. Electronics, Computer and Computation (ICECCO), 2013 International Conference on, 212-215.

47. Bilge, H. Ş., Kerimbekov, Y. (2015). *Classification with Lorentzian distance metric*. 23th Signal Processing and Communications Applications Conference (SIU), 2106-2109.
48. Bilge, H. Ş., Kerimbekov, Y., Uğurlu, H. H. (2015). *A new classification method by using Lorentzian distance metric*. Innovations in Intelligent SysTems and Applications (INISTA), 2015 International Symposium on, 1-6.
49. Kerimbekov, Y., Bilge, H. Ş. (2016). *A feature selection method based on Lorentzian metric*. Big Data, 2016 IEEE International Conference on, 1760-1767.
50. Kerimbekov, Y., Bilge, H. Ş. (2017). *Lorentzian Distance Classifier for Multiple Features*, 6th International Conference on Pattern Recognition Applications and Methods, (ICPRAM 2017).
51. Kerimbekov, Y., Bilge, H. Ş. (2017). *Improved Feature Selection in Lorentzian Space*. 9th International Conference on Machine Learning and Computing (ICMLC 2017).
52. Kerimbekov, Y., Bilge, H. Ş., Uğurlu, H.H. (2017). *DVM ile Lorentz Uzayında Sınıflandırma*. IEEE 25. Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SİU2017).
53. Kerimbekov, Y., Bilge, H. Ş., (2017). *Lorentz Metriği ile Yüz Doğrulama*. IEEE 25. Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SİU2017).

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : KERİMBEKOV, Yerzhan
 Uyuğu : Kazakistan
 Doğum tarihi ve yeri : 07.05.1986, Ankara
 Medeni hali : Evli
 e-mail : kerimbekov@ogr.yesevi.edu.tr



Eğitim

Derece	Eğitim Birimi	Mezuniyet tarihi
Doktora	Gazi üniversitesi/ Bilgisayar Mühendisliği	devam ediyor
Yüksek lisans	Ahmet Yesevi Üniversitesi /Matematik	2010
Lisans	Ahmet Yesevi Üniversitesi /Matematik	2008
Lise	1. Mayıs Lisesi	2003

İş Deneyimi

Yıl	Yer	Görev
2010-Halen	Ahmet Yesevi Üniversitesi	Araştırma Görevlisi

Yabancı Dil

İngilizce, Rusça

Yayınlar

SCI, SCI-EXP., SSCI Tarafından Taranan Makaleler:

1. Kerimbekov, Y., Bilge, H. S., Uğurlu H. H. (2016). The use of Lorentzian distance metric in classification problems, *Pattern Recognition Letters*, 84 (2016), 170-176.

Bilimsel Toplantı- Kongre ve Görevler

1. Bilge, H.Ş., Kerimbekov, Y., Uğurlu, H.H. (2013). Dimensionality reduction based on Lorentz Manifold for face recognition, *2013 International Conference on Electronics, Computer and Computation, (ICECCO)*, 212 – 215.

2. Bilge, H.Ş. Kerimbekov, Y. (2015) Lorentz Uzaklık Ölçütü ile Sınıflandırma”, *IEEE 23. Sinyal İşleme ve İletişim Uygulamaları Kurultayı*, (SİU2015), 2003-2008.
3. Bilge, H.Ş., Kerimbekov, Y., Uğurlu, H.H. (2015). A new classification method by using Lorentzian distance metric, *2015 International Symposium on Innovations in Intelligent SysTems and Applications* (INISTA 2015-Madrid), 1-6.
4. Kerimbekov, Y., Bilge, H.Ş. (2016). A feature selection method based on Lorentzian metric, *2016 IEEE International Conference on Big Data*, (BigData-Washington D.C.), 1760-1767.
5. Kerimbekov, Y., Bilge, H.Ş. (2017). Lorentzian Distance Classifier for Multiple Features, *6th International Conference on Pattern Recognition Applications and Methods*, (ICPRAM 2017), 24-26 February 2017, Portugal.
6. Kerimbekov, Y., Bilge, H.Ş. (2017). Improved Feature Selection in Lorentzian Space, *9th International Conference on Machine Learning and Computing*, (ICMLC 2017), 24-26 February 2017, Singapore.
7. Kerimbekov, Y., Bilge, H.Ş., Uğurlu, H.H. (2017). DVM ile Lorentz Uzayında Sınıflandırma, *IEEE 25. Sinyal İşleme ve İletişim Uygulamaları Kurultayı* (SİU2017), 15-18 Mayıs 2017, Antalya.
8. Kerimbekov, Y., Bilge, H.Ş. (2017). Lorentz Metriği ile Yüz Doğrulama, *IEEE 25. Sinyal İşleme ve İletişim Uygulamaları Kurultayı* (SİU2017, 15-18 Mayıs 2017, Antalya.

Yer Alınan Bilimsel Projeler

1. “Lorentz Uzayına Dayalı Yeni Bir Uzaklık Ölçütü Kullanarak Sınıflandırma Başarısının Artırılması”, 115E181 nolu TUBİTAK 3001 projesi (01.06.2015-2017), Araştırmacı.

Hobiler

Yüzme, Futbol

DİZİN

A

ayırt etme ölçütü · 21, 33, 78, 99

B

benzerlik · iv, 8, 48, 52, 109

D

diklik · xvi, 54, 55

diyagonal matris · 36

döndürme · iv, vii, xv, 21, 22, 23,
28, 29, 30, 40, 47, 62, 63, 68,
69, 73, 83, 84, 86, 88, 89, 91,
107, 108

E

en iyi parametre · 36, 37, 38, 72,
108

en yakın komşu · xviii, 14, 34,
35, 45, 48, 58, 63, 88

H

hiperbolik · 23

hiperdüzlem · xv, 24, 49, 51, 54,
99

K

karar çizgisi · xv, 24, 25, 26, 27,
28, 30, 63, 67

kernel fonksiyon · 99

komşuluk kavramı · iv, 99

kovarians matrisi · 43

L

LDVM · x, xvi, xvii, xviii, 21,
56, 57, 99, 100, 101, 102, 103,
104, 105, 106, 110

Lorentz manifoldu · 12

Lorentz metriği · iv, viii, 11, 12,
21, 24, 34, 36, 39, 42, 43, 48,
55, 56, 57, 58, 62, 75, 80, 96,
101, 107, 108, 109

Lorentz Sınıflandırıcısı · iv, vii,
viii, xviii, 21, 33, 34, 35, 36,
38, 39, 41, 44, 45, 57, 62, 69,
71, 77, 107, 108

Lorentz uzaklığı · iv, vii, 6, 13,
15, 31, 33, 34, 57, 107, 108,
109

Lorentz uzayı · iv, viii, xv, 5, 6,
8, 11, 12, 33, 36, 41, 42, 57,
69

LYD · iv, x, xviii, 17, 48, 57, 93,
94, 95, 96, 97, 98, 99, 109

M

marjin · 49, 50

metrik tensörü · 7, 8, 11

N

null eksenı · xv, 15, 28, 30

null-like · 12

Ö

önışlem · iv, 21, 24, 25, 32, 36,
39, 40, 47, 62, 73, 79, 86, 94,
100, 108, 109, 110

öteleme · iv, 21, 27, 35, 107

P

pozitif tanımlı · 9, 11

projeksiyon · 5, 22, 32, 36

S

sıkıştırma · iv, vii, x, 21, 22, 35,
36, 37, 38, 40, 46, 69, 70, 71,
72, 73, 79, 80, 81, 83, 85, 88,
89, 107, 108

sınıf etiketi · xviii, 7, 45

space-like · 12

T

time-like · 12

U

uzaklık ölçütü · iv, 5, 6, 15, 33,
35, 45, 48, 57, 60, 69, 109

Y

yüz doğrulama · iv, xvi, 41, 47,
48, 93, 94, 95, 96, 97, 98, 99,
109



GAZİ GELECEKTİR..