



**T.C.
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**



DOKTORA TEZİ

**HİPERSPEKTRAL İMGELERİN İZGE-UZAMSAL
SINIFLANDIRILMASI İÇİN OLASILIKSAL MODELLER**

Sezer KUTLUK

**DANIŞMAN
Prof. Dr. Aydın AKAN**

**II. DANIŞMAN
Doç. Dr. Koray KAYABOL**

Elektrik-Elektronik Mühendisliği Anabilim Dalı

Elektrik-Elektronik Mühendisliği Programı

İSTANBUL-2021

Bu çalışma 23.09.2021 tarihinde ařağıdaki jüri tarafından Elektrik-Elektronik Mühendisliğı Anabilim Dalı, Elektrik-Elektronik Mühendisliğı Programı Doktora Tezi olarak kabul edilmiştir.

TEZ JÜRİSİ

Prof. Dr. Aydın AKAN
İzmir Ekonomi Üniversitesi
Mühendislik Fakültesi

Prof. Dr. Ahmet SERTBAŞ
İstanbul Üniversitesi-Cerrahpařa
Mühendislik Fakültesi

Prof. Dr. Mustafa Ersel KAMAŞAK
İstanbul Teknik Üniversitesi
Bilgisayar ve Biliřim Fakültesi

Prof. Dr. Fırat KAÇAR
İstanbul Üniversitesi-Cerrahpařa
Mühendislik Fakültesi

Doç. Dr. Yalçın ÇEKİÇ
Bahçeşehir Üniversitesi
Mühendislik ve Doğı Bilimleri Fakültesi



20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince; Bu Lisansüstü teze, İstanbul Üniversitesi-Cerrahpaşa’nın abonesi olduğu intihal yazılım programı kullanılarak Lisansüstü Eğitim Enstitüsü’nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.

Bu tez, 114E535 numaralı TÜBİTAK projesi ile desteklenmiştir.

ÖNSÖZ

Lisans öğrenimimden başlayarak bana sinyal işleme ve olasılık teorisini sevdirdikleri ve doktora öğrenimim boyunca her zaman olumlu yaklaşımlarıyla destek oldukları için danışman hocalarım Prof. Dr. Aydın Akan ve Doç. Dr. Koray Kayabol'a teşekkür ederim.

Tez İzleme Komitesi'ndeki değerli hocalarım Prof. Dr. Ahmet Sertbaş ve Prof. Dr. Mustafa Ersel Kamaşak'a olumlu katkıları ve yönlendirmeleri için teşekkür ederim. Tez savunma jürisindeki diğer hocalarım Prof. Dr. Fırat Kaçar ve Doç. Dr. Yalçın Çekiç'e değerli katkıları için teşekkür ederim.

Tez süresince bana destek olan arkadaşlarım Dr. Soner Kotan'a, Serkan Tüzün'e ve Huawei'de geçirdiğim süre boyunca birlikte çalışma şansı yakaladığım arkadaşlarıma teşekkür ederim.

Değerli arkadaşım Onur Ural'a doktora boyunca destek olduğu için teşekkür ederim.

Hayatım boyunca bana olan güvenini ve desteğini her an yanımda hissettiğim aileme teşekkür ederim.

Eylül 2021

Sezer KUTLUK

İÇİNDEKİLER

Sayfa No

ÖNSÖZ.....	iv
İÇİNDEKİLER	v
ŞEKİL LİSTESİ.....	vii
TABLO LİSTESİ.....	x
SİMGE VE KISALTMA LİSTESİ	xi
ÖZET.....	xii
SUMMARY	xiv
1. GİRİŞ.....	1
1.1. TEZİN KATKILARI	7
2. GENEL KISIMLAR	12
2.1. YAPAY ÖĞRENME	12
2.2. OLASILIKSAL MODELLEME VE SINIFLANDIRMA	13
2.2.1. Bayesçi Yaklaşım	13
2.2.2. Parametre Kestirimi.....	14
2.2.3. Sınıflandırma.....	15
2.2.4. Üretici ve Ayrımsayıcı Modeller.....	15
2.2.5. Önsel Dağılımlar ve Düzenleştirme	15
2.3. HİPERSPEKTRAL İMGELERİN SINIFLANDIRILMASI	16
2.3.1. Olasılıksal Yaklaşım ile Problem Tanımı	16
2.3.2. İzge-Uzamsal Sınıflandırma	16
3. MALZEME VE YÖNTEM.....	18
3.1. AZ SAYIDA ÖRNEKLE EĞİTİM VE İZGE-UZAMSAL SINIFLANDIRMA	19
3.1.1. Uzamsal Olarak Değişen Student-t Karışım Modeli	25
3.1.2. Uzamsal Olarak Değişen Gauss Karışım Modeli.....	29
3.2. ÜRETİCİ ALT-UZAY YÖNTEMİYLE İZGE-UZAMSAL SINIFLANDIRMA.....	32
3.3. ÇOKTERİMLİ LOJİSTİK BAĞLANIM MODELLERİYLE İZGE-UZAMSAL SINIFLANDIRMA	35
3.3.1. Üretici Model	36
3.3.2. Ayrımsayıcı Model.....	39
3.4. EVRİŞİMSSEL SINIR AĞI İLE İZGE-UZAMSAL SINIFLANDIRMA	43

3.4.1. Ağ Mimarisi	47
3.4.2. Sınıflandırma.....	48
3.5. HİPERSPEKTRAL İMGE VERİ KÜMELERİ	49
3.5.1. Indian Pines İmgesi	49
3.5.2. Pavia Centre İmgesi.....	50
3.5.3. Pavia University İmgesi.....	51
3.5.4. Salinas İmgesi	52
3.5.5. Kennedy Space Center İmgesi	54
4. BULGULAR	55
4.1. AZ SAYIDA ÖRNEKLE EĞİTİM YAPILAN MODELLERLE ELDE EDİLEN SONUÇLAR.....	55
4.1.1. Uzamsal Olarak Değişen Student-t Karışım Modeli ile Elde Edilen Sonuçlar	55
4.1.2. Uzamsal Olarak Değişen Gauss Karışım Modeli ile Elde Edilen Sonuçlar	59
4.2. ÜRETİCİ ALT-UZAY MODELİ İLE ELDE EDİLEN SONUÇLAR.....	65
4.3. ÇOKTERİMLİ LOJİSTİK BAĞLANIM MODELLERİ İLE ELDE EDİLEN SONUÇLAR.....	68
4.3.1. Üretici Model ile Elde Edilen Sonuçlar.....	68
4.3.2. Ayrımsayıcı Model ile Elde Edilen Sonuçlar	71
4.4. EVRİŞİMSSEL SİNİR AĞI MODELİ İLE ELDE EDİLEN SONUÇLAR.....	75
5. TARTIŞMA VE SONUÇ	83
KAYNAKLAR	86
EKLER.....	94
EK-1. ÇOKTERİMLİ STUDENT-T DAĞILIMININ PARAMETRE KESTİRİMİ.....	94
EK-2. SINIF-KOŞULLU DAĞILIMLARIN NORMALİZE EDİLMİŞ ÜSTEL DAĞILIMLAR ŞEKLİNDE TANIMLANMASI İLE ÇOKTERİMLİ LOJİSTİK BAĞLANIM SINIFLANDIRICISININ DENKLİĞİ	97
EK-3. SÖZLÜK.....	99
ÖZGEÇMİŞ	100

ŞEKİL LİSTESİ

Sayfa No

Şekil 3.1: Farklı özgürlük derecesi değerleri için Student-t dağılımı ile normal dağılım olasılık yoğunluk fonksiyonları. Normal dağılımın ortalaması 0, standart sapması 1, Student-t dağılımlarının ise lokasyon parametreleri 0 ve ölçek parametreleri 1 olup özgürlük dereceleri 1, 3 ve 10 olarak seçilmiştir.	26
Şekil 3.2: Sınıf-koşullu olasılıkların Student-t dağılımı seçilmesi durumunda izge-uzamsal üretici model.....	27
Şekil 3.3: Sınıf-koşullu olasılıkların Gauss dağılımı seçilmesi durumunda izge-uzamsal üretici model.	30
Şekil 3.4: Normal dağılım ve Laplace dağılımının olasılık yoğunluk fonksiyonları.	41
Şekil 3.5: Spektral bant sayısı 200 olan bir imge için tasarlanan ağdaki evrişim katmanlarının yapısı.	48
Şekil 3.6: Spektral bant sayısı 103 olan bir imge için tasarlanan ağdaki evrişim katmanlarının yapısı.	48
Şekil 3.7: Indian Pines imgesinin gerçek sınıflandırma haritası.	50
Şekil 3.8: Pavia Centre imgesinin gerçek sınıflandırma haritası.....	51
Şekil 3.9: Pavia University imgesinin gerçek sınıflandırma haritası.....	52
Şekil 3.10: Salinas imgesinin gerçek sınıflandırma haritası.	53
Şekil 3.11: Kennedy Space Center imgesinin gerçek sınıflandırma haritası.	54
Şekil 4.1: Indian Pines imgesi üzerinde a) BGMM b) BStMM c) p-LDA d) d-QDA ile elde edilen örnek sınıflandırma haritaları.....	56
Şekil 4.2: Pavia Centre imgesi üzerinde a) BGMM b) BStMM c) p-LDA ve d-QDA ile elde edilen örnek sınıflandırma haritaları.....	57
Şekil 4.3: Pavia University imgesi üzerinde a) BGMM b) BStMM c) p-LDA d) d-QDA ile elde edilen örnek sınıflandırma haritaları.	58
Şekil 4.4: Salinas imgesi üzerinde a) BGMM b) BStMM c) p-LDA d) d-QDA ile elde edilen örnek sınıflandırma haritaları.	59
Şekil 4.5: Indian Pines imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kayabol ve Kutluk, 2016). Parantez içindeki değerler bu haritalar için hesaplanmış yüzde doğruluk değerleridir.....	61

Şekil 4.6: Pavia Centre imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kayabol ve Kutluk, 2016). Parantez içindeki değerler bu haritalar için hesaplanmış yüzde doğruluk değerleridir.....	62
Şekil 4.7: Pavia University imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kayabol ve Kutluk, 2016). Parantez içindeki değerler bu haritalar için hesaplanmış yüzde doğruluk değerleridir.....	63
Şekil 4.8: Salinas imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kayabol ve Kutluk, 2016). Parantez içindeki değerler bu haritalar için hesaplanmış yüzde doğruluk değerleridir.....	64
Şekil 4.9: Eğitim kümesi boyutuna göre sınıflandırma doğruluğunun değişimi (Kayabol ve Kutluk, 2016).	65
Şekil 4.10: Indian Pines imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kutluk ve diğ., 2016a).....	66
Şekil 4.11: Pavia Centre imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kutluk ve diğ., 2016a).....	67
Şekil 4.12: Pavia University imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kutluk ve diğ., 2016a).....	67
Şekil 4.13: Salinas imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kutluk ve diğ., 2016a).	67
Şekil 4.14: Indian Pines imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kutluk ve diğ., 2018). Parantez içinde verilen değerler bu haritalarda hesaplanan yüzde doğruluk değerleridir.....	69
Şekil 4.15: Pavia University imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kutluk ve diğ., 2018). Parantez içinde verilen değerler bu haritalarda hesaplanan yüzde doğruluk değerleridir.....	70
Şekil 4.16: Ayrımsayıcı model ile Indian Pines imgesinde elde edilen örnek sınıflandırma haritaları.	72
Şekil 4.17: Ayrımsayıcı model ile Pavia University imgesinde elde edilen örnek sınıflandırma haritaları.	73
Şekil 4.18: Ayrımsayıcı model ile Salinas imgesinde elde edilen örnek sınıflandırma haritaları.....	74
Şekil 4.19: Indian Pines imgesi ile elde edilen örnek sınıflandırma haritaları. Parantez içindeki değerler sınıflandırma doğruluğunu göstermektedir.	76
Şekil 4.20: Pavia University imgesi ile elde edilen örnek sınıflandırma haritaları. Parantez içindeki değerler sınıflandırma doğruluğunu göstermektedir.	77
Şekil 4.21: Devir sayısı ile ortalama sınıflandırma doğruluğunun değişimi.....	79

Şekil 4.22: Eğitim süresi ile sınıflandırma doğruluğunun değişimi. 80



TABLO LİSTESİ

Sayfa No

Tablo 3.1: Indian Pines imgesindeki sınıflar ve piksel sayıları.	50
Tablo 3.2: Pavia Centre imgesindeki sınıflar ve piksel sayıları.....	51
Tablo 3.3: Pavia University imgesindeki sınıflar ve piksel sayıları.....	52
Tablo 3.4: Salinas imgesindeki sınıflar ve piksel sayıları.	53
Tablo 3.5: Kennedy Space Center imgesindeki sınıflar ve piksel sayıları.	54
Tablo 4.1: Student-t karışım modeliyle elde edilen ortalama yüzde doğruluk değerleri.	55
Tablo 4.2: $N_e = 15$ için elde edilen ortalama yüzde doğruluk ve standart sapma değerleri.....	60
Tablo 4.3: Olasılıksal temel bileşen analizi karışım modeli ile elde edilen ortalama yüzde doğruluk değerleri.....	66
Tablo 4.4: Yaklaşık çokterimli lojistik bağlanım modeli ile elde edilen ortalama yüzde doğruluk değerleri.	69
Tablo 4.5: Ayrırımsayıcı model ile elde edilen sınıflandırma sonuçları (ortalama yüzde doğruluk $\pm$ standart sapma).	71
Tablo 4.6: Evrişimsel sinir ağı ile elde edilen sonuçlar (% ortalama doğruluk $\pm$ standart sapma).	76
Tablo 4.7: Indian Pines, Pavia University, Salinas ve Kennedy Space Center imgeleri için kullanılan ağılardaki öznetelik çıkarma katmanlarının yapısı.	79
Tablo 4.8: Evrişimsel sinir ağı ile 100. devirde elde edilen ortalama sınıflandırma doğruluğu ve standart sapma değerleri.....	81
Tablo 4.9: Evrişimsel sinir ağı modeli için istatistiksel analiz sonuçları.	81

SİMGE VE KISALTMA LİSTESİ

Simgeler

Açıklama

$\mathbf{s}_n$	: n . piksele ait spektral vektör
$\mathbf{z}_n$	: n . piksele ait etiket vektörü
$z_{n,k}$	: n . piksele ait etiket vektörünün k . elemanı
L	: İmgedeki spektral bant sayısı
N	: İmgedeki etiketli piksel sayısı
N_k	: k . sınıftaki piksel sayısı

Kısaltmalar

Açıklama

ASMLR	: Approximate sparse multinomial logistic regressin – Yaklaşık çokterimli seyrek lojistik bağlanım
CNN	: Convolutional neural network – Evrişimsel sinir ağı
EM	: Expectation maximization – Beklenti en büyükleme
ICM	: Iterated conditional mode – Yinelemeli koşullu mod
LDA	: Linear discriminant analysis – Doğrusal ayırtaç analizi
ML	: Maximum likelihood – En büyük olabilirlik
MAP	: Maximum a posteriori – En büyük sonsal
PCA	: Principal component analysis – Temel bileşen analizi
PPCA	: Probabilistic principal component analysis – Olasılıksal temel bileşen analizi
QDA	: Quadratic discriminant analysis – Karesel ayırtaç analizi
ReLU	: Rectified linear unit – Doğrultulmuş doğrusal birim
SV-GMM	: Spatially-varying Gaussian mixture model – Uzamsal olarak değişen Gauss karışım modeli
SVM	: Support vector machine – Destek vektör makinesi

ÖZET

DOKTORA TEZİ

HİPERSPEKTRAL İMGELERİN İZGE-UZAMSAL SINIFLANDIRILMASI İÇİN OLASILIKSAL MODELLER

Sezer KUTLUK

İstanbul Üniversitesi-Cerrahpaşa

Lisansüstü Eğitim Enstitüsü

Elektrik-Elektronik Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Aydın AKAN

II. Danışman: Doç. Dr. Koray KAYABOL

Son yıllarda hiperspektral uzaktan algılama teknolojisinde önemli gelişmeler olmuştur. Havadan ve uzaydan algılama yapan araçlara yerleştirilen hiperspektral sensörler ile Dünya yüzeyinin yüksek spektral, uzamsal ve zamansal çözünürlüklerde görüntülenmesi mümkün olmakta ve böylece görüntülerden daha ayrıntılı bilgi çıkarımı yapılabilmektedir. Diğer yandan, hiperspektral imge verisinin büyük boyutlu olması, eğitici yapay öğrenme yöntemleri için veri etiketlemenin güçlüğü, spektral karışım ve gürültü ile atmosferik etkilerin sebep olduğu bozulmalar bu imgelerin işlenmesinde karşılaşılan zorluklardır.

Hiperspektral imgeler geniş bir spektral aralıkta, birbirine yakın spektral bantlardan oluşmaktadır. Hiperspektral imge sınıflandırma probleminde, imgedeki her pikselin hangi sınıfa ait olduğu tahmin edilmekte, dolayısıyla imge için bir bölütleme haritası elde edilmektedir. İzgesel sınıflandırmada yalnızca piksellerin spektral imzaları kullanılırken, izge-uzamsal sınıflandırmada bir pikselin hem spektral imzası hem de bu piksele komşu olan piksellerin etiketleri dikkate alınmaktadır. Atmosferik koşulların değişimi ve diğer gürültü kaynakları nedeniyle, yapay öğrenme algoritmalarının hiperspektral imge sınıflandırma probleminde uygulanmasında genellikle aynı imgeden hem eğitim hem test pikselleri seçilmektedir. Hem bu nedenle, hem spektral karışım nedeniyle, hem de spektral bant sayısının fazla olması nedeniyle, standart imge işleme yöntemlerinin olduğu gibi uygulanması çoğu

zaman mümkün olmamakta, bu veri türüne uygun yöntem ve modellerin geliştirilmesi gerekmektedir.

Hiperspektral imge sınıflandırmada kullanılan klasik yapay öğrenme yöntemlerinde boyut indirgeme, öznitelik çıkarımı ve sınıflandırma gibi işlemler ayrık bir işlem dizisi şeklinde uygulanmaktadır. Olasılıksal yaklaşımda verideki rastgeleliğin veya belirsizliğin modellenmesi mümkün olmakta, önsel dağılımlar tanımlanarak düzenlileştirme kısıtları da eklenebilmekte ve böylece parametre kestirimi kolaylaşmaktadır. Son yıllarda sıkça kullanılmakta olan evrimsel sinir ağlarının en önemli özelliği ise sınıflandırma ve öznitelik çıkarımı parametrelerinin birlikte optimize edilmesiyle daha iyi öznitelikler ve dolayısıyla daha iyi sınıflandırma başarımı elde edilmesidir.

Bu tezde hiperspektral imgelerde yeryüzü bölgelerinin sınıflandırılması amacıyla izge-uzamsal yöntemler geliştirilmiştir. Olasılıksal yaklaşım temel alınmış olup, az sayıda eğitim pikseliyle eğitim yapılabilmesi için farklı karışım modelleri, boyut indirgeme amacıyla bir üretici alt-uzay yöntemi, çokterimli lojistik bağlanım kullanılan üretici ve ayırmsayıcı modeller ile bir evrimsel sinir ağı modeli geliştirilmiş, bu modeller için parametre kestirimi ve sınıflandırma algoritmaları elde edilmiştir. Önerilen yöntemlerin başarımları literatürde kullanılan hiperspektral imge veri kümeleri üzerinde yapılan testlerle ölçülmüş, başka algoritmalarla karşılaştırılmış ve önerilen yöntemlerle iyi sonuçlar elde edildiği gözlenmiştir.

Eylül 2021, 116 sayfa.

Anahtar kelimeler: hiperspektral imge, yapay öğrenme, olasılıksal modelleme, görüntü işleme

SUMMARY

Ph.D. THESIS

PROBABILISTIC MODELS FOR SPECTRAL-SPATIAL CLASSIFICATION OF HYPERSPECTRAL IMAGES

Sezer KUTLUK

Istanbul University-Cerrahpasa

Institute of Graduate Studies

Department of Electrical and Electronic Engineering

Supervisor: Prof. Dr. Aydın AKAN

Co-Supervisor: Assoc. Prof. Dr. Koray KAYABOL

Significant developments have occurred in the hyperspectral remote sensing technology in recent years. With hyperspectral sensors deployed on airborne and spaceborne vehicles, it has been possible to gather the images of the surface of the Earth with high spectral, spatial, and temporal resolutions, and thus, more detailed information can be gathered via these images. On the other hand, the size of hyperspectral image data being large, difficulties in data annotation for supervised machine learning methods, spectral mixing, and distortion caused by noise and atmospheric effects are the difficulties encountered while processing these images.

Hyperspectral images consist of spectral bands close to each other in a wide spectrum. In the hyperspectral image classification problem, which class each pixel in the image belongs to is predicted; therefore, a segmentation map is obtained for the image. While only the spectral signatures of pixels are used in spectral classification, both spectral signature and labels of the neighbors of a pixel are considered in spectral-spatial classification. Because of the changes in atmospheric conditions and other sources of noise, both training and test pixels are generally chosen from the same image in the application of machine learning algorithms to the hyperspectral image classification problem. Both for this reason and because of spectral mixing alongside the number of spectral bands being large, it is usually not possible to apply standard

image processing methods as they are, and methods and models suitable to this type of data need to be developed.

In classical machine learning algorithms used in hyperspectral image classification, operations like dimension reduction, feature extraction and classification are applied as a series of separate processes. In the probabilistic approach, the randomness or uncertainty in the data can be modelled, regularization constraints can be integrated by defining prior distributions, and thus, parameter estimation becomes easier. The most important property of convolutional neural networks which have frequently been used in recent years is that with classification and feature extraction parameters optimized together, better features and therefore better classification performance are obtained.

In this dissertation, spectral-spatial methods have been developed for land cover classification in hyperspectral images. While probabilistic approach is taken, different mixture models for training with a small number of training pixels, a generative subspace method for dimension reduction, a generative and a discriminative model by using multinomial logistic regression, and a convolutional neural network model have been developed, and parameter estimation and classification algorithms for these models have been derived. Performance of the proposed methods were measured by tests on the hyperspectral image datasets used in the literature; they were compared with other algorithms, and it has been observed that good results were obtained with the proposed methods.

September 2021, 116 pages.

Keywords: hyperspectral image, machine learning, probabilistic modeling, image processing

1. GİRİŞ

Son yıllarda uzaktan algılamada kullanılan sensör teknolojisindeki gelişmeler, havadan ve uzaydan görüntüleme yapan araçlara yerleştirilen sensörlerle, Dünya yüzeyinin yüksek izgesel, uzamsal ve zamansal çözünürlükle görüntülenmesine olanak sağlamıştır (Bioucas-Dias ve diğ., 2013). Hiperspektral görüntüleme de bu alanda önemli bir araştırma konusu durumuna gelmiştir. Hiperspektral sensörlerle elde edilen imgelerde, her piksel yüzlerce spektral banttı oluşarı bir vektörle temsil edilmektedir. Elde edilen bu vektörlerle yer yüzü bölgelerinin bölütlenmesi ve nesnelerin tanınması gibi çeşitli uygulamalar gerçekleştirilebilmekte olup bu uygulamalar tarım arazilerinin denetlenmesi, şehir planlama, bitki örtüsünün sınıflandırılması gibi farklı alanlarda katkı sağlamaktadır.

Hiperspektral imgelerde her spektral bant elektromanyetik spektrumun farklı bir bölgesini göstermektedir. Yüksek sayıda spektral bant olması, eğitimci algoritmalar için eğitim verisi elde etmek amacıyla veri etiketleme işleminin zor olması, spektral karışım, çeşitli gürültü kaynaklarının varlığı ve atmosferik etkiler, bu imgelerin işlenmesinde karşılaşılan temel zorluklardır (Bioucas-Dias ve diğ., 2013).

Hiperspektral imge işleminin önemli problemlerinden biri yeryüzü bölgelerinin ve arazi örtüsünün sınıflandırılmasıdır. Bu problemde amaç imgedeki her pikselin hangi sınıfa ait olduğuna ilişkin bir tahmin yapılmasıdır.

Hiperspektral imge sınıflandırma yöntemleri genel olarak izgesel veya izge-uzamsal olarak ikiye ayrılmaktadır. İzgesel sınıflandırmada piksellerin yalnızca spektral imzaları kullanılarak sınıflandırıcı eğitimi ve sınıflandırma yapılmaktadır. İzge-uzamsal sınıflandırmada ise hem spektral imzalardan yararlanılmakta hem de piksel etiketleri komşu piksellerin etiketlerine bağlı olarak belirlenmektedir (Audebert ve diğ., 2019).

İstatistiksel modellerde parametre boyutu arttıkça parametre kestiriminin güvenilir şekilde yapılabilmesi için gerekli veri miktarı daha hızlı bir şekilde artmaktadır (Bellman, 1961). Yüksek boyutlu veri türlerinde bu problemle sıkça karşılaşılmakta ve gerekli eğitim verisi çok büyük boyutlara ulaşmaktadır. Yapay öğrenmede, eğitim kümesindeki örnek sayısı sabit tutulduğunda, parametre boyutu artarken model başarımı bir süre artmakta, daha sonra

azalmaktadır ve bu durum Hughes fenomeni olarak adlandırılmaktadır (Hughes, 1968). Hiperspektral imgelerde genellikle etiketli piksel sayısı az, buna karşılık spektral bant sayısı çok olduğundan bu durum ortaya çıkabilmektedir. Örneğin, Gauss karışım modeli kullanıldığında, kovaryans matrisi kestirimi spektral bant sayısının fazla, etiketli örnek sayısının ise sınırlı olması nedeniyle güçleşmektedir. Bu nedenle hiperspektral imgelerin olasılıksal modellenmesinde farklı dağılımlar ile boyut indirgeme ve düzenlileştirme yöntemleriyle bu problem giderilmeye çalışılmaktadır.

Bu tezde hiperspektral imgelerde yeryüzü bölgelerinin izge-uzamsal sınıflandırılması amacıyla çeşitli modeller ve çıkarsama yöntemleri önerilmektedir. İmgelerin olasılıksal olarak modellenmesi için çeşitli yaklaşımlar uygulanmış, geliştirilen modeller için parametre kestirimi ve sınıflandırma algoritmaları elde edilmiştir. İlk yaklaşımda az örnekle eğitim yapılabilmesi için Bayesçi yaklaşımla durağan olmayan Student-t ve Gauss karışım modelleri geliştirilmiştir. Bu modellerde piksel etiketleri üzerinde önsel dağılımlar tanımlanarak izge-uzamsal sınıflandırma yapılmaktadır. İkinci yaklaşımda bir üretici alt-uzay yöntemi ile daha düşük boyutta bir durağan olmayan Gauss karışım modeli elde edilmekte, piksel etiket önselleriyle birlikte izge-uzamsal sınıflandırma yapılmaktadır. Boyut indirgeme ile hem parametre kestirimi kolaylaştırılmakta, hem işlemsel karmaşıklık azaltılmakta, hem de daha yüksek başarımlar elde edilebilmektedir. Üçüncü yaklaşımda, çokterimli lojistik bağlanım modeli ile bir üretici bir de ayımsayıcı model geliştirilmiştir. Üretici modelde sınıf-koşullu olasılıklar çokterimli lojistik bağlanım fonksiyonlarıyla modellenmekte, sınıf etiketleri üzerinde tanımlanan önsel dağılımlarla da bir durağan olmayan karışım modeli elde edilerek izge-uzamsal sınıflandırma yapılmaktadır. Ayımsayıcı modelde ise sınıf etiketlerinin koşullu olasılığı doğrudan modellenerek, izge-uzamsal sınıflandırma yapılan bir koşullu rastgele alan modeli elde edilmektedir. Son olarak, evrişimsel sinir ağlarının eğitimi için bir algoritma önerilmekte, bu algoritmayla farklı yapıya sahip evrişimsel sinir ağları ve uzamsal model birlikte kullanılarak izge-uzamsal sınıflandırma yapılmaktadır.

Hiperspektral imge sınıflandırma amacıyla, Li ve diğ. (2014) tarafından önerilen yöntemde Gauss karışım modeli kullanılmış, az örnekle kovaryans kestiriminde karşılaşılan problemin önlenmesi amacıyla bir boyut indirgeme yöntemi uygulanmıştır. Bandos ve diğ. (2009) doğrusal ayırtaç analizinde kovaryans kestirimi sorununu gidermek için düzenlileştirilmiş LDA algoritmasını kullanmıştır. Berge ve Solberg (2006) tarafından önerilen yöntemde kovaryans

matrisinin yapısı için bir kısıtlama uygulanarak kestirilmesi gereken parametre sayısı sınırlandırılmıştır. Tadjudin ve Landgrebe (1999) az sayıda veri ile kovaryans matrisi kestirimi için bLOOC adında bir yöntem önermiştir. Kayabol (2015) ortalama ve kovaryans matrisi için normal ters-Wishart önseli kullanarak hiperspektral imge sınıflandırma için bir Bayeşçi Gauss karışım modeli önermiştir.

Veride aykırı örneklerin bulunması eğitim ve sınıflandırmayı zorlaştırmaktadır. Bu sorunun çözümü için farklı olasılık dağılımları kullanılmaktadır. Sfikas ve diğ. (2007) Student-t karışım modeli ile bir gürbüz imge bölütleme yöntemi önermiştir. Farrell ve Mersereau (2005) hiperspektral imgelerin öbeklenmesi için bir model ve EM algoritmasıyla bir parametre kestirim yöntemi önermiştir. Bu çalışmada eliptik olarak sınırlandırılmış Student-t dağılımı kullanılmıştır. Veracini ve diğ. (2010) tarafından önerilen yöntemde Bayeşçi yaklaşımla Student-t karışım modeli kullanılmış ve parametreler üzerinde önsel dağılımlar tanımlanmıştır.

Yapay öğrenme problemlerinin çözümünde boyut indirgeme önemli bir yer tutmaktadır. Hiperspektral imgelerde de spektral bant sayısının yüksek olması nedeniyle boyut indirgemeye sıkça ihtiyaç duyulmaktadır. En sık kullanılan boyut indirgeme yöntemlerinden biri temel bileşen analizi (PCA) olup hiperspektral imgelerde de kullanılmıştır. Rodarmel ve Shan (2002) tarafından önerilen yöntemde spektral vektörlere PCA uygulanmış ve yalnızca önemli bantlar seçilerek sınıflandırma yapılmıştır. Agarwal ve diğ. (2007) tarafından önerilen yöntemde hiperspektral imge izgesel ve uzamsal parçalara ayrılarak PCA uygulanmış ve sınıflandırma yapılmıştır. Li ve diğ. (2014) bir boyut indirgeme yöntemi uygulayarak çıkarılan öznitelik vektörlerini Gauss karışım modeli kullanarak modellemiştir.

Tipping ve Bishop (1999) temel bileşen analizini olasılıksal yaklaşımla, bir gizli değişken modeli olarak formüle etmiştir. Xia ve diğ. (2014) bu yöntemi eğiticili ve yarı-eğiticili olarak hiperspektral imgelerin sınıflandırılması için kullanmıştır, ancak bu yöntemde uzamsal bilgi kullanılmamıştır.

Tezde izge-uzamsal sınıflandırma için geliştirilen durağan olmayan karışım modellerinde piksel etiketleri üzerinde önsel dağılımlar tanımlanmakta ve bu önsel dağılımlarla karışım oranları piksel konumuna göre komşu piksel etiketlerine bağlı olarak değişmektedir. Durağan olmayan karışım modelleri imge bölütleme problemi için birçok çalışmada kullanılmıştır. Örneğin Sanjay-Gopal ve Hebert (1998), Zhang ve diğ. (2001), Blekas ve diğ. (2005) ile

Diplaros ve diğ. (2007) tarafından önerilen yöntemlerde Markov rastgele alan (MRF) önsel dağılımları kullanılarak uzamsal olarak değişen karışımlar elde edilmiştir. Fernandez ve Green (2002) Poisson dağılımına sahip ve uzamsal bağlantılı veri için bir karışım modeli önermiştir. Figueiredo (2005) Bayeşçi yaklaşımla Gauss rastgele alanı kullanarak bir imge bölütleme yöntemi önermiştir.

Kayabol ve diğ. (2011) ile Kayabol ve Zerubia (2013) tarafından sentetik açıklıklı radar imgelerinin sınıflandırılması için pikseller arası uzamsal bağlantıların kullanılması amacıyla çokterimli özlojistik bağlanım modeli geliştirilmiştir. Bu uzamsal model Besag (1974) tarafından önerilen özlojistik modelinin genelleştirilmiş hâli olup hiperspektral imgelerin sınıflandırılması için Kayabol (2015), Kayabol ve Kutluk (2016), Kutluk ve diğ. (2016a; 2016b; 2018) tarafından önerilen karışım modellerinde, ayrıca Kutluk ve diğ. (2019; 2021) tarafından önerilen evrişimsel sinir ağları ile izge-uzamsal sınıflandırma yöntemlerinde kullanılmıştır.

İmge bölütleme probleminde uzamsal bilginin kullanıldığı bir diğer model Potts modeli olarak adlandırılan bir Markov rastgele alan modelidir (Geman ve Geman, 1984). Potts modeli hiperspektral imge analizi için Bali ve Mohammad-Djafari (2008) ile Eches ve diğ. (2011) tarafından kullanılmıştır. Tarabalka ve diğ. (2010) hiperspektral imgelerin sınıflandırılması için destek vektör makinesi ile birlikte piksel etiketleri üzerinde tanımlanan Markov rastgele alan önseli kullanmıştır. Trianni ve Gamba (2005) ile Plaza ve diğ. (2009) tarafından önerilen yöntemlerde hiperspektral imge sınıflandırma için Gauss karışım modeli ve Markov rastgele alan önselleri kullanılmıştır.

Yapay öğrenmedeki sıkça kullanılan sınıflandırma algoritmalarından biri lojistik bağlanım algoritmasıdır. Bu bir ayırmsayıcı model olup, sınıf sayısının ikiden fazla olması durumunda softmax veya çokterimli lojistik bağlanım modeli olarak adlandırılmaktadır. Hiperspektral imge sınıflandırma amacıyla, Borges ve diğ. (2006; 2011) tarafından önerilen yöntemlerde Krishnapuram ve diğ. (2005) tarafından önerilen seyrek çokterimli lojistik bağlanım modeli kullanılmış olup, bu çalışmadan farklı olarak Gauss-Siedel (Quarteroni, 2000) yöntemi ile daha hızlı bir algoritma önerilmiştir. Kayabol (2020) hiperspektral imgelerin sınıflandırılması için seyrek çokterimli lojistik bağlanım modelini kullanmış, bu modelin eğitimi için bir yaklaşık parametre kestirim yöntemi önermiştir.

Tezde ayrımsayıcı bir model de geliştirilmiş olup, sınıf etiketinin koşullu olasılığı izgesel vektörlere ve komşu piksel etiketlerine bağlı olarak bir çokterimli lojistik bağlanım modeliyle doğrudan modellenmektedir. Uzamsal bilginin piksel etiket önselleriyle modele dahil edildiği durağan olmayan karışım modellerinden farklı olarak, bu yaklaşımda sınıf-koşullu olasılıklar ve sınıf etiket önselleri tanımlanmamakta, bunun yerine sınıf etiketinin sonsal dağılımı doğrudan modellenmektedir. Bu model bir koşullu rastgele alan (Lafferty ve diğ., 2001) olarak da düşünülebilir. Alam ve diğ. (2016) koşullu rastgele alan ve evrişimsel sinir ağı kullanılan süperpiksel tabanlı bir hiperspektral sınıflandırma yöntemi önermiştir.

Son yıllarda yapay öğrenme alanında derin öğrenme yaklaşımı yaygınlaşmış ve imge sınıflandırma, nesne tanıma, doğal dil işleme gibi çeşitli problemler için uygulanmıştır. 2015 yılından bu yana hiperspektral imge sınıflandırma literatüründe de derin öğrenme yöntemleri git gide daha fazla kullanılmaya başlanmıştır. Birçok imge işleme probleminde olduğu gibi hiperspektral imgelerde de sıkça kullanılan bir yöntem olan evrişimsel sinir ağları, evrişim ve havuz (pooling) işlemleri sayesinde otomatik olarak öznelilik çıkarmada oldukça başarılı olup, filtrelerin imgenin her yerinde kullanılabilmesi nedeniyle de parametre sayısı ve model karmaşıklığı açısından oldukça etkindir (Zhu ve diğ., 2017).

Hiperspektral imge sınıflandırmada evrişimsel sinir ağlarının kullanıldığı ilk çalışmalardan biri Hu ve diğ. (2015) tarafından yapılmış olup, bu çalışmada spektral vektörler üzerinde tek boyutlu evrişimsel katmanlar kullanılmıştır. Hiperspektral imge sınıflandırmada tek boyutlu evrişimsel sinir ağları Wang ve diğ. (2019), Chen ve diğ. (2016) ile Haut ve diğ. (2018) tarafından da kullanılmıştır

Hiperspektral imgelerden izge-uzamsal özneliliklerin öğrenilmesi amacıyla iki ve üç boyutlu evrişimsel ağlar da kullanılmaktadır (Cao ve diğ., 2018; Chen ve diğ., 2016; Yue ve diğ., 2015; Zhang ve diğ., 2019). Yapay sinir ağı kullanılan sınıflandırma yöntemlerinde uzamsal bilgiden yararlanmak için Markov rastgele alan (Cao ve diğ., 2018) ve koşullu rastgele alan (Alam ve diğ., 2016; 2019; Zhong ve diğ. 2020) kullanılan çalışmalar da bulunmaktadır.

Chen ve diğ. (2016) tarafından bir, iki ve üç boyutlu derin evrişimsel sinir ağları spektral, uzamsal ve izge-uzamsal öznelilik çıkarma için kullanılmış, aşırı öğrenmeyi önlemek için L_2 normu, dropout ve ReLU kullanılmış, eğitim kümesi büyüklüğünün artırılması için de yapay örnekler üretilmiştir. Li ve diğ. (2018) evrişim ve ters-evrişim fonksiyonlarını kullanarak, aşırı

öğrenme makinesi algoritmasını uygulamıştır. Jia ve diğ. (2016) hiperspektral imge sınıflandırma için tek ve çift evrişim katmanlı sinir ağlarını kullanmış ve farklı parametreler için deney sonuçlarını raporlamıştır. Bu çalışmada spektral vektörlerin elemanları matris oluşturacak şekilde yeniden sıralanmıştır. Alipourfard ve diğ. (2008) hiperspektral imgeye önce bir alt-uzay öğrenme yöntemiyle boyut indirgeme uygulamış, sonra imgeyi parçalar hâlinde bir evrişimsel sinir ağına vermiştir. Lee ve Kwon (2016) farklı boyutlarda filtreler içeren bir üç boyutlu evrişimsel sinir ağı ile izgesel ve uzamsal öznitelik öğrenme sağlayarak sınıflandırma yapmıştır. Cao ve diğ. (2018) üç boyutlu evrişimsel sinir ağı ile izge-uzamsal sınıflandırma yapmış, sonra da Markov rastgele alan önseli ile sınıflandırma haritasına düzleştirme uygulamıştır.

Yapay sinir ağlarının eğitiminde genellikle geri yayılım algoritması (Rumelhart ve diğ., 1986) kullanılmaktadır. Bu algorithmada ağın çıkışında hesaplanan maliyet fonksiyonunun parametrelere geri yayılması ve genellikle en dik iniş algoritması temel alınarak bu parametrelerin optimize edilmesi sağlanmaktadır. Maliyet fonksiyonunun genellikle karmaşık ve doğrusal olmayan yapısı nedeniyle kapalı formda çözüm bulunamamakta, bu nedenle yinelemeli optimizasyon yöntemlerine ihtiyaç duyulmaktadır.

Geri yayılım algoritması ile sıkça kullanılan en dik iniş algoritmasında adım büyüklüğünün seçimi büyük önem taşımaktadır. Adım büyüklüğünün küçük seçilmesi durumunda ihtiyaç duyulan yineleme sayısı artmakta, büyük seçilmesi durumunda ise lokal minimumun bulunamaması durumuyla karşılaşilmektedir. Bunun için AdaGrad (Duchi ve diğ., 2011), ADAM (Kingma ve Ba, 2015), Adadelta (Zeiler, 2012) gibi yöntemlerde olduğu gibi uyarlanabilir adım büyüklüğü ile adım büyüklüğünün eğitim iterasyon sayısına göre değiştirilmesi gibi çözümler önerilmiştir.

Birinci derece optimizasyon algoritmalarından farklı olarak, ikinci derece yöntemlerde adım boyu algoritmik olarak hesaplanmakta ve algoritmanın daha az yineleme sayısı ile yakınsaması mümkün olmaktadır. Buna karşın, ikinci derece türevlerin hesaplanmasının maliyetli veya bazı durumlarda mümkün olmaması nedeniyle, ikinci derece optimizasyon algoritmalarının uygulanması daha zor olmaktadır. Örneğin Newton-Raphson algoritmasında Hessian matrisinin hesaplanması yüksek işlemci gücü ve depolama gereksinimi nedeniyle zor olabilmektedir. Bu nedenle sabit bir yaklaşık Hessian matrisi (Böhning, 1992), quasi-Newton yöntemleri ve eşlenik gradyan gibi yaklaşık hesaplama yöntemleri önerilmiştir.

Yapay sinir ağlarının eğitiminde genellikle stokastik en dik iniş tabanlı algoritmalar kullanılsa da ikinci derece yöntemlerin kullanıldığı bazı algoritmalar önerilmiştir. Le ve diğ. (2011) eşlenik gradyan, bir quasi-Newton algoritması olan sınırlı hafızalı Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algoritması ve stokastik en dik iniş algoritmaları ile bir karşılaştırma yapmıştır. Martens (2010) derin oto-kodlayıcılar için Hessian-bağımsız bir algoritma önermiştir. Martens ve Grosse (2015) tarafından önerilmiş olan K-FAC adlı algoritma ile ikinci derece bilgi için Fisher matrisi yaklaşık olarak hesaplanmaktadır. Gauss-Newton (Botev ve diğ., 2017; Tan ve Lim, 2019), Levenberg-Marquardt (Wilamowski ve Yu, 2010a; 2010b; Hagan ve Menhaj, 1994), yaklaşık en büyük iniş (Tan ve diğ., 2017), ölçeklenmiş eşlenik gradyan (Møller, 1993) ve Fletcher-Reeves (Chen ve diğ., 2019) gibi başka algoritmalar da yapay sinir ağlarının eğitiminde kullanılmıştır.

Hiperspektral imge sınıflandırma modellerinin geliştirilmesinde veri kümelerinin sınırlı olması karşılaşılan önemli problemlerden biridir. Modellerin başarımı ve başarımların ölçümünün doğruluğu verinin doğru şekilde eğitim ve test kümelerine ayrılmasına oldukça bağlıdır. Literatürde modellerin eğitimi ve başarımların ölçümü genellikle aynı imgeler üzerinde yapılmakta ve bir imge hem eğitim hem test için kullanılmaktadır. Bu nedenle, hiperspektral imgelerin sınıflandırılması amacıyla uzamsal veya izge-uzamsal modellerin uygulanmasında karşılaşılan önemli bir sorun imgenin eğitim ve test kümelerine ayrılmasıdır. İmge içinde bir bölgenin eğitim için seçilmesi durumunda imgedeki tüm sınıflardan eğitim veya test pikseli bulunamamaktadır, bu nedenle imgenin farklı bölgelerinden piksellerin seçilmesi gerekmektedir. Eğitim kümesindeki bir pikselin test kümesindeki komşularının göz ardı edilerek parametre kestiriminde kullanılmaması, test sırasında ise bir test pikselinin komşularından eğitim kümesinde bulunanların göz ardı edilmesi gerekmektedir. Ancak literatürde hem koşullu rastgele alan gibi uzamsal modellerin hem de iki veya üç boyutlu evrimsel sinir ağlarının uygulanmasında bu gereklilik göz ardı edilmekte ve yanlış deney kurgularıyla olması gerekenden daha yüksek başarımlar raporlanmaktadır. Bu şekilde yanlış kurgulanan deneylerle başarımların ölçümünün doğru yapılmadığı da bazı çalışmalarda açıklanmıştır (Audebert ve diğ., 2019; Nalepa ve diğ., 2019).

1.1. TEZİN KATKILARI

Tezin hiperspektral imge işleme literatürüne katkıları, az örnekle eğitim yapılarak izge-uzamsal sınıflandırma, bir üretici alt-uzay yöntemiyle boyut indirgeme ve izge-uzamsal sınıflandırma,

çokterimli lojistik bağlanım modeli temel alınarak geliştirilen üretici ve ayımsayıcı modellerle izge-uzamsal sınıflandırma ve evrimsel sinir ağları kullanılarak izge uzamsal sınıflandırma olmak üzere dört gruba ayrılmaktadır.

Az örnekle sınıflandırma amacıyla iki farklı durağan olmayan karışım modeli kullanılmış, piksel etiketleri üzerinde tanımlanan önsel dağılımlarla birlikte izge-uzamsal sınıflandırma modelleri elde edilmiştir. Karışım oranları pikselin konumuna göre değiştiğinden, bu modeller uzamsal olarak değişen veya durağan olmayan modeller olarak adlandırılmaktadır. Bu yöntemlerde spektral vektörler Gauss ve Student-t dağılımları ile modellenmektedir. Hiperspektral imgelerde eğitim örnek sayısı az olduğundan, parametreler üzerinde tanımlanan önsel dağılımlarla kestirim işlemi kolaylaştırılmaktadır. Verideki aykırı örneklerin sınıflandırma başarımını etkilemesinin önlenmesi amacıyla, geliştirilen modellerden birinde sınıf-koşullu olasılıklar Student-t dağılımlarıyla modellenmiştir.

Az örnekle parametre kestiriminin zorluğu ve spektral bant sayısının büyük olmasına bağlı olarak model karmaşıklığının fazla olması sorunlarının giderilmesi amacıyla, bir üretici alt-uzay modeli geliştirilmiş olup, bu modelde olasılıksal temel bileşen analizi (Tipping ve Bishop, 1999) kullanılmış, böylece spektral vektörler bir alt-uzay Gauss karışımı ile modellenmiştir. Piksel etiket vektörleri üzerinde çokterimli özlojistik bağlanım önselleriyle tanımlanarak bir üretici sınıflandırma modeli elde edilmiş olup, karışım oranlarının piksel konumuna bağlı olarak değişmesi nedeniyle bu model de uzamsal olarak değişen bir modeldir. Bu yöntemde olasılıksal yaklaşımla boyut indirgeme ve izge-uzamsal sınıflandırma tek bir modelde gerçekleştirilmektedir.

Tezin üçüncü katkısı çokterimli lojistik bağlanım modelinin temel alındığı, üretici ve ayımsayıcı olmak üzere iki ayrı izge-uzamsal sınıflandırma modelidir. Üretici modelde sınıf-koşullu olasılıklar çokterimli lojistik bağlanım modeli ile modellenmekte, piksel etiketleri üzerinde çokterimli özlojistik bağlanım önselleri tanımlanarak bir lojistik bağlanım karışımı elde edilmektedir. Ayımsayıcı sınıflandırıcıda ise sınıf-koşullu olasılıklar ve etiket önseli tanımlanmamakta, etiket değişkeninin sonsal dağılımı spektral vektöre ve komşu piksel etiketlerine bağlı olarak doğrudan bir çokterimli lojistik bağlanım modeli şeklinde modellenmekte, dolayısıyla bir koşullu rastgele alan modeli elde edilmektedir.

Dördüncü yaklaşımda ise, spektral vektörler bir tek boyutlu evrişimsel sinir ağı ile modellenmekte, piksel etiketleri üzerinde uzamsal önsel dağılım tanımlanarak da uzamsal düzleştirme yapılmaktadır. Bu yöntemle sınıflandırıcı eğitimiyle bütünleşik olarak spektral vektörlerden özniteliklerin çıkarılması sağlanmaktadır. Ayrıca bu modelin eğitilmesi için hem birinci hem de ikinci derece optimizasyon yöntemlerinin birlikte kullanıldığı bir eğitim yöntemi önerilmekte, bu yöntemle yalnız birinci derece optimizasyon yapılması durumuna göre daha kısa sürede eğitim yapılabilir.

Tez çalışmalarında temel alınan ilk makalede (Kayabol, 2015) hiperspektral imgedeki piksel vektörleri bir Bayesçi Gauss karışım modeli ile modellenmiş, pikseller arası komşuluk ilişkileri de çokterimli özlojistik bağlanım modeli (Kayabol ve diğ., 2011, Kayabol ve Zerubia, 2013) ile modellenmiştir. Kayabol (2015) tarafından önerilen Bayesçi Gauss karışım modelinde ortalama ve kovaryans matrisi için normal-ters Wishart önsel dağılımı kullanılmıştır. Tezdeki ilk çalışma, bu model temel alınarak geliştirilmiş olan, Bayesçi Student-t karışım modelidir. Student-t dağılımı, normal dağılıma göre aykırı örnekleri daha iyi modelleyebilmekte, dolayısıyla bu dağılım ile gürbüz sınıflandırma modelleri elde edilebilmektedir (Bishop, 2006). Ortalama ve kovaryans matrisi için önsel olasılık dağılımları tanımlanarak bir Bayesçi Student-t karışım modeli tanımlanmış ve bu model için parametre kestirim algoritmaları elde edilmiştir.

Bu çalışmadan sonra, Bayesçi Gauss karışım modelinde ortalama vektörü ve kovaryans matrisi için tanımlanan normal-ters Wishart önsel dağılımı için farklı bir parametre daha tanımlanmış, bu parametreyle örnek kovaryans matrisi ve kovaryans önseli arasında bir ağırlıklandırma yapılarak daha iyi parametre kestirimi sağlanmıştır (Kayabol ve Kutluk, 2016).

Bayesçi Gauss karışım modeli ile Bayesçi Student-t karışım modelinde, piksellerin spektral vektörleri üzerinde bir öznitelik çıkarma işlemi uygulanmamış, bu vektörler doğrudan sınıflandırıcıya verilerek eğitim ve test amacıyla kullanılmıştır.

Yapay öğrenmede boyut indirgeme önemli konulardan biridir. Bu yöntemlerle verinin daha düşük boyutlu bir uzayda temsil edilmesi ve bu temsil ile hem veri işleme kolaylığının sağlanması hem de sınıflandırma doğruluğunun artırılması hedeflenmektedir. Bu doğrultuda yapılan alt-uzay öğrenme çalışmalarıyla geliştirilen yöntemde, izgesel vektörlerin olasılıksal temel bileşen analizi ile modellenmesiyle bu vektörlerin alt-uzay temsilleri elde edilmekte ve vektör boyutları küçültülmektedir. Olasılıksal temel bileşen analizi, Tipping ve Bishop (1999)

tarafından önerilmiş bir model olup, temel bileşen analizinin olasılıksal olarak formüle edilmesiyle geliştirilmiştir. Temel bileşen analizi verinin daha düşük boyutlu bir alt-uzaya izdüşümü olarak tanımlanabilirken, olasılıksal temel bileşen analizi bu yöntemin bir olasılıksal gizli değişken modeli ve en büyük olabilirlik çözümü olarak düşünülebilir (Bishop, 2006). İzgesel vektörlerin az sayıda temel bileşenle tanımlanan bir uzaydaki izdüşümü ile düşük boyutlu temsiller elde edilmekte ve bunlarla bir Gauss karışım modeli oluşturulmaktadır. Aynı zamanda bu vektörler ile yapılan sınıflandırmada daha iyi sonuçlar elde edildiği de deneylerle gösterilmiştir.

Tez kapsamında piksellerin uzamsal ilişkilerinin modellenmesi için iki ayrı yaklaşım uygulanmıştır. İlk yaklaşımda, spektral imzaların modellenmesi için kullanılan yöntemlerle bütünleşik olarak bir özlojistik bağlanım modelinden yararlanılmaktadır. Bu yaklaşımda piksel etiketleri üzerinde bir önsel dağılım tanımlanmakta, izgesel model ile birlikte, piksel etiketlerinin sonsal olasılığı en büyüklenmektedir. Özlojistik bağlanım modelinde, eğitim aşamasında uzamsal parametre kestirimi yapılmadığından, piksel etiketlerinin sonsal olasılığının en büyüklenmesi, spektral sınıflandırma sonrasında uzamsal düzleştirme işlemi olarak düşünülebilir. Bu modelde uzamsal olarak değişen karışım modelleri oluşturulmaktadır. Bu ilk yaklaşım bir üretici modelleme yaklaşımıdır. Diğer yaklaşımda ise bir koşullu rastgele alan modeli kullanılmakta, izgesel ve uzamsal model tek bir çokterimli lojistik bağlanım modeli şeklinde tanımlanmakta ve bunların parametre kestirimi eğitim aşamasında birlikte yapılmaktadır. Bu modelde sınıf etiketlerinin koşullu olasılığı doğrudan modellenmekte, yani sınıf-koşullu olasılıklar ve etiket önselleri ayrı ayrı modellenmemektedir, dolayısıyla bu bir ayrımsayıcı modelleme yaklaşımıdır.

Son yıllarda yapay öğrenme ve bilgisayarlı görü gibi pek çok alanda derin öğrenme modelleri çok yaygın kullanım alanı bulmuştur. Sinir ağları, karmaşık dağılıma sahip verinin modellenmesini sağlayan esnek istatistiksel modellerdir (Neal, 1996). Tez kapsamında da evrimsel sinir ağları (LeCun, 1989) kullanılarak derin öğrenmeye yer verilmiştir. En az bir katmanında evrim işlemi uygulanan yapay sinir ağları evrimsel sinir ağları olarak adlandırılmaktadır (Goodfellow ve diğ., 2016). İmge sınıflandırma gibi problemlerde çokça kullanılan evrimsel sinir ağlarının başarılı olmasının en önemli nedeni öznel çıkarma ve sınıflandırıcı eğitiminin bütünleşik olarak gerçekleştirilmesidir. Bu ağların eğitiminde genellikle birinci derece optimizasyon algoritmaları kullanılmaktadır.

Tezde kullanılan evrişimsel sinir ağlarının eğitimi için birinci ve ikinci derece optimizasyon algoritmalarının birlikte kullanıldığı bir eğitim yöntemi geliştirilmiştir. Bu yöntemde, ağın son katmanındaki softmax sınıflandırıcısı için Kayabol (2020) tarafından önerilen yaklaşık ikinci derece parametre kestirimi yöntemi kullanılmakta, ağın diğer katmanları içinse birinci derece optimizasyon yöntemleri kullanılmaktadır. Böylece optimizasyon maliyetindeki artış sınırlı tutularak eğitim için gerekli iterasyon sayısı ve eğitim süresi azaltılmaktadır. Özlojistik bağlanım modeli ile uzamsal düzleştirme de yapılarak, hiperspektral imgelerin sınıflandırılması için derin öğrenme tabanlı bir izge-uzamsal sınıflandırma yöntemi geliştirilmiştir. Geliştirilen eğitim yöntemi yalnızca evrişimsel sinir ağları için değil, son katmanı çokterimli lojistik bağlanım fonksiyonundan oluşan herhangi bir derin ağa da uygulanabilmekte, diğer mimariler ve eğitim yöntemleriyle birlikte kullanılabilmekte ve çeşitli problemlere uygulanabilmektedir.

Tezde kullanılan derin öğrenme modelleri tek boyutlu evrişimsel sinir ağlarıdır. Literatürde iki ve üç boyutlu evrişimsel sinir ağlarının kullanıldığı çalışmalar bulunmasına karşın, Audebert ve diğ. (2019) ile Nalepa ve diğ. (2019) tarafından da açıklandığı gibi, bu yöntemlerde evrişim penceresine hem eğitim hem test piksellerinin girmesiyle bu kümeler arasında bilgi geçişi olmakta, dolayısıyla yöntem başarımları doğru değerlendirilememektedir. Bu nedenle tezde önerilen derin öğrenme yöntemlerinde tek boyutlu evrişimsel sinir ağları kullanılmış, piksel etiketleri üzerinde tanımlanan uzamsal önsellerle de izge-uzamsal sınıflandırma yapılması sağlanmıştır.

2. GENEL KISIMLAR

2.1. YAPAY ÖĞRENME

Yapay öğrenme, bir kurallar dizisi şeklinde tanımlanamayan bir algoritmanın veriden öğrenilebilmesini amaçlayan bir araştırma dalıdır. Temel olarak, bu tez kapsamında incelenen algoritmalar bakımından, yapay öğrenme algoritmaları eğitici ve eğitici olmayan öğrenme olarak ikiye ayrılabilir. Eğitici öğrenme algoritmaları ise sınıflandırma ve bağlanım olarak ikiye ayrılmaktadır.

Eğitici öğrenmede, algoritmanın etiketlenmiş bir eğitim verisi üzerinde giriş-çıkış ilişkisini öğrenmesi sağlanırken, eğitici olmayan öğrenmede bir etiketli eğitim verisi bulunmaz. Sınıflandırma ve bağlanım algoritmaları arasındaki temel fark ise, sınıflandırma algoritmalarıyla yapılan tahmin sonucunda, girişteki örneğe karşılık bir kategorik çıkış elde edilmesi, bağlanım algoritmalarında ise sürekli bir değer elde edilmesidir.

Bir eğitici yapay öğrenme modeli, en basit şekliyle, bir x gözlemi için, bu gözleme ilişkin y sınıf etiketini tahmin eden bir $f(x)$ fonksiyonu olarak tanımlanabilir. Bu fonksiyon, (x, y) çiftlerinden oluşan eğitim veri kümesi $\{(x, y)\}$ ile tanımlanmaktadır. Bu fonksiyonun yaptığı tahmin gerçek sınıf etiketine ne kadar yakınsa, sınıflandırma doğruluğu o kadar yüksektir. Modelin eğitim verisi dışındaki örnekler üzerindeki sınıflandırma doğruluğu modelin genelleme başarısı olarak tanımlanmaktadır (Bishop, 2006). Karmaşık bir modelin az veriyle eğitilmesi durumunda, modelin eğitim verisi üzerinde yüksek sınıflandırma başarımı göstermesi, buna karşın genelleme başarımının düşmesi mümkündür. Bu duruma aşırı öğrenme denmektedir (Alpaydın, 2013).

Sınıflandırılması istenen bir örneğe ilişkin yapılan ölçümler doğrudan sınıflandırma modeline verilebileceği gibi, bu ölçümlerin modele verilmeden önce işlenmesi de mümkündür. Bu ön işlemlere öznitelik çıkarma denmektedir. Öznitelik çıkarmada amaç bazen sınıflandırma doğruluğunun artırılması, bazen de sınıflandırma doğruluğu mümkün olduğunca korunarak işlem karmaşıklığının azaltılmasıdır (Bishop, 2006).

2.2. OLASILIKSAL MODELLEME VE SINIFLANDIRMA

Olasılıksal yaklaşımda, model parametreleri ve verideki belirsizlik olasılık dağılımları ile modellenmektedir. Yapay öğrenmede olasılıksal yaklaşımla modelin bileşenleri rastgele değişkenler olarak tanımlanmakta ve bu değişkenlerin birbiriyle ilişkileri olasılık dağılımları yoluyla modellenmektedir (Barber, 2012). Modeldeki belirsizliğin gözlenen veriyle güncellenmesiyle daha doğru olasılık dağılımları elde edilir. Olasılıksal yaklaşımda, Bayes kuralı ve diğer olasılık kuralları çıkarsama yapmak için kullanılır. Olasılıksal modelleme ile veri gözlenmeden önce tanımlanan önsel dağılımların, gözlem sonrası elde edilen sonsal dağılımlara dönüştürülmesi için bir yöntem elde edilmektedir (Ghahramani, 2015).

2.2.1. Bayesçi Yaklaşım

Bayesçi yaklaşım, modelin bileşenlerine ilişkin önsel bilginin, gözlenen veriyle güncellenmesini ve sonsal olasılık dağılımlarının elde edilmesini sağlamaktadır.

Bağımsız ve özdeşçe dağılan N adet gözlemden oluşan ve $D = \{x_1, \dots, x_N\}$ şeklinde gösterilen bir veri kümesi olsun. Gözlemlerin θ parametre kümesine sahip bir dağılıma göre dağıldığı kabul edilirse, bu dağılım şöyle gösterilebilir:

$$x_i \sim p(x_i|\theta) \quad (2.1)$$

Tüm örneklerin bileşik olasılığı ise şöyle yazılabilir:

$$p(D|\theta) = \prod_{i=1}^N p(x_i|\theta) \quad (2.2)$$

Burada $p(D|\theta)$, olabilirlik fonksiyonu olarak adlandırılır ve θ 'nın bir fonksiyonu olarak düşünüldüğünde, θ 'nın olasılık dağılımı değildir. Olabilirlik fonksiyonu, gözlenen veri için, parametre değerlerinin uygunluğunun sayısal olarak ölçülmesini mümkün kılmaktadır (Bishop, 2006).

Parametrelerin olabilirlik fonksiyonuyla birlikte, parametrelere ilişkin önsel olasılık dağılımı $p(\theta)$ tanımlandığında, gözlenen veriden yararlanılarak parametrelerin sonsal olasılık dağılımı Bayes kuralı ile şöyle yazılabilir (Bishop, 2006):

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \quad (2.3)$$

Paydadaki terim $p(D)$, D değişkeninin marjinal olasılık dağılımı olup şöyle yazılabilir:

$$p(D) = \int_{\theta} p(D|\theta)p(\theta) \quad (2.4)$$

Bu terim normalizasyon terimi olup, bazı durumlarda göz ardı edilebilir ve Bayes eşitliği oransallık şeklinde şöyle ifade edilebilir:

$$p(\theta|D) \propto p(D|\theta)p(\theta) \quad (2.5)$$

2.2.2. Parametre Kestirimi

Eğitici bir yapay öğrenme modelinin eğitimi, model tanımlandıktan sonra, etiketli veri kullanılarak model parametrelerinin kestirilmesidir.

N adet örnek içeren bir veri kümesi $D = \{x_1, \dots, x_N\}$ için, bu örnekler bağımsız ve özdeşçe dağılıyorsa, olabilirlik fonksiyonu şöyle yazılabilir:

$$p(D|\theta) = \prod_{i=1}^N p(x_i|\theta) \quad (2.6)$$

Olabilirlik fonksiyonu, parametrelerin bir fonksiyonudur, ancak parametrelerin olasılık dağılımı değildir. Parametrelerin kestirilmesi için olabilirlik fonksiyonunu en büyükleyen parametreler seçilebilir, buna en büyük olabilirlik (ML: maximum likelihood) kestirimi adı verilmektedir.

Parametrelerle ilgili önbilginin de kullanılması durumunda, parametrelerin sonsal olasılık dağılımı en büyüklenir ve bu yöntem de en büyük sonsal (MAP: maximum a posteriori) kestirimi olarak adlandırılmaktadır. Parametrelerin sonsal dağılımı şöyle yazılabilir:

$$p(\theta|D) \propto p(D|\theta)p(\theta) \quad (2.7)$$

Parametre kestirimi yapılırken, işlem kolaylığı için genellikle en büyüklenecek ifadelerin logaritması alındıktan sonra optimizasyon yapılır (Bishop, 2006).

2.2.3. Sınıflandırma

Bir sınıflandırma modelinde, gözlenen veri x , sınıf etiketleri y , model parametreleri de θ rastgele değişkenleriyle gösterildiğinde, sınıf-koşullu olasılıklar $p(x|y, \theta)$ ve sınıf etiketlerinin önsel dağılımı $p(y)$ ile Bayes eşitliğinden yararlanılarak, sınıf etiketlerinin sonsal dağılımı şöyle yazılabilir:

$$p(y|x, \theta) = \frac{p(x|y, \theta)p(y)}{p(x|\theta)} \quad (2.8)$$

Bir olasılıksal modelde sınıflandırma yapılabilmesi için, $p(y|x, \theta)$ olasılığını en büyükleyen y değeri seçilmektedir.

2.2.4. Üretici ve Ayrımsayıcı Modeller

Üretici sınıflandırma modellerinde sınıf-koşullu olasılık $p(x|y, \theta)$ ile sınıf etiketlerinin önsel dağılımı $p(y)$ ayrı ayrı tanımlanarak, giriş değişkenlerinin bileşik olasılığı $p(x, y|\theta)$ modellenmekte ve sınıflandırma için $p(y|x, \theta)$ olasılığı elde edilmektedir. Ayrımsayıcı modellerde ise, sınıf-koşullu olasılıklar ve önsel dağılımlar ayrı ayrı tanımlanmamakta, $p(y|x, \theta)$ dağılımı doğrudan modellenmektedir (Ng ve Jordan, 2002).

Bir ayrımsayıcı sınıflandırma modelinin eğitiminde koşullu log-olabilirlik fonksiyonu $\sum_{n=1}^N \ln p(y_n|x_n, \theta)$ en büyüklenecek şekilde model parametreleri belirlenirken, üretici sınıflandırma modelinde bileşik log-olabilirlik fonksiyonu $\sum_{n=1}^N \ln p(x_n, y_n|\theta)$ kullanılır (Murphy, 2012).

2.2.5. Önsel Dağılımlar ve Düzenleştirme

$p(x|y, \theta)$ şeklinde bir dağılım için, $p(\theta)$ şeklinde bir önsel dağılım tanımlandığında, $p(\theta|x, y)$ sonsal dağılımı önsel ile aynı şekle sahipse, bu önsel dağılım, $p(x|y, \theta)$ olabilirlik fonksiyonu için eşlenik önsel olarak tanımlanır. Üstel ailedeki tüm dağılımlar için bir eşlenik önsel bulunabilir. Eşlenik önsel kullanılması durumunda, önsel dağılımın analitik ifadesi biliniyorsa, sonsal dağılım da analitik olarak bulunabilir. Eşlenik önsel kullanılmaması durumunda

genellikle sonsal dağılımın analitik olarak bulunması mümkün olmamakta, bu nedenle rastgele örnekleme ve varyasyonel yöntemlerle sonsal dağılımın yaklaşık olarak hesaplanması mümkün olmaktadır (Bishop, 2006).

En büyük olabilirlik yöntemi genel olarak aşırı öğrenmeye açık bir yöntemdir. Bayesçi yaklaşımla, parametreler üzerinde tanımlanan önsel dağılımlarla parametrelerin alabileceği değerler sınırlandırılmakta, dolayısıyla düzenlileştirilme yapılmaktadır. Bir diğer ifadeyle, model karmaşıklığı sınırlandırmakta ve aşırı öğrenme durumu engellenmektedir. Yapay öğrenmede sıkça kullanılan düzenlileştirme yöntemlerinden L_2 normu Gauss önsel dağılımına, L_1 normu ise Laplace önsel dağılımına karşılık gelmektedir.

2.3. HİPERSPEKTRAL İMGELERİN SINIFLANDIRILMASI

2.3.1. Olasılıksal Yaklaşımla Problem Tanımı

Olasılıksal yaklaşımla, hiperspektral imgenin her pikseline ilişkin spektral vektör ve pikselin sınıf etiketi, birer rasgele değişken ile ilişkilendirilmektedir. $N_1 \times N_2$ boyutunda bir hiperspektral imgede K adet sınıf, k . sınıfta N_k adet piksel ve L adet izgesel bant olsun. Bu imgenin n . pikseline ilişkin izgesel vektör $\mathbf{s}_n$ ile, bu pikselin sınıf etiketi $\mathbf{z}_n$ ile, model parametreleri kümesi de θ ile gösterilsin. Sınıf etiketinin, k . elemanı $z_{n,k}$ ile gösterildiğinde, $z_{n,k} \in \{0,1\}$ ve $\sum_{k=1}^K z_{n,k} = 1$ koşullarını sağladığı kabul edilmektedir, böylece $\mathbf{z}_n$ etiket vektörlerinin alabileceği değerler kümesi $\mathcal{Z} = \{[1,0, \dots, 0], [0,1, \dots, 0], \dots, [0,0, \dots, 1]\}$ şeklinde tanımlanmaktadır. $\mathbf{s}_n$ pikselinin sınıflandırılması için $p(\mathbf{z}_n|\mathbf{s}_n, \theta)$ olasılığını en büyükleyen etiket vektörü seçilmektedir.

Yapay öğrenmede sınıflandırma bir eğitici öğrenme yöntemi olup, etiketli veri üzerinde giriş-çıkış ilişkisinden yararlanılarak model parametreleri kestirilmekte ve bu parametreler etiketi bilinmeyen örneklerin sınıflandırılması için kullanılmaktadır.

2.3.2. İzge-Uzamsal Sınıflandırma

İzgesel sınıflandırma, piksellerin spektral vektörlerinin modellenmesi ve sınıflandırılması olarak tanımlanırken, uzamsal sınıflandırma piksellerin komşuluk ilişkilerine göre sınıflandırılmasıdır. Dolayısıyla uzamsal sınıflandırmada spektral bilgidan

yararlanılmamaktadır ve bu durumda piksel etiketlerinin birbirlerinden bağımsız olduğu söylenebilir.

İzge-uzamsal sınıflandırmada hem spektral bilgiden yararlanılmakta hem de birbirine yakın olan piksellerin aynı sınıftan olması olasılığının yüksek olduğu kabulüyle, spektral bilgiye ek olarak komşu piksellerin etiketlerinden de yararlanılmaktadır (Audebert ve diğ., 2019).



3. MALZEME VE YÖNTEM

Bu tezde hiperspektral imgelerin izge-uzamsal sınıflandırılması için üretici ve ayırmsayıcı olasılıksal modeller geliştirilmiştir. Bu modeller temel olarak dört gruba ayrılmaktadır.

Hiperspektral imge sınıflandırmada karşılaşılan sorunlardan biri veri etiketlemenin zorluğu ve dolayısıyla eğitim kümesi için sınırlı sayıda örnek bulunabilmesidir. Uygulanan ilk yaklaşımda, az sayıda örnekle eğitim yapılabilmesi için model parametreleri üzerinde tanımlanan önsel dağılımlarla parametre kestirimi yapılmaktadır. Bayeşçi yaklaşımda, önsel dağılımların tanımlanması model parametrelerinin düzenlenilmesini ve parametre kestirimini kolaylaştırılmasını sağlamaktadır. Bu yaklaşımda, izge-uzamsal olasılıkların Student-t ve Gauss dağılımı şeklinde seçildiği iki ayrı durağan olmayan karışım modeli geliştirilmiştir. Her iki yöntemde de sınıf etiketleri üzerinde tanımlanan çokterimli özlojistik bağlanım önselleri (Kayabol ve diğ., 2011; Kayabol ve Zerubia, 2013) ile uzamsal düzleştirme de yapılarak daha iyi sınıflandırma haritaları elde edilmiştir.

İkinci yaklaşımda, bir üretici alt-uzay yöntemi ile veri boyutunun indirgenmesi sağlanmakta ve bir durağan olmayan karışım modeli elde edilmektedir. Bu alt-uzay yönteminde spektral vektörlerin gizli değişken modeli şeklinde modellenmesi ve olasılıksal temel bileşen analizi ile düşük boyutlu temsillerinin elde edilmesi sağlanmaktadır. Böylece hem model parametrelerinin kestirimi kolaylaşmakta hem de iyi sınıflandırma sonuçları elde edilmektedir. Bu yaklaşımda da sınıf etiketleri üzerinde tanımlanan özlojistik önselleri ile bir üretici izge-uzamsal sınıflandırma modeli elde edilmektedir.

Üçüncü yaklaşımda iki farklı çokterimli lojistik bağlanım modeli ile izge-uzamsal sınıflandırma yapılmaktadır. İlk model, bir üretici model olup, sınıf-koşullu olasılıklar çokterimli lojistik bağlanım modelleri şeklinde modellenmekte, sınıf etiketleri üzerinde de özlojistik bağlanım önselleri tanımlanarak bir üretici karışım modeli elde edilmektedir. İkinci yaklaşımda ise hem spektral vektörlere hem de piksel etiketlerine bağlı bir çokterimli lojistik bağlanım modeli önerilmektedir. Bu yaklaşımda sınıf etiketinin pikselin spektral vektörüne ve diğer piksellerin etiketlerine bağlı sonsal dağılımı doğrudan modellendiğinden, bu ayırmsayıcı bir sınıflandırma modelidir.

Son yaklaşımda ise, evrişimsel sinir ağlarının öznetelik öğrenmedeki başarısından yararlanılması hedeflenmiştir. Bu amaçla, farklı evrişimsel sinir ağı modelleri ile sınıflandırma yapılmıştır. Burada tek boyutlu evrişimsel katmanlar kullanılmış olup, uzamsal sınıflandırma için piksel etiketleri üzerinde tanımlanan özlojistik bağlanım önselleri ile uzamsal bilgi de sınıflandırma modeline dahil edilmiştir. Evrişimsel sinir ağlarının eğitiminde birinci ve ikinci derece optimizasyon yöntemlerinin birlikte kullanıldığı bir eğitim algoritması geliştirilmiş, stokastik gradyan iniş algoritmasının kullanıldığı geri yayılım algoritmasıyla eğitime göre daha az sayıda yineleme ile ve daha kısa sürede eğitim yapılabilirdiği gösterilmiştir.

3.1. AZ SAYIDA ÖRNEKLE EĞİTİM VE İZGE-UZAMSAL SINIFLANDIRMA

Hiperspektral imgelerse spektral bant sayısı fazla, buna karşın, etiketli piksel sayısı azdır. Dolayısıyla bu problemde parametre kestirimi zordur. Bu nedenle, tezde önerilen ilk yaklaşımda, iki ayrı üretici sınıflandırıcı geliştirilmiş olup, bunlar durağan olmayan Bayesçi Gauss ve Student-t karışım modelleridir. Bayesçi yaklaşımla, parametreler üzerinde önsel dağılımlar tanımlanarak parametre kestirimi kolaylaştırılmaktadır. Sınıf etiketleri üzerinde tanımlanan önsel dağılımlarla da üretici modeller elde edilmekte olup, bu önsel dağılımlara bağlı olarak karışım oranlarının konuma göre değişmesi nedeniyle bu modeller durağan olmayan veya uzamsal olarak değişen modellerdir.

Bölüm 2.2.4'te açıklandığı gibi, bir üretici sınıflandırma modelinde giriş, sınıf etiketi ve model parametrelerinin bileşik olasılığı $p(\mathbf{s}_n, \mathbf{z}_n, \theta)$ sınıf-koşullu olasılıklar ve önsel dağılımların çarpımı şeklinde ifade edilir:

$$p(\mathbf{s}_n, \mathbf{z}_n, \theta) = p(\mathbf{s}_n | \mathbf{z}_n, \theta) p(\mathbf{z}_n) p(\theta) \quad (3.1)$$

Böyle bir modelde sınıflandırma yapılabilmesi için, sınıf etiketinin sonsal dağılımını en büyükleyen sınıf etiketi seçilmektedir:

$$\begin{aligned} \hat{\mathbf{z}}_n &= \operatorname{argmax}_{\mathbf{z}_n} p(\mathbf{z}_n | \mathbf{s}_n, \theta) = \operatorname{argmax}_{\mathbf{z}_n} \frac{p(\mathbf{s}_n | \mathbf{z}_n, \theta) p(\mathbf{z}_n) p(\theta)}{\sum_{\mathbf{z}_n} p(\mathbf{s}_n | \mathbf{z}_n, \theta) p(\mathbf{z}_n) p(\theta)} \\ &= \operatorname{argmax}_{\mathbf{z}_n} \frac{p(\mathbf{s}_n | \mathbf{z}_n, \theta) p(\mathbf{z}_n)}{\sum_{\mathbf{z}_n} p(\mathbf{s}_n | \mathbf{z}_n, \theta) p(\mathbf{z}_n)} \end{aligned} \quad (3.2)$$

Burada sınıf etiketlerinin gizli değişken olduğu varsayılmaktadır.

Tezde kullanılan üretici modellerde, izge-uzamsal sınıflandırma yapılabilmesi için hem piksellerin izgesel imzaları için bir olasılık dağılımı tanımlanmakta, hem de piksellerin sınıf etiketleri üzerinde tanımlanan önsel dağılımlarla etiketlerin komşu piksellerin etiketlerine bağlı olarak modellenmesi sağlanmaktadır. Kullanılan üretici modeller temel olarak durağan olmayan veya konuma bağlı olarak değişen bir karışım modeli olup, karışım katsayıları piksel konumuna bağlı olarak değişmektedir. Tezde kullanılan tüm üretici modellerde sınıf etiketleri üzerinde tanımlanan önsel dağılım aynı kalmakla birlikte, sınıf-koşullu olasılıklar farklı dağılımlar şeklinde modellenerek problem için farklı çözümler geliştirilmiştir. Bir diğer ifadeyle, izgesel sınıflandırma farklı modellerle yapılmakta, etiket önselleri kullanılarak da uzamsal düzleştirme yapılmaktadır.

Karışım modelleri, karmaşık dağılıma sahip verinin modellenmesi için, tek bir basit olasılık dağılımı yerine, birden çok dağılımın doğrusal kombinasyonu ile elde edilen olasılık modelleridir (Bishop, 2006).

Bir sonlu karışım modelinde, bir pikselin olasılık dağılımı şöyle tanımlanmaktadır (Bishop, 2006):

$$p(\mathbf{s}_n) = \sum_{k=1}^K \pi_k p(\mathbf{s}_n | z_{n,k} = 1) \quad (3.3)$$

Burada π_k k . sınıfın karışım katsayısı olarak adlandırılmakta ve şöyle tanımlanmaktadır:

$$\pi_k = p(z_{n,k} = 1) \quad (3.4)$$

Bir piksele ilişkin etiket vektörü $\mathbf{z}_n$ 'in marjinal olasılık dağılımı kategorik bir dağılım olup şöyle yazılabilir:

$$p(\mathbf{z}_n) = \prod_{k=1}^K \pi_k^{z_{n,k}} \quad (3.5)$$

Piksellerin sınıf-koşullu dağılımı, karışımı oluşturan modellerin nasıl seçildiğine bağlıdır. Örneğin, Gauss karışım modelinde, bir piksele ilişkin sınıf-koşullu dağılım şöyle tanımlanmaktadır:

$$p(\mathbf{s}_n | z_{n,k} = 1) = \mathcal{N}(\mathbf{s}_n | \mu_k, \Sigma_k) \quad (3.6)$$

Sınıf etiket vektörünün dağılımına benzer şekilde, bu koşullu olasılık da şöyle yazılabilir:

$$p(\mathbf{s}_n | \mathbf{z}_n) = \prod_{k=1}^K p(\mathbf{s}_n | z_{n,k} = 1)^{z_{n,k}} \quad (3.7)$$

Burada $z_{n,k}$, k 'nin alabileceği değerlerin yalnızca biri için 1, diğerleri için 0 değerini alacağından, bu çarpım bir seçim fonksiyonu olarak düşünülebilir. Sınıf etiket değişkeninin önsel dağılımı ile sınıf-koşullu olasılıklardan elde edilen bileşik olasılık dağılımının marjinalize edilmesiyle, bir pikselin olasılık dağılımı şöyle yazılabilir:

$$\begin{aligned} p(\mathbf{s}_n) &= \sum_{\mathbf{z}_n} p(\mathbf{s}_n | \mathbf{z}_n) p(\mathbf{z}_n) = \sum_{\mathbf{z}_n} \prod_{k=1}^K \left(p(\mathbf{s}_n | z_{n,k} = 1) p(z_{n,k} = 1) \right)^{z_{n,k}} \\ &= \sum_{k=1}^K p(\mathbf{s}_n | z_{n,k} = 1) \pi_k \end{aligned} \quad (3.8)$$

Kayabol (2015) bir Bayesçi Gauss karışım modeli önermiş ve bu modeli hiperspektral imge sınıflandırma problemine uygulamıştır. Bu yöntem ile verideki her sınıf bir Gauss dağılımı ile modellenmektedir. Çokterimli özlojistik bağlanım modeli ile, piksel etiket vektörleri üzerinde bir önsel dağılım tanımlanmakta ve böylece karışım katsayıları her piksel için değişkenlik gösterdiğinden, durağan olmayan bir karışım modeli elde edilmektedir (Kayabol, 2015). Tezdeki üretici modeller bu çalışma temel alınarak çokterimli özlojistik bağlanım modeli ile elde edilen farklı karışım modelleridir. Bu modeller, sınıf-koşullu olasılıkların farklı şekillerde tanımlanmasıyla birbirinden ayrılmaktadır.

Çokterimli özlojistik bağlanım modelinde, hiperspektral imgedeki tüm piksellerin spektral vektörleri ile etiket vektörlerinin bileşik olasılığı şöyle yazılabilir (Kayabol ve Kutluk, 2016):

$$p(\mathbf{s}_{1:N}, \mathbf{z}_{1:N} | \theta_{1:K}, \beta) = p(\mathbf{s}_{1:N} | \mathbf{z}_{1:N}, \theta_{1:K}) p(\mathbf{z}_{1:N} | \beta) \quad (3.9)$$

Bu eşitlikte $\theta_{1:K}$ tüm sınıflara ilişkin izgesel model parametrelerini, β ise uzamsal modelin parametresini göstermektedir. Bu modelde piksel etiket vektörlerinin bağımlı, izgesel vektörlerin ise etiket vektörleri verildiğinde koşullu bağımsız olduğu kabul edilmektedir

(Kayabol ve Kutluk, 2016). Buna göre, denklem (3.9)'daki bileşik olasılık dağılımı şöyle yazılabilir:

$$p(\mathbf{s}_{1:N}, \mathbf{z}_{1:N} | \theta_{1:K}, \beta) = \left[\prod_{n=1}^N \prod_{k=1}^K p(\mathbf{s}_n | z_{n,k} = 1, \theta_k)^{z_{n,k}} \right] p(\mathbf{z}_{1:N} | \beta) \quad (3.10)$$

Piksel etiket vektörleri üzerinde tanımlanan $p(\mathbf{z}_{1:N} | \beta)$ önsel dağılımı, piksel etiket vektörlerinin uzamsal bağımlılığını tanımladığı için, piksellerin komşuluk bilgisi de modele dahil edilmektedir.

İzgesel vektörlerin koşullu bağımsız, etiket vektörlerinin ise bağımlı olduğu kabul edildiğinden, diğer pikseller verildiğinde bir pikselin spektral vektörü ile etiket vektörünün bileşik olasılığı şöyle tanımlanmaktadır:

$$p(\mathbf{s}_n, \mathbf{z}_n | \mathbf{s}_{\bar{n}}, \mathbf{z}_{\bar{n}}, \theta_{1:K}, \beta) = p(\mathbf{s}_n | \mathbf{z}_n, \theta_{1:K}) p(\mathbf{z}_n | \mathbf{z}_{\bar{n}}, \beta) \quad (3.11)$$

Burada $\bar{n}$, n piksel indisinin tümleyen kümesidir ve $\bar{n} = \{1, 2, \dots, N\} \setminus \{n\}$ olarak tanımlanmaktadır.

Uzamsal modelin Markov özelliğine sahip olduğu, dolayısıyla piksellerin yalnızca bir pencere içindeki komşu piksellere bağımlı olduğu kabul edilmektedir. Buna göre, (3.11)'deki eşitlik, n . pikselin komşuları kümesi $\tilde{n}$ ile gösterildiğinde, şöyle yazılabilir:

$$p(\mathbf{s}_n, \mathbf{z}_n | \mathbf{s}_{\tilde{n}}, \mathbf{z}_{\tilde{n}}, \theta_{1:K}, \beta) = p(\mathbf{s}_n | \mathbf{z}_n, \theta_{1:K}) p(\mathbf{z}_n | \mathbf{z}_{\tilde{n}}, \beta) \quad (3.12)$$

Bu bileşik olasılık dağılımı marjinalize edilirse, şu eşitlik yazılabilir:

$$\begin{aligned} p(\mathbf{s}_n | \mathbf{z}_{\tilde{n}}, \theta_{1:K}, \beta) &= \sum_{\mathbf{z}_n} \prod_{k=1}^K [p(\mathbf{s}_n | z_{n,k} = 1, \theta_k) \omega_{n,k}]^{z_{n,k}} \\ &= \sum_{k=1}^K p(\mathbf{s}_n | z_{n,k}, \theta_k) \omega_{n,k} \end{aligned} \quad (3.13)$$

Bu eşitlik, denklem (3.8)'de tanımlanmış olan karışım modeline benzediği için bu modelin de bir karışım modeli olduğu görülmektedir. Burada $\omega_{n,k}$ karışım katsayılarını göstermekte olup şöyle yazılabilir:

$$\omega_{n,k} = p(z_{n,k} = 1 | z_{\tilde{n},k}, \beta) \quad (3.14)$$

Bu karışım katsayıları, denklem (3.4)'te tanımlanmış olan karışım katsayılarından farklı olarak, komşu piksellerin sınıf etiketlerine de bağlı olduğundan, piksellerin uzamsal ilişkisi de dikkate alınmakta ve her piksel için değişmektedir. Karışım katsayılarının bu özelliğinden dolayı, bu model *uzamsal olarak değişen* veya *durağan olmayan karışım modeli* olarak tanımlanmaktadır (Kayabol ve Kutluk, 2016).

Özlojistik bağlanım modelinde etiket değişkenleri bağımlı kabul edildiğinden, bir piksel etiketinin komşu piksellere bağlı dağılımı şöyle tanımlanmaktadır (Besag, 1974):

$$p(z_{n,k} | z_{\tilde{n},k}, \beta) \propto e^{\beta(z_{n,k} + z_{n,k} \sum_{m \in \tilde{n}} z_{m,k})} \quad (3.15)$$

Tüm piksellerin etiket vektörlerinin bileşik olasılık dağılımı şöyle yazılabilir:

$$\begin{aligned} p(\mathbf{z}_{1:N} | \beta) &= \frac{\prod_{n=1}^N \prod_{k=1}^K e^{\beta z_{n,k} (1 + \frac{1}{2} \sum_{m \in \tilde{n}} z_{m,k})}}{\mathcal{Z}(\beta)} \\ &= \frac{\prod_{k=1}^K e^{\beta \sum_{n=1}^N z_{n,k} (1 + \frac{1}{2} \sum_{m \in \tilde{n}} z_{m,k})}}{\mathcal{Z}(\beta)} \end{aligned} \quad (3.16)$$

Burada $\mathcal{Z}(\beta)$ normalizasyon terimidir.

Bir etiket vektörünün diğer piksellerin etiket vektörlerine bağlı koşullu dağılımı şöyle yazılabilir:

$$p(\mathbf{z}_n | \mathbf{z}_{\tilde{n}}, \beta) = \frac{p(\mathbf{z}_{1:N} | \beta)}{\sum_{\mathbf{z}_n \in \mathcal{Z}} p(\mathbf{z}_{1:N}, \beta)} = \frac{\prod_{k=1}^K (e^{\beta v_{n,k}})^{z_{n,k}}}{\sum_{j=1}^K e^{\beta v_{n,j}}} \quad (3.17)$$

Burada $v_{n,k} = 1 + \sum_{m \in \tilde{n}} z_{m,k}$ olarak yazılabilir ve bu değişken n . pikseli çevreleyen pencere içinde k . sınıftan olan piksellerin sayısını vermektedir.

Bunlara göre, sınıfların ağırlık katsayıları şöyle yazılabilir:

$$\omega_{n,k} = p(z_{n,k} = 1 | z_{\tilde{n},k}, \beta) = \frac{e^{\beta v_{n,k}}}{\sum_{j=1}^K e^{\beta v_{n,j}}} \quad (3.18)$$

Eğitim kümesi kullanılarak model parametreleri kestirildikten sonra, bir test pikselinin sınıflandırılması için sınıf etiketlerinin sonsal olasılığı en büyüklenmektedir. Bir sınıf etiketinin koşullu olasılığı şöyle yazılabilir:

$$p(\mathbf{z}_n | \mathbf{z}_{\bar{n}}, \mathbf{s}_B, \hat{\theta}_{1:K}, \beta) = \frac{p(\mathbf{s}_n | \mathbf{z}_n, \hat{\theta}_{1:K})}{p(\mathbf{s}_n | \hat{\theta}_{1:K}, \beta)} p(\mathbf{z}_n | \mathbf{z}_{\bar{n}}, \beta) \quad (3.19)$$

Burada $\hat{\theta} = \{\hat{\mathbf{m}}_{1:K}, \hat{\Sigma}_{1:K}\}$ ve B test piksellerinin indisleri kümesini göstermektedir. Piksel etiketlerinin sonsal dağılımı birlikte en büyüklenemediğinden, yinelemeli koşullu mod algoritması (Besag, 1974) kullanılmaktadır. Bu algoritma ile, yinelemeli olarak $p(\mathbf{s}_n | \mathbf{z}_n, \hat{\theta}_{1:K})p(\mathbf{z}_n | \mathbf{z}_{\bar{n}}, \beta)$ çarpımı en büyüklenmektedir. Buna göre, yinelemeli olarak sınıf etiketi $\mathbf{z}_n$ ve β parametresi, t güncel iterasyon sayısını göstermek üzere, şöyle güncellenebilir:

$$\begin{aligned} \mathbf{z}_n^t &\leftarrow \underset{\mathbf{z}_n}{\operatorname{argmax}} p(\mathbf{s}_n | \mathbf{z}_n, \theta_{1:K}^{t-1}) p(\mathbf{z}_n | \mathbf{z}_{\bar{n}}^{t-1}, \beta^{t-1}) \\ \beta^t &\leftarrow \underset{\beta}{\operatorname{argmax}} \prod_{n=1}^N p(\mathbf{z}_n^t | \mathbf{z}_{\bar{n}}^t, \beta) \end{aligned} \quad (3.20)$$

Sınıf etiket önselinin parametresi β 'nın kestirimi denklem (3.16)'da paydada bulunan normalizasyon terimi nedeniyle zordur. Kayabol ve Kutluk (2016), bunun için Besag (1974) tarafından önerilen sözde olabilirlik fonksiyonunu kullanmıştır. Buna göre, bağımsızlık kabulü yapılarak, denklem (3.16)'da verilen olabilirlik yerine şu sözde olabilirlik yazılabilir:

$$p_{PL}(\beta | \mathbf{z}_{1:N}) = \prod_{n=1}^N \frac{\prod_{k=1}^K (e^{\beta v_{n,k}})^{z_{n,k}}}{\sum_{j=1}^K e^{\beta v_{n,j}}} \quad (3.21)$$

Bu sözde olabilirliğin en büyüklenmesi için, sözde log-olabilirlik fonksiyonu şöyle yazılabilir:

$$Q(\beta | \mathbf{z}_{1:N}^t) = \frac{\beta}{2} \sum_{n=1}^N \sum_{k=1}^K z_{n,k} (1 + v_{n,k}) - \sum_{n=1}^N \ln \sum_{j=1}^K e^{\frac{\beta}{2}(1+v_{n,j})} \quad (3.22)$$

Bu denklem β için kapalı formda çözümü olmadığından yinelemeli olarak optimize edilebilir ve buna göre güncelleme kuralı şöyle yazılabilir:

$$\beta^{t+1} = \beta^t - \eta \frac{Q'(\beta^t)}{Q''(\beta^t)} \quad (3.23)$$

Burada Q' ve Q'' sırayla Q fonksiyonunun birinci ve ikinci türevleri, η ise öğrenme katsayısıdır.

Buraya kadar tanımlanmış olan uzamsal olarak değişen karışım modelinde, sınıf-koşullu olasılıklar için bir dağılım belirlenmemiştir. Takip eden alt başlıklarda, sınıf-koşullu olasılıkların Student-t ve Gauss dağılımlarıyla modellenmesi ve parametreler üzerinde tanımlanan önsel dağılımlarla parametre kestirim yöntemleri açıklanmaktadır.

3.1.1. Uzamsal Olarak Değişen Student-t Karışım Modeli

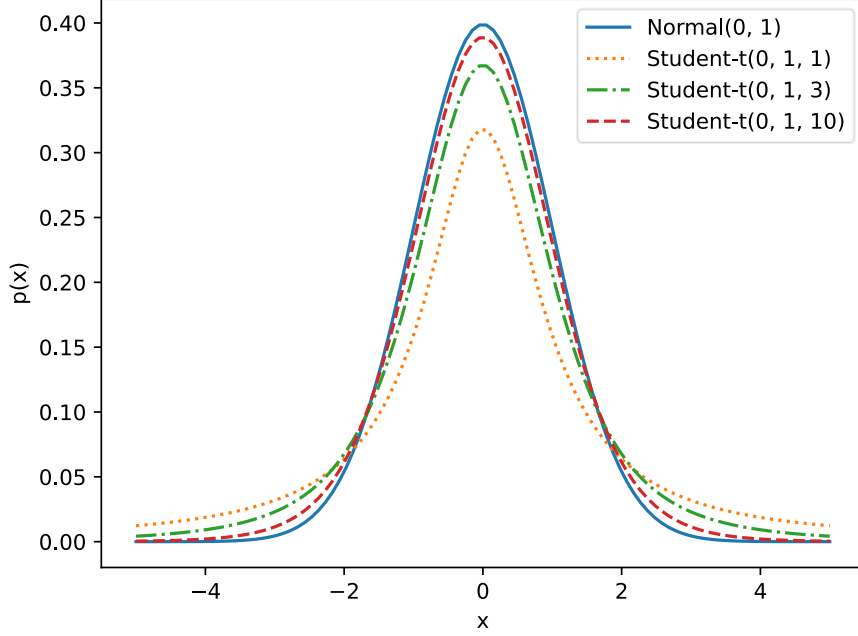
Kayabol (2015), hiperspektral imgelerin izge-uzamsal sınıflandırılması için bir üretici model önermiştir. Bu modelde her sınıf bir Gauss dağılımı ile modellenmekte, sınıf etiketleri üzerine tanımlanan uzamsal önsel dağılımlarla da piksel komşuluklarından yararlanılarak daha iyi sınıflandırma sonuçları elde edilmiştir. Ortalama ve kovaryans matrisleri için önsel dağılımlar tanımlanmış, böylece az sayıda eğitim örneği ile parametre kestirimi de mümkün hâle gelmiştir. Bu çalışmada önerilen karışım modeli, karışım katsayılarının her piksel için konuma bağlı olarak değişkenlik göstermesi nedeniyle, uzamsal olarak değişen ya da durağan olmayan bir karışım modeli olarak adlandırılmaktadır.

Tez kapsamında yapılan ilk çalışmada, Kayabol (2015) çalışması temel alınarak alınarak, her sınıfın çok değişkenli Gauss yerine çok değişkenli Student-t dağılımı ile modellenmesiyle benzer bir karışım modeli oluşturulmuştur. Bu bölümde bu model ve parametre kestirim yöntemleri açıklanmaktadır.

Student-t dağılımı, normal dağılıma göre daha kalın kuyruklu bir simetrik dağılım olup, bir sonsuz Gauss karışım modeli olarak düşünülebilir (Bishop, 2006). Normal dağılıma göre daha kalın kuyruklu bir dağılım olduğundan, Student-t dağılımı ile tasarlanan sınıflandırıcılar verideki aykırı örneklerden daha az etkilenmekte, dolayısıyla gürbüz algoritmalar elde edilebilmektedir (Bishop, 2006).

Student-t dağılımı, konum, ölçek ve özgürlük derecesi parametreleri ile tanımlanmaktadır. Özgürlük derecesi arttıkça, aynı konum ve ölçek (ortalama ve standart sapma) değerlerine sahip normal dağılıma yaklaşmaktadır. Şekil 3.1'de bir X rastgele değişkeninin tek boyutlu standart

normal dağılıma sahip olması durumu ile özgürlük derecesi 1, 3 ve 10 olan tek boyutlu Student-t dağılımına sahip olması durumundaki olasılık yoğunluk fonksiyonları görülmektedir.



Şekil 3.1: Farklı özgürlük derecesi değerleri için Student-t dağılımı ile normal dağılım olasılık yoğunluk fonksiyonları. Normal dağılımın ortalaması 0, standart sapması 1, Student-t dağılımlarının ise lokasyon parametreleri 0 ve ölçek parametreleri 1 olup özgürlük dereceleri 1, 3 ve 10 olarak seçilmiştir.

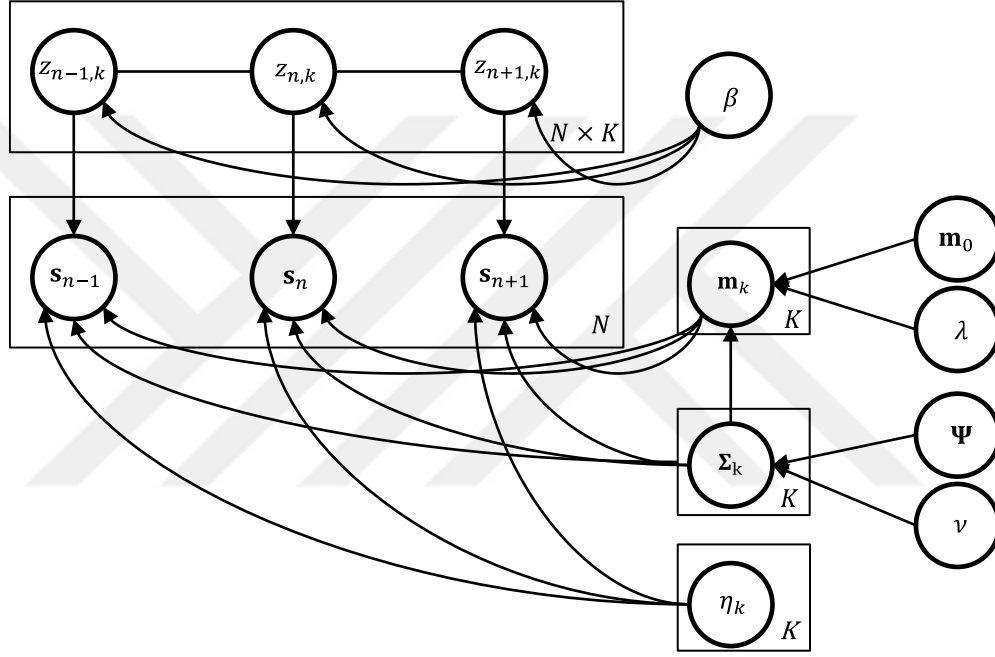
Student-t dağılımının parametrelerinin kestirilmesi için beklenti en büyükleme (EM) algoritması kullanılmıştır. Dempster ve diğ. (1977) tarafından önerilmiş olan bu algoritma, gizli değişken içeren istatistiksel modellerin ML veya MAP kestiriminin analitik çözümünün zor olduğu durumlarda, bu kestirimin yapılabilmesini sağlayan genel ve yinelemeli bir algoritmadır (Bishop, 2006; McLachlan ve Krishnan, 2008). Gözlenen verinin $\mathbf{X}$, model parametrelerinin θ ve gizli değişkenlerin $\mathbf{Z}$ ile gösterildiği durumda, $p(\mathbf{X}|\theta)$ olabilirlik fonksiyonunun en büyüklenmesi zor olabilmekte, bu durumda $p(\mathbf{X}, \mathbf{Z}|\theta)$ şeklinde ve tam olabilirlik fonksiyonu olarak adlandırılan bir fonksiyonu en büyüklenmesi mümkün olabilmektedir. Örneğin, logaritma fonksiyonu içinde toplam veya integrasyon olması durumunda, ML ya da MAP yöntemlerinin analitik çözümü bulunamamaktadır. Bu durumda EM algoritması kullanılarak, gizli değişkenlerin modele dahil edilmesiyle, problem bir alt sınır optimizasyonuna dönüştürülebilmektedir (Bishop, 2006).

EM algoritması, yinelemeli olarak sırayla uygulanan bir beklenti ve bir en büyükleme adımından oluşmaktadır. Beklenti adımında gizli değişkenlerin sonsal olasılığı hesaplanmakta

ve bu sonsal dağılıma göre tam log-olabilirlik fonksiyonunun beklenen değeri hesaplanmaktadır. En büyükleme adımında ise bu beklenen değeri en büyükleyen parametre değerleri hesaplanmaktadır.

EM algoritması, model parametreleri için önsel dağılımların tanımlanması durumunda MAP kestirimi için de kullanılabilmektedir (Bishop, 2006).

Şekil 3.2’de önerilen izge-uzamsal üretici modele ilişkin grafik model görülmektedir.



Şekil 3.2: Sınıf-koşullu olasılıkların Student-t dağılımı seçilmesi durumunda izge-uzamsal üretici model.

Sınıf-koşullu dağılımların Student-t dağılımı ile modellenmesi durumunda, bir pikselin sınıf-koşullu olasılık dağılımı şöyle tanımlanmaktadır:

$$p(\mathbf{s}_n | z_{n,k} = 1, \mathbf{m}_k, \Sigma_k, \eta_k) = \int_0^\infty \mathcal{N}(\mathbf{s}_n | \mathbf{m}_k, \frac{\Sigma_k}{\tau_n}) \text{Gamma}(\tau_n | \frac{\eta_k}{2}, \frac{\eta_k}{2}) d\tau_n \quad (3.24)$$

Burada τ_n gizli değişken olarak düşünülerek EM algoritması uygulanabilir ve buna göre tam olabilirlik fonksiyonu şöyle yazılabilir:

$$\begin{aligned}
& \prod_{n=1}^{N_k} p(\mathbf{s}_n, \tau_n | z_{n,k} = 1, \mathbf{m}_k, \mathbf{\Sigma}_k, \eta_k) \\
&= \prod_{n=1}^{N_k} \mathcal{N}(\mathbf{s}_n | \mathbf{m}_k, \frac{\mathbf{\Sigma}_k}{\tau_n}) \text{Gamma}(\tau_n | \frac{\eta_k}{2}, \frac{\eta_k}{2})
\end{aligned} \tag{3.25}$$

Gamma dağılımı, normal dağılım için eşlenik önsel olduğundan, gizli değişkenin sonsal dağılımı da Gamma dağılımıdır:

$$\begin{aligned}
& \tau_n | \mathbf{s}_n, z_{n,k} \\
&= 1, \mathbf{m}_k, \mathbf{\Sigma}_k, \eta_k \sim \text{Gamma}\left(\tau_n \middle| \frac{L + \eta_k}{2}, \frac{(\mathbf{s}_n - \mathbf{m}_k)^T \mathbf{\Sigma}_k^{-1} (\mathbf{s}_n - \mathbf{m}_k) + \eta_k}{2}\right)
\end{aligned} \tag{3.26}$$

En büyükleme adımında optimize edilmesi gereken fonksiyon, bu sonsal dağılım ile şöyle yazılabilir:

$$\begin{aligned}
Q(\theta, \theta^{t-1}) &= \int_{\tau_n} (p(\tau_n | \mathbf{s}_n, z_{n,k} = 1, \mathbf{m}_k^{t-1}, \mathbf{\Sigma}_k^{t-1}, \eta_k^{t-1}) \\
&\quad \times \ln p(\mathbf{s}_n, \tau_n | z_{n,k} = 1, \mathbf{m}_k, \mathbf{\Sigma}_k, \eta_k))
\end{aligned} \tag{3.27}$$

Burada θ^{t-1} bir önceki adımda kestirilen parametre kümesini göstermektedir.

Bu çalışmada modelin ortalama ve kovaryans matrisi parametreleri için normal-ters Wishart önsel dağılımları tanımlanmıştır. Buna göre, ortalama ve kovaryans matrisinin bileşik önsel dağılımı şöyle yazılabilir:

$$\begin{aligned}
p(\mathbf{m}_k, \mathbf{\Sigma}_k) &= \mathcal{N}\left(\mathbf{m}_k \middle| \mathbf{m}_0, \frac{1}{\lambda} \mathbf{\Sigma}_k\right) W^{-1}(\mathbf{\Sigma}_k | \mathbf{\Psi}, \nu) \\
&= \frac{1}{(2\pi)^{L/2} \left|\frac{\mathbf{\Sigma}_k}{\lambda}\right|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{m}_k - \mathbf{m}_0)^T \left(\frac{\mathbf{\Sigma}_k}{\lambda}\right)^{-1} (\mathbf{m}_k \right. \\
&\quad \left. - \mathbf{m}_0)\right\} \frac{|\mathbf{\Psi}|^{\frac{\nu}{2}}}{2^{\frac{\nu L}{2}} \Gamma\left(\frac{\nu}{2}\right)} |\mathbf{\Sigma}_k|^{-\frac{\nu+L+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{\Psi} \mathbf{\Sigma}_k^{-1})\right\}
\end{aligned} \tag{3.28}$$

Burada $\Gamma()$ simgesi ile gamma fonksiyonu gösterilmektedir.

Önsel dağılımla birlikte optimize edilmesi gereken fonksiyon $Q(\theta, \theta^{t-1}) + \ln p(\mathbf{m}_k, \mathbf{\Sigma}_k)$ olarak yazılabilir (Bishop, 2006).

Buna göre, $\mathbb{E}[\]$ sembolü beklenen değer olmak üzere, ortalama vektörünün kestirimi şöyle elde edilebilir:

$$\hat{\mathbf{m}}_k = \frac{\lambda \mathbf{m}_0 + \sum_{n=1}^N \mathbb{E}[\tau_n] \mathbf{s}_n}{\lambda + \sum_{n=1}^N \mathbb{E}[\tau_n]} \quad (3.29)$$

Benzer şekilde, kovaryans matrisinin kestirimi şöyle elde edilebilir:

$$\hat{\mathbf{\Sigma}}_k = \frac{\lambda(\mathbf{m}_k - \mathbf{m}_0)(\mathbf{m}_k - \mathbf{m}_0)^T + \Psi^T + \sum_{n=1}^N \mathbb{E}[\tau_n](\mathbf{s}_n - \mathbf{m}_k)(\mathbf{s}_n - \mathbf{m}_k)^T}{N_k + \eta_k + L + 2} \quad (3.30)$$

τ_n 'in beklenen değeri ise şöyle hesaplanır:

$$\mathbb{E}[\tau_n] = \frac{L + \eta_0}{\eta_0 + (\mathbf{s}_n - \mathbf{m}_0)^T \mathbf{\Sigma}_k^{-1} (\mathbf{s}_n - \mathbf{m}_k)} \quad (3.31)$$

η parametresi için kapalı formda çözüm olmadığı için, yapılan testlerde deneysel olarak belirlenmiştir.

EM algoritması ile parametre kestirim adımları EK-1'de verilmektedir

3.1.2. Uzamsal Olarak Değişen Gauss Karışım Modeli

İzgesel vektörlerin Gauss dağılımlarıyla modellenmesi ile bir uzamsal olarak değişen Gauss karışım modeli elde edilmiştir. Bu modelde her sınıf bir çokterimli normal dağılım ile modellendiğinden, sınıf-kışullu olasılıklar şöyle tanımlanmaktadır:

$$p(\mathbf{s}_n | z_{n,k} = 1, \mathbf{m}_k, \mathbf{\Sigma}_k) = \mathcal{N}(\mathbf{s}_n | \mathbf{m}_k, \mathbf{\Sigma}_k) \quad (3.32)$$

Burada $\mathbf{m}_k$ ve $\mathbf{\Sigma}_k$ sırayla k . sınıfın ortalama vektörü ve kovaryans matrisidir. Her sınıf bir Gauss dağılımı ile modellendiğinden, hiperspektral imgenin sınıf sayısı kadar Gauss dağılımının karışımı şeklinde modellendiği düşünülebilir.

Kayabol (2015) tarafından önerilen Bayesçi Gauss karışım modelinde, az örnekle eğitim yapılabilmesi için ortalama vektörü ve kovaryans matrisi üzerinde normal-ters Wishart önsel dağılımı tanımlanmıştır. Bu önsel dağılım şöyle gösterilmektedir:

$$p(\mathbf{m}_k, \mathbf{\Sigma}_k) = \mathcal{N}\left(\mathbf{m}_k \middle| \mathbf{m}_0, \frac{1}{\lambda} \mathbf{\Sigma}_k\right) W^{-1}(\mathbf{\Sigma}_k | \Psi, \nu) \quad (3.33)$$

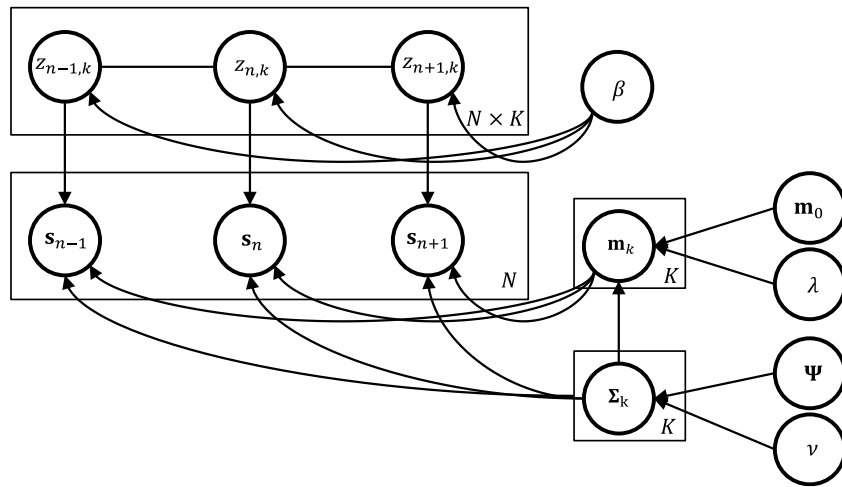
Normal-ters Wishart önsel dağılımı normal dağılım için eşlenik önsel dağılım olduğundan, ortalama vektörü ve kovaryans matrisinin sonsal dağılımı da normal-ters Wishart dağılımına sahip olmaktadır, dolayısıyla sonsal dağılım analitik olarak bulunabilmektedir (Gelman ve diğ., 2014).

Tezde önerilen Gauss karışım modelinde ise, normal-ters Wishart önsel dağılımı için bir parametre daha tanımlanarak, kovaryans matrisi için daha iyi bir kestirim yöntemi geliştirilmiştir (Kayabol ve Kutluk, 2016). Bu önsel dağılım şöyledir:

$$p(\mathbf{m}_k, \mathbf{\Sigma}_k) = \mathcal{N}\left(\mathbf{m}_k \middle| \mathbf{m}_0, \frac{1}{\lambda} \mathbf{\Sigma}_k\right) W^{-1}(\mathbf{\Sigma}_k | (\tau + L + 1)\Psi, \tau) \quad (3.34)$$

Burada $\mathbf{m}_0$, Ψ ve τ önsel dağılım parametreleridir. Denklem (3.33) ve (3.34)'te verilen dağılımların ikisi de normal-ters Wishart olsa da, (3.34)'te Ψ parametresi için farklı bir ağırlıklandırma uygulanmıştır (Kayabol ve Kutluk, 2016).

Şekil 3.3'te önerilen izge-uzamsal modele ilişkin grafik model görülmektedir.



Şekil 3.3: Sınıf-koşullu olasılıkların Gauss dağılımı seçilmesi durumunda izge-uzamsal üretici model.

Önsel dağılım tanımlandığından, ortalama vektörü ve kovaryans matrisinin kestirimi için MAP yöntemi kullanılmaktadır. k . sınıfın eğitim piksellerinin indisleri kümesi D_k ile gösterildiğinde bu parametrelerin önsel dağılımları şöyle yazılabilir:

$$p(\mathbf{m}_k | \mathbf{s}_{D_k}, \mathbf{\Sigma}_k) = \frac{p(\mathbf{s}_{D_k} | \mathbf{m}_k, \mathbf{\Sigma}_k) p(\mathbf{m}_k, \mathbf{\Sigma}_k)}{\int p(\mathbf{s}_{D_k} | \mathbf{m}'_k, \mathbf{\Sigma}_k) p(\mathbf{m}'_k, \mathbf{\Sigma}_k) d\mathbf{m}'_k} \quad (3.35)$$

$$p(\mathbf{\Sigma}_k | \mathbf{s}_{D_k}, \mathbf{m}_k) = \frac{p(\mathbf{s}_{D_k} | \mathbf{m}_k, \mathbf{\Sigma}_k) p(\mathbf{m}_k, \mathbf{\Sigma}_k)}{\int p(\mathbf{s}_{D_k} | \mathbf{m}_k, \mathbf{\Sigma}'_k) p(\mathbf{m}_k, \mathbf{\Sigma}'_k) d\mathbf{\Sigma}'_k} \quad (3.36)$$

Bu sonsal dağılımlar analitik olarak hesaplandığında şu denklemler elde edilmektedir:

$$\mathbf{m}_k | \mathbf{s}_{D_k}, \mathbf{\Sigma}_k \sim \mathcal{N} \left(\frac{N_k \bar{\mathbf{s}}_k + \lambda \mathbf{m}_0}{N_k + \lambda}, \frac{1}{N_k + \lambda} \mathbf{\Sigma}_k \right) \quad (3.37)$$

$$\mathbf{\Sigma}_k | \mathbf{s}_{D_k} \sim W^{-1}(\mathbf{P}_k, N_k + \tau) \quad (3.38)$$

Burada $N_k = |D_k|$ olup k . sınıftaki piksel sayısını göstermektedir. $\mathbf{P}_k$ ise şöyle hesaplanmaktadır:

$$\mathbf{P}_k = (\tau + L + 1)\Psi + N_k \mathbf{S}_k + \frac{N_k \lambda}{N_k + \lambda} (\bar{\mathbf{s}}_k - \mathbf{m}_0)(\bar{\mathbf{s}}_k - \mathbf{m}_0)^T \quad (3.39)$$

$\bar{\mathbf{s}}_k$ örnek ortalama vektörünü, $\mathbf{S}_k$ ise örnek kovaryans matrisini göstermekte olup bunlar şöyle hesaplanabilir:

$$\bar{\mathbf{s}}_k = \frac{1}{N_k} \sum_{n \in D_k} \mathbf{s}_n \quad (3.40)$$

$$\mathbf{S}_k = \frac{1}{N_k} \sum_{n \in D_k} (\mathbf{s}_n - \bar{\mathbf{s}}_k)(\mathbf{s}_n - \bar{\mathbf{s}}_k)^T \quad (3.41)$$

Ortalama vektörü ve kovaryans matrisinin sonsal dağılımlarını en büyükleyen kestirimler şöyle yazılabilir:

$$\hat{\mathbf{m}}_k = \frac{N_k \bar{\mathbf{s}}_k + \lambda \mathbf{m}_0}{N_k + \lambda} \quad (3.42)$$

$$\hat{\mathbf{\Sigma}}_k = \frac{1}{N_k + \tau + L + 1} \mathbf{P}_k \quad (3.43)$$

Kayabol ve Kutluk (2016) şu şekilde bir α parametresi tanımlamıştır:

$$\alpha = \frac{\tau + L + 1}{N_k + \tau + L + 1}, \quad 0 \leq \alpha \leq 1 \quad (3.44)$$

Bu parametre ile kovaryans matrisinin kestirimi şöyle yazılabilir:

$$\hat{\mathbf{\Sigma}}_k = \alpha \Psi + (1 - \alpha) \mathbf{S}_k + \frac{N_k \lambda (\bar{\mathbf{s}}_k - \mathbf{m}_0)(\bar{\mathbf{s}}_k - \mathbf{m}_0)^T}{(N_k + \lambda)(N_k + \tau + L + 1)} \quad (3.45)$$

Tanımlanan α parametresi kullanılarak τ parametresi de şöyle yazılabilir:

$$\tau = \frac{\alpha(N_k + L + 1) - L - 1}{1 - \alpha}, \quad 0 \leq \alpha \leq 1 \quad (3.46)$$

Denklem (3.44)'te verilmiş olan α parametresi ile örnek kovaryans matrisi $\mathbf{S}_k$ ile önsel kovaryans matris Ψ için bir ağırlıklandırma yapılmaktadır. Eğitim kümesinde az sayıda piksel olması durumunda örnek kovaryans matrisinin tersi alınamayacağından, α parametresi ile önsel kovaryans matrisine daha çok ağırlık verilerek kovaryans matrisinin kestirimi mümkün olmaktadır (Kayabol ve Kutluk, 2016).

Bu modelin hiperparametreleri $\mathbf{m}_0$, Ψ , α ve λ olmak üzere dört tanedir. $\mathbf{m}_0$ parametresi eğitim kümesinin ortalaması olarak seçilmiştir. Ψ ise eğitim kümesinin örnek kovaryans matrisine bağlı olarak $\Psi = 0,75\mathbf{S} + 0,25\text{diag}\{\mathbf{S}\}$ şeklinde seçilmiştir. Burada D eğitim kümesindeki örneklerin indisleri kümesi olmak üzere, örnek kovaryans matrisi şöyle hesaplanabilir:

$$\mathbf{S} = \frac{1}{N} \sum_{n \in D} (\mathbf{s}_n - \mathbf{m}_0)(\mathbf{s}_n - \mathbf{m}_0)^T \quad (3.47)$$

3.2. ÜRETİCİ ALT-UZAY YÖNTEMİYLE İZGE-UZAMSAL SINIFLANDIRMA

Bu modelde olasılıksal temel bileşen analizi (Tipping ve Bishop, 1999) ile her sınıfın bir normal dağılım ile modellenmesi, aynı zamanda spektral vektör boyutlarının indirgenmesi amaçlanmaktadır.

Hiperspektral imgelerde izgesel bant sayısı oldukça büyüktür. Buna karşılık etiketli veri elde etmek yüksek maliyetli olduğundan, literatürde kullanılan veri kümelerindeki piksel sayısı genellikle küçüktür. Bu da model parametrelerinin kestirilmesini zorlaştırmaktadır. Bu nedenle modelin eğitiminden önce boyut indirgeme yöntemleriyle spektral vektörler için bir alt-uzay temsili elde edilmesi ile parametre kestiriminin yapılabilmesi, model karmaşıklığının düşürülmesi ve bazı durumlarda daha iyi sınıflandırma başarımı elde edilmesi mümkündür.

Boyut indirgeme için en sık kullanılan yöntemlerden biri temel bileşen analizidir (Hotelling, 1933; Pearson, 1901). Bu yöntem veri için temel bileşenlerin bulunmasını ve değişintiyi en büyükleyecek eksenler seçilerek verinin bu uzaya izdüşürülmesini sağlayan bir ortogonal dönüşümdür.

Olasılıksal temel bileşen analizi, tüm marjinal ve koşullu dağılımların Gauss olduğu bir doğrusal Gauss modelidir (Bishop, 2006). Bu modelde temel bileşen alt-uzayında bir gizli değişken tanımlanmakta, doğrusal Gauss modeliyle de gözlemin bu gizli değişkenin bir doğrusal fonksiyonu ile hatanın toplamından oluştuğu varsayılmaktadır. Bu doğrusal model şöyle yazılabilir:

$$\mathbf{s}_n = \mathbf{W}_k \mathbf{t}_n + \mathbf{m}_k + \boldsymbol{\epsilon}_n \quad (3.48)$$

Burada $\mathbf{s}_n$ bir piksele ilişkin spektral imzayı, $\mathbf{W}_k$ gözlem ile gizli değişken arasındaki ilişki matrisini, $\mathbf{t}_n$ bu piksele ilişkin gizli değişkeni, $\mathbf{m}_k$ ortalamayı, $\boldsymbol{\epsilon}_n$ ise hatayı göstermektedir. Gizli değişken $\mathbf{t}_n$, $q < L$ olmak üzere, q boyutlu bir vektördür. Dolayısıyla $\mathbf{W}_k$ matrisi de $L \times q$ boyutludur.

Gizli değişken $\mathbf{t}_n$ 'in sıfır ortalamalı ve birim değişintili bir Gauss dağılımına sahip olduğu, hatanın da sıfır ortalamalı $\sigma^2 \mathbf{I}_q$ kovaryans matrisine sahip olduğu kabul edilmektedir:

$$\begin{aligned} \mathbf{t}_n &\sim \mathcal{N}(0, \mathbf{I}_q) \\ \boldsymbol{\epsilon}_n &\sim \mathcal{N}(0, \sigma^2 \mathbf{I}_q) \end{aligned} \quad (3.49)$$

Dolayısıyla, $p(\mathbf{s}_n | \mathbf{t}_n)$ dağılımı da bir Gauss dağılımıdır:

$$p(\mathbf{s}_n | \mathbf{t}_n) = \mathcal{N}(\mathbf{W}_k \mathbf{t}_n + \mathbf{m}_k, \sigma^2 \mathbf{I}_L) \quad (3.50)$$

Piksel vektörü $\mathbf{s}_n$ 'in marjinal yoğunluğu da bir normal dağılım olup şöyle yazılabilir:

$$p(\mathbf{s}_n) = \int p(\mathbf{s}_n|\mathbf{t}_n)p(\mathbf{t}_n)d\mathbf{t}_n = \mathcal{N}(\mathbf{m}_k, \mathbf{\Sigma}_k) \quad (3.51)$$

Piksel vektörlerinin dağılımı da Gauss olduğu için, sınıf etiketleri üzerinde tanımlanan önsel dağılımla birlikte elde edilen durağan olmayan karışım modeli bir Gauss karışım modeli olmaktadır.

Tipping ve Bishop (1999) kovaryans matrisi $\mathbf{\Sigma}_k$ 'nın şöyle hesaplanabileceğini göstermiştir:

$$\mathbf{\Sigma}_k = \mathbf{W}_k \mathbf{W}_k^T + \sigma_k^2 \mathbf{I}_L \quad (3.52)$$

Log-olabilirlik fonksiyonu şöyle yazılabilir:

$$\mathcal{L} = -\frac{N_k}{2} \{d \ln(2\pi) + \ln|\mathbf{\Sigma}_k| + \text{tr}(\mathbf{\Sigma}_k^{-1} \mathbf{S}_k)\} \quad (3.53)$$

Burada N_k k . sınıftaki piksel sayısı, $\mathbf{S}_k$ ise örnek kovaryans matrisidir ve şöyle hesaplanabilir:

$$\mathbf{S}_k = -\frac{1}{N_k} \sum_{n=1}^{N_k} (\mathbf{s}_n - \hat{\mathbf{m}}_k)(\mathbf{s}_n - \hat{\mathbf{m}}_k)^T \quad (3.54)$$

Ortalamanın ML kestirimi şöyle hesaplanabilir:

$$\hat{\mathbf{m}}_k = \frac{1}{N_k} \sum_{n=1}^{N_k} \mathbf{s}_n \quad (3.55)$$

Tipping ve Bishop (1999) log olabilirlik fonksiyonunu en büyükleyen $\mathbf{W}_k$ matrisinin şöyle bulunabileceğini göstermiştir:

$$\hat{\mathbf{W}}_k = \mathbf{U}_q (\mathbf{\Lambda}_q - \sigma_k^2 \mathbf{I})^{1/2} \mathbf{R} \quad (3.56)$$

Burada $\mathbf{U}_q$ $L \times q$ boyutlu bir matris olup sütunlarında $\mathbf{S}_k$ kovaryans matrisinin özvektörleri bulunmaktadır. $\mathbf{\Lambda}_q$ matrisi ise köşegen bir kare matris olup, köşegen üzerinde bu özvektörlere ilişkin özdeğerler $\lambda_{k,1}, \dots, \lambda_{k,q}$ bulunmaktadır. $\mathbf{R}$ matrisi $q \times q$ boyutlu gelişigüzel bir ortogonal dönme matrisi olup bu çalışmada birim matris olarak seçilmiştir.

Tipping ve Bishop (1999) tarafından önerilen yöntemden farklı olarak, bu çalışmada kullanılan modelde varyans için bir önsel dağılım tanımlanmış ve MAP kestirimi uygulanmıştır. Tanımlanan önsel dağılım ters-Gamma dağılımı olup şöyle gösterilebilir:

$$\sigma_k^2 \sim \text{Ters} - \text{Gamma}(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma_k^2)^{-\alpha-1} \exp\left(-\frac{\beta}{\sigma_k^2}\right) \quad (3.57)$$

Bu önsel dağılımla birlikte, varyansın MAP kestirimi şöyle hesaplanabilir:

$$\sigma_k^2 = \frac{N_k \sum_{i=q+1}^L \lambda_{k,i} + 2\beta}{N_k(L - q) + 2(\alpha + 1)} \quad (3.58)$$

Bölüm 3.1’de açıklanan durağan olmayan karışım modellerinde olduğu gibi, bu modelde de piksel etiketleri üzerinde tanımlanan çokterimli özlojistik bağlanım önselleri ile bir üretici model elde edilmektedir. Dolayısıyla bu yöntemde alt-uzay temsilleriyle bir durağan olmayan Gauss karışım modeli elde edilmekte ve bu modelle izge-uzamsal sınıflandırma yapılabilmektedir.

3.3. ÇOKTERİMLİ LOJİSTİK BAĞLANIM MODELLERİYLE İZGE-UZAMSAL SINIFLANDIRMA

Lojistik bağlanım modeli, bir ayımsayıcı ve olasılıksal sınıflandırma modelidir. Bu yöntemde sınıf etiket değişkeninin sonsal olasılığı giriş değişkeninin doğrusal fonksiyonları ile modellenir (Hastie ve diğ., 2009). Etiket değişkeninin alabileceği değerlerin, yani sınıf sayısının ikiden fazla olduğu durum için çokterimli lojistik bağlanım modeli tanımlanmıştır. Bu model, girişteki veriye bağlı olarak çıkış değişkeninin alabileceği değerlerin normalize olasılık değerlerini verdiğinden bir olasılıksal sınıflandırıcı olarak tanımlanmaktadır.

Çokterimli lojistik bağlanım modeli şöyle ifade edilebilir:

$$p(z_{n,k} = 1 | \mathbf{s}_n, \omega_k) = \frac{\exp(\omega_k^T \mathbf{s}_n)}{\sum_{j=1}^K \exp(\omega_j^T \mathbf{s}_n)} \quad (3.59)$$

Kayabol (2020) sınıf-koşullu olasılıkları çokterimli lojistik bağlanım modelleri ile modellemiş, sınıf etiketlerinin de kategorik dağılıma uyduğunu kabul ederek bir üretici model tanımlamıştır.

Bu model, sınıf etiketleri üzerinde başka önsel dağılımların tanımlanmasını ve böylece uzamsal sınıflandırma da yapılabilmesini mümkün kılmıştır.

Tezde çokterimli lojistik bağlanım modeli izge-uzamsal sınıflandırma amacıyla iki ayrı şekilde kullanılmış olup, bunlardan biri üretici, diğeri ise ayırmsayıcı bir modeldir.

3.3.1. Üretici Model

Bu bölümde, izgesel vektörlerin çokterimli lojistik bağlanım ile modellenmesi ile elde edilen üretici model tanıtılmaktadır.

Kayabol (2020) bir üretici model önermiş olup sınıf-koşullu olasılıkları çokterimli lojistik bağlanım modelleri şeklinde modellemiş, sınıf etiketleri üzerinde de kategorik önsel dağılım tanımlamıştır. Bu durumda imgenin normalize edilmiş üstel dağılımların karışımı şeklinde modellendiği düşünülebilir. EK-2’de gösterildiği gibi, sınıf ağırlıklarının eşit olması durumunda, bu model ayırmsayıcı bir model olan çok terimli lojistik bağlanım modeline eşdeğerdir. Sınıf-koşullu olasılıkların aynı üstel dağılıma sahip olması durumunda elde edilen modelin lojistik bağlanım modeline eşit olduğu gösterilmiştir (Banerjee, 2007; Jordan, 1995). EK-2’de ise, sınıf-koşullu olasılıkların normalize üstel dağılıma sahip olması durumunda da lojistik bağlanım modelinin elde edildiği gösterilmiştir.

Bu çalışmada piksel etiket değişkenleri üzerinde çokterimli özlojistik bağlanım önseli tanımlanmakta, bu nedenle Kayabol (2020)’den farklı olarak piksellerin uzamsal olarak bağımlı olduğu kabul edilmekte, dolayısıyla izge-uzamsal sınıflandırma yapılmaktadır.

Bu modelde, K adet sınıflı bir imge için, etiket vektörünün $z_n \in \{0,1\}^K$ şeklinde tanımlandığı ve $\sum_{k=1}^K z_{n,k} = 1$ koşulu sağladığı kabul edilirse, $n = 1, \dots, N$ olmak üzere, $\mathbf{s}_n$ piksel vektörleri ile $\mathbf{z}_n$ etiket vektörlerinin bileşik olasılık dağılımı şöyle tanımlanmaktadır:

$$p(\mathbf{s}_{1:N}, \mathbf{z}_{1:N} | \boldsymbol{\omega}_{1:K}, \beta) = \left[\prod_{n=1}^N \prod_{k=1}^K p(\mathbf{s}_n | \boldsymbol{\omega}_k)^{z_{n,k}} \right] p(\mathbf{z}_{1:N} | \beta) \quad (3.60)$$

Sınıf-koşullu olasılıklar normalize edilmiş üstel fonksiyonlar ile şöyle tanımlanmaktadır:

$$p(\mathbf{s}_n | z_{n,k} = 1, \boldsymbol{\omega}_{1:K}) = \frac{e^{\boldsymbol{\omega}_k^T \mathbf{s}_n}}{\sum_{j=1}^K e^{\boldsymbol{\omega}_j^T \mathbf{s}_n}} \quad (3.61)$$

Bu ifade piksel etiketi $\mathbf{z}_n$ vektörü ile gösterildiğinde şöyle yazılabilir:

$$p(\mathbf{s}_n | \mathbf{z}_n, \boldsymbol{\omega}_{1:K}) = \prod_{k=1}^K \left(\frac{e^{\boldsymbol{\omega}_k^T \mathbf{s}_n}}{\sum_{j=1}^K e^{\boldsymbol{\omega}_j^T \mathbf{s}_n}} \right)^{z_{n,k}} \quad (3.62)$$

Modeldeki izgesel bağlanım parametrelerinin kestirimi ML kestirimi yöntemiyle yapılmaktadır.

$\boldsymbol{\omega} = [\boldsymbol{\omega}_1^T, \dots, \boldsymbol{\omega}_K^T]^T$ olmak üzere, bağlanım parametrelerinin log-olabilirlik fonksiyonu şöyle yazılabilir:

$$L(\boldsymbol{\omega}) = \sum_{n=1}^N \left[\sum_{k=1}^K z_{n,k} \boldsymbol{\omega}_k^T \mathbf{s}_n - \ln \sum_{j=1}^K \exp(\boldsymbol{\omega}_j^T \mathbf{s}_n) \right] \quad (3.63)$$

Bu log-olabilirlik fonksiyonunun ikinci derece Taylor serisi açılımı şöyle yazılabilir:

$$\begin{aligned} L(\boldsymbol{\omega}) &\approx L(\boldsymbol{\omega}^{(t)}) + (\boldsymbol{\omega} - \boldsymbol{\omega}^{(t)})^T \nabla L(\boldsymbol{\omega}^{(t)}) \\ &\quad + \frac{1}{2} (\boldsymbol{\omega} - \boldsymbol{\omega}^{(t)})^T \nabla^2 L(\boldsymbol{\omega}^{(t)}) (\boldsymbol{\omega} - \boldsymbol{\omega}^{(t)}) \end{aligned} \quad (3.64)$$

Burada $\nabla L(\boldsymbol{\omega}^{(t)})$ log-olabilirlik fonksiyonunun $\boldsymbol{\omega}^{(t)}$ için hesaplanan gradyan vektörünü, $\nabla^2 L(\boldsymbol{\omega}^{(t)})$ ise Hessian matrisini göstermektedir.

Çokterimli özlojistik bağlanım modelinin log-olabilirlik fonksiyonu içbükey olduğundan birinci derece Taylor serisi açılımı ile sınırlıdır. Bu durumda, Hessian matrisi için bir $\mathbf{B}$ alt sınırı tanımlanabilirse, optimizasyon bir alt sınır optimizasyonu problemine dönüştürülebilir (Krishnapuram, 2005). Taylor serisi açılımı ile bu alt sınır şöyle yazılabilir:

$$L(\boldsymbol{\omega}) \geq L(\boldsymbol{\omega}^{(t)}) + (\boldsymbol{\omega} - \boldsymbol{\omega}^{(t)})^T \nabla L(\boldsymbol{\omega}^{(t)}) + \frac{1}{2} (\boldsymbol{\omega} - \boldsymbol{\omega}^{(t)})^T \mathbf{B} (\boldsymbol{\omega} - \boldsymbol{\omega}^{(t)}) \quad (3.65)$$

Böhning (1992), çokterimli lojistik bağlanım modelinin log-olabilirlik fonksiyonu için Hessian matrisinin alt sınırlı olduğunu göstermiştir. Bu alt sınır matrisi $KL \times KL$ boyutlu olup, her $L \times L$ boyutlu bölümü şöyle hesaplanabilir (Kayabol, 2020):

$$\mathbf{B}_{kj} = -\frac{1}{2} \left(\delta_{kj} - \frac{1}{K} \right) \mathbf{S}^T \mathbf{S} \quad (3.66)$$

Burada $k = 1, \dots, K, j = 1, \dots, K$ ve $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_N]^T$ olup, δ_{kj} Kronecker delta fonksiyonunu göstermektedir.

Bu alt sınırla bağlanım parametrelerinin yinelemeli güncelleme kuralı şöyle yazılabilir:

$$\boldsymbol{\omega}^{(t+1)} = \boldsymbol{\omega}^{(t)} - \mathbf{B}^{-1} \nabla L(\boldsymbol{\omega}^{(t)}) \quad (3.67)$$

Bu güncelleme kuralındaki matris tersi işlemi yüksek maliyetli olduğundan, Kayabol (2020) güncellemenin bileşenlere ayrı ayrı uygulanabileceğini göstermiştir. Bir sınıfa ilişkin güncelleme kuralı şöyle yazılabilir:

$$\boldsymbol{\omega}_k^{(t+1)} = \boldsymbol{\omega}_k^{(t)} - \mathbf{B}_{kk}^{-1} \left[\mathbf{g}_k(\boldsymbol{\omega}_k^{(t)}) + \frac{1}{2} \sum_{j \neq k} \mathbf{B}_{kj} (\boldsymbol{\omega}_j^{(t)} - \boldsymbol{\omega}_j^{(t-1)}) \right] \quad (3.68)$$

Burada log-olabilirlik fonksiyonunun k . sınıfın bağlanım parametre vektörü $\boldsymbol{\omega}_k$ 'ya göre gradyan vektörü şöyle hesaplanabilir:

$$\mathbf{g}_k(\boldsymbol{\omega}_k^{(t)}) = \sum_{n=1}^N \left(z_{n,k} - \frac{e^{(\boldsymbol{\omega}_k^{(t)})^T \mathbf{s}_n}}{\sum_{j=1}^K e^{(\boldsymbol{\omega}_j^{(t)})^T \mathbf{s}_n}} \right) \mathbf{s}_n \quad (3.69)$$

Yeni bir pikselin sınıflandırılması için, kestirilen bağlanım parametreleri $\hat{\boldsymbol{\omega}}_{1:K}$ kullanılarak, pikselin spektral imzası $\mathbf{s}_n$ verildiğinde, $\mathbf{z}_n$ etiket vektörünün sonsal olasılığı en büyüklenmektedir. Sınıf etiketleri üzerinde tanımlanan çokterimli özlojistik bağlanım önseli ile, $\mathbf{s}_B$ test pikselleri kümesini göstermek üzere, bu sonsal olasılık şöyle yazılabilir:

$$\begin{aligned}
p(\mathbf{z}_n | \mathbf{z}_{\bar{n}}, \mathbf{s}_B, \hat{\boldsymbol{\omega}}_{1:K}, \beta) &\propto p(\mathbf{s}_n | \mathbf{z}_n, \hat{\boldsymbol{\omega}}_{1:K}) p(\mathbf{z}_n | \mathbf{z}_{\bar{n}}, \beta) \\
&= \prod_{k=1}^K \left[\frac{e^{\hat{\boldsymbol{\omega}}_k^T \mathbf{s}_n} e^{\beta v_{n,k}}}{\sum_{j=1}^K e^{\hat{\boldsymbol{\omega}}_j^T \mathbf{s}_n} \sum_{j=1}^K e^{\beta v_{n,j}}} \right]^{z_{n,k}}
\end{aligned} \tag{3.70}$$

Burada $\bar{n} = \{1, 2, \dots, N\} \setminus \{n\}$, $\tilde{n}$ n . pikselin komşu piksel indisleri ve $v_{n,k} = 1 + \sum_{m \in \tilde{n}} z_{m,k}$. Sınıf etiketlerinin bileşik olasılığı en büyüklenemediğinden, yinelemeli koşullu mod algoritması (Besag, 1974) kullanılmaktadır.

3.3.2. Ayırmsayıcı Model

Bölüm 2.2.4'te olasılıksal sınıflandırma modellerinin üretici ve ayırmsayıcı olarak ikiye ayrılabilirliği belirtilmişti. Bu bölümde izge-uzamsal sınıflandırma için geliştirilen ayırmsayıcı model tanıtılmaktadır.

Ayırmsayıcı modellerde, üretici modellerden farklı olarak sınıf-koşullu olasılıklar ve sınıf etiket önsel dağılımları tanımlanmaz, bunun yerine, $p(\mathbf{z}_n | \mathbf{s}_n, \theta)$ koşullu dağılımı doğrudan modellenir ve bu dağılımı en büyükleyen sınıf etiketi seçilerek sınıflandırma yapılır (Bishop ve Lasserre, 2007).

Önerilen ayırmsayıcı modelde, imgedeki tüm piksellerin sınıf etiketlerinin bileşik koşullu olasılık dağılımı şöyle tanımlanmaktadır:

$$\begin{aligned}
p(\mathbf{z}_{1:N} | \mathbf{s}_{1:N}, \boldsymbol{\omega}_{1:K}, \boldsymbol{\beta}_{1:K}) \\
&= \prod_{n=1}^N \prod_{k=1}^K \left(\frac{e^{\boldsymbol{\omega}_k^T \mathbf{s}_n + \boldsymbol{\beta}_k^T \mathbf{m}_{n,k}}}{\sum_{j=1}^K e^{\boldsymbol{\omega}_j^T \mathbf{s}_n + \boldsymbol{\beta}_j^T \mathbf{m}_{n,j}}} \right)^{z_{n,k}} \\
&= \prod_{n=1}^N \frac{\prod_{k=1}^K e^{z_{n,k} (\boldsymbol{\omega}_k^T \mathbf{s}_n + \boldsymbol{\beta}_k^T \mathbf{m}_{n,k})}}{\sum_{j=1}^K e^{\boldsymbol{\omega}_j^T \mathbf{s}_n + \boldsymbol{\beta}_j^T \mathbf{m}_{n,j}}}
\end{aligned} \tag{3.71}$$

Burada $\boldsymbol{\omega}_k$ k . sınıfa ilişkin spektral katsayı vektörünü, $\boldsymbol{\beta}_k$ k . sınıfa ilişkin uzamsal katsayı vektörünü, $\mathbf{m}_{n,k}$ ise n . pikselin merkezde olduğu bir pencere içindeki piksellerin $z_{n,k}$ etiket değerlerinden oluşan vektörü göstermektedir ve $\boldsymbol{\beta}_k$ ile $\mathbf{m}_{n,k}$ etiketlerin ve katsayıların aynı şekilde sıralanmasıyla elde edilmektedir.

Bu modelin eğitimi için, Bölüm 3.3.1'deki gibi yaklaşık ikinci derece optimizasyon yöntemi uygulanmaktadır. Modelin log-olabilirlik fonksiyonunun hem spektral bağlanım parametreleri ω_k 'ya göre hem de uzamsal bağlanım parametreleri β_k 'ya göre gradyan vektörü ve Hessian matrisi aynı biçimde olacağından hem spektral hem uzamsal parametrelerin kestirimi için aynı alt sınır optimizasyon yöntemi kullanılabilir.

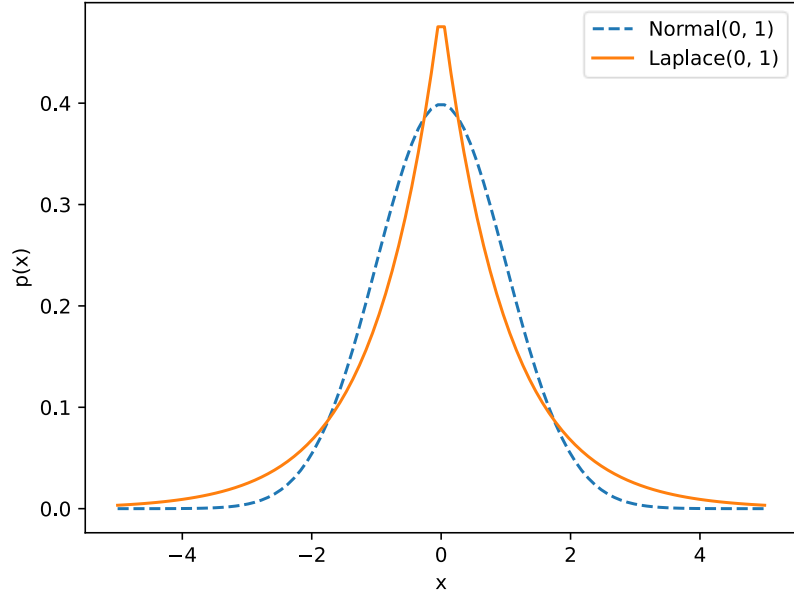
Spektral bağlanım parametrelerinin seyrekleştirilmesi amacıyla, Kayabol (2020) ile Krishnapuram ve diğ. (2005) tarafından önerildiği gibi, bu parametreler üzerinde Laplace önsel dağılımı tanımlanmıştır. Bu önsel dağılım şöyle ifade edilebilir:

$$p(\omega_{1:K}|\lambda) = \prod_{k=1}^K \frac{\lambda}{2} \exp(-\lambda \|\omega_k\|_1) \quad (3.72)$$

Burada $\|\omega_k\|_1$ L_1 normunu göstermekte olup şöyle hesaplanabilir:

$$\|\omega_k\|_1 = \sum_{l=1}^L |\omega_{k,l}| \quad (3.73)$$

Laplace önsel dağılımı, parametrelerin sıfır civarında yoğunlaşmasını sağlayarak seyrekleştirilmesini sağlamaktadır. Parametreler üzerinde Laplace önsel dağılımı tanımlanması, L_1 düzenleme yönteminin Bayesçi yaklaşımdaki karşılığıdır (Barber, 2012). Tek boyutlu normal dağılım ile Laplace dağılımının olasılık yoğunluk fonksiyonları Şekil 3.4'te görülmektedir. Burada normal dağılımın ortalaması 0, standart sapması ise 1 iken, Laplace dağılımının konum parametresi 0, ölçek parametresi ise 1'dir. Bu şekilde görüldüğü gibi, Laplace dağılımına sahip bir değişkenin sıfır değerini alma olasılığı normal dağılıma göre daha yüksektir, dolayısıyla Laplace önseli tanımlanan parametreler seyrekleşmektedir.



Şekil 3.4: Normal dağılım ve Laplace dağılımının olasılık yoğunluk fonksiyonları.

Tanımlanan önsel dağılımla birlikte, $\boldsymbol{\omega} = [\boldsymbol{\omega}_1^T, \dots, \boldsymbol{\omega}_K^T]^T$ olmak üzere, spektral bağlanım parametrelerinin log-sonsal fonksiyonu şöyle yazılabilir:

$$L(\boldsymbol{\omega}) = \sum_{n=1}^N \left[\sum_{k=1}^K z_{n,k} \boldsymbol{\omega}_k^T \mathbf{s}_n - \ln \sum_{j=1}^K \exp(\boldsymbol{\omega}_j^T \mathbf{s}_n) \right] - \lambda \sum_{k=1}^K \|\boldsymbol{\omega}_k\|_1 \quad (3.74)$$

Buradaki olabilirlik fonksiyonu ve Laplace dağılımı log-içbükey olduğundan, log-sonsal da bir içbükey fonksiyon olup tek bir maksimum değere sahiptir (Krishnapuram ve diğ., 2005). Dolayısıyla log-sonsal fonksiyon da birinci derece Taylor serisi ile sınırlıdır ve Bölüm 3.3.1’de açıklandığı gibi, Hessian matrisinin de sınırlı olması durumunda alt sınır optimizasyon yöntemi uygulanabilmektedir.

L_1 normunun sıfırda türevi tanımsız olduğundan, Kayabol (2020) bir alt-gradyan çözümü uygulamış ve $\nabla L(\boldsymbol{\omega}^{(t)})$ gradyan vektörü yerine bir $\mathbf{g}_L(\boldsymbol{\omega}^{(t)})$ alt-gradyan terimi tanımlamıştır. Bu terim log-olabilirlik ve log-önsel gradyanlarının toplamı şeklinde yazılabilir:

$$\mathbf{g}_L(\boldsymbol{\omega}^{(t)}) = \mathbf{g}_l(\boldsymbol{\omega}^{(t)}) + \mathbf{g}_p(\boldsymbol{\omega}^{(t)}) \quad (3.75)$$

Log-önselin gradyanı sıfırda tanımsız olduğundan, signum fonksiyonu alt-gradyan olarak kullanılabilir (Kayabol, 2020):

$$\mathbf{g}_p(\boldsymbol{\omega}^{(t)}) = -\lambda \text{sign}(\boldsymbol{\omega}^{(t)}) \quad (3.76)$$

Kayabol (2020) tarafından önerilen yöntemde, alt-gradyan yöntemiyle birlikte, Hessian matrisi için de $\nabla^2 L(\boldsymbol{\omega}^{(t)}) \geq \mathbf{H}_L(\boldsymbol{\omega}^{(t)})$ şeklinde bir alt sınır kullanılmaktadır ve bu alt sınır log-olabilirlik ve log-önsele ilişkin Hessian alt sınırlarının toplamı şeklinde yazılabilir:

$$\mathbf{H}_L(\boldsymbol{\omega}^{(t)}) = \mathbf{H}_l(\boldsymbol{\omega}^{(t)}) + \mathbf{H}_p(\boldsymbol{\omega}^{(t)}) \quad (3.77)$$

Burada $\mathbf{H}_l(\boldsymbol{\omega}^{(t)})$ log-olabilirlik fonksiyonunun, $\mathbf{H}_p(\boldsymbol{\omega}^{(t)})$ is log-önselin Hessian alt sınırını göstermektedir. $\mathbf{H}_l(\boldsymbol{\omega}^{(t)})$ alt sınırı, Bölüm 3.3.1’de açıklandığı gibi $KL \times KL$ boyutlu bir sabit matris olup, her $L \times L$ boyutlu bölümü şöyle hesaplanabilir (Kayabol, 2020):

$$\mathbf{B}_{kj} = -\frac{1}{2} \left(\delta_{kj} - \frac{1}{K} \right) \mathbf{S}^T \mathbf{S} \quad (3.78)$$

Kayabol (2020), log-önselin Hessian matrisi için de Krishnapuram ve diğ. (2005) tarafından önerilmiş olan, $\mathbf{H}_p(\boldsymbol{\omega}^{(t)}) = -\lambda \boldsymbol{\Lambda}(\boldsymbol{\omega}^{(t)})$ şeklinde bir yaklaşık hesaplama yöntemi uygulamıştır. $\boldsymbol{\Lambda}(\boldsymbol{\omega}^{(t)})$ bir köşegen matris olup, köşegen üzerindeki terimler $k = 1, \dots, K$ olmak üzere şöyle hesaplanabilir:

$$\boldsymbol{\Lambda}(\boldsymbol{\omega}_k) = \text{diag}\{|\omega_{k,1}|^{-1}, |\omega_{k,2}|^{-1}, \dots, |\omega_{k,L}|^{-1}\} \quad (3.79)$$

Bu yaklaşık hesaplamalarla, bağlanım parametrelerinin yinelemeli güncellenme kuralı şöyle yazılabilir:

$$\boldsymbol{\omega}^{(t+1)} = \boldsymbol{\omega}^{(t)} - \left(\mathbf{B} - \lambda \boldsymbol{\Lambda}(\boldsymbol{\omega}^{(t)}) \right)^{-1} \mathbf{g}_L(\boldsymbol{\omega}^{(t)}) \quad (3.80)$$

Bölüm 3.3.1’deki gibi, burada da güncelleme her sınıf için ayrı ayrı yapılabilir. Burada k . sınıfın bağlanım parametre vektörü $\boldsymbol{\omega}_k$ ’ya ilişkin sub-gradyan vektörü şöyle hesaplanabilir:

$$\mathbf{g}_k(\boldsymbol{\omega}_k^{(t)}) = \sum_{n=1}^N \left(z_{n,k} - \frac{e(\boldsymbol{\omega}_k^{(t)})^T \mathbf{s}_n}{\sum_{j=1}^K e(\boldsymbol{\omega}_j^{(t)})^T \mathbf{s}_n} \right) \mathbf{s}_n - \lambda \text{sign}(\boldsymbol{\omega}_k^{(t)}) \quad (3.81)$$

Buna göre, k . sınıfın bağlanım parametreleri için güncelleme kuralı şöyle yazılabilir:

$$\boldsymbol{\omega}_k^{(t+1)} = \boldsymbol{\omega}_k^{(t)} - [\mathbf{B}_{kk} - \lambda \Lambda(\boldsymbol{\omega}_k^{(t)})]^{-1} \left[\mathbf{g}_k(\boldsymbol{\omega}_k^{(t)}) + \sum_{j \neq k} \mathbf{B}_{kj}(\boldsymbol{\omega}_j^{(t)} - \boldsymbol{\omega}_j^{(t-1)}) \right] \quad (3.82)$$

Uzamsal bağlanım parametre vektörü $\boldsymbol{\beta}_k$ üzerinde önsel dağılım tanımlanmamaktadır. Dolayısıyla bu parametre vektörünün kestirimi için Bölüm 3.3.1’de açıklanan yaklaşık optimizasyon yöntemi uygulanabilmektedir.

Eğitim aşamasında izgesel ve uzamsal parametrelerin kestirimi yinelemeli olarak yapılabilmektedir. Her yineleme adımında sırayla izgesel ve uzamsal parametrelerin güncellenmesi veya önce uzamsal parametrelerin sabit tutularak izgesel parametrelerin kestirilmesi, daha sonra uzamsal parametrelerin kestirilmesi gibi farklı optimizasyon stratejileri uygulanabilmektedir.

Test piksellerinin sınıflandırılması için denklem (3.71)’de verilen $p(\mathbf{z}_{1:N} | \mathbf{s}_{1:N}, \boldsymbol{\omega}_{1:K}, \boldsymbol{\beta}_{1:K})$ koşullu olasılığın en büyüklenmesi gerekmektedir. Piksellerin uzamsal bağımlılığı nedeniyle bu bileşik olasılık en büyüklenememekte, bu nedenle yinelemeli koşullu mod algoritması (Besag, 1974) kullanılmakta ve yineleme adımında bir pikselin diğer pikseller verildiğinde koşullu olasılığı en büyüklenmektedir.

3.4. EVRİŞİMSSEL SİNİR AĞI İLE İZGE-UZAMSAL SINIFLANDIRMA

Evrışimsel sinir ağlarında veriden öznitelik çıkarımı ile sınıflandırıcı eğitimi işlemleri bir arada yapılmaktadır. Öznitelik çıkarımının da parametrik bir model ile yapılması ve sınıflandırıcı eğitimiyle birlikte bu parametrelerin de kestirilmesi ile, öznitelik çıkarımı ve sınıflandırma algoritmalarının ayrı uygulanması durumuna göre daha iyi başarımlar elde edilebilmektedir.

Krizhevsky ve diğ. (2012) tarafından önerilen AlexNet adlı ağ mimarisiyle imge sınıflandırma alanında elde edilen başarıdan sonra evrışimsel sinir ağları yapay öğrenmede, özellikle imge işleme problemlerinde çok yaygın olarak kullanılmaya başlanmış, farklı ağ mimarileriyle nesne tanıma, yüz tanıma gibi çok çeşitli problemlerin çözümünde bu yaklaşımdan yararlanılmıştır. 2015 yılından bu yana hiperspektral imge sınıflandırma probleminin çözümü içinde derin ağlar sıkça kullanılmaktadır.

Sınıflandırma için tasarlanan bir evrışimsel sinir ağı evrişim katmanları, aktivasyon fonksiyonları, havuz (pooling) fonksiyonları ve normalizasyon fonksiyonları ve bunları takip

eden en az bir doğrusal katman ve softmax fonksiyonundan oluşmaktadır. Ağın sonundaki doğrusal katman ve softmax fonksiyonu bir çokterimli lojistik bağlanım modeli oluşturmaktadır. Ağın yapısı, probleme uygun şekilde, değişen sayıda ve farklı bileşenlerden oluşan katmanlar ile tasarlanmaktadır.

Yapay sinir ağlarında ağın yapısının yanı sıra eğitim yöntemi de, giriş ve çıkış arasında çok katmanlı ve doğrusal olmayan, karmaşık bir ilişki bulunması nedeniyle büyük önem taşımaktadır. Bu ağlarda parametre kestirimi genellikle geri yayılım algoritması ile yapılmaktadır (Rumelhart ve diğ., 1986). Bu algoritmayla, ağ çıkışında hesaplanan amaç fonksiyonu ağ parametrelerine zincir türev kuralıyla yayılmaktadır ve parametreler bir optimizasyon algoritmasıyla güncellenmektedir. Yaygın olarak, birinci derece bir optimizasyon algoritması olan en dik iniş algoritmasının türevleri kullanılmaktadır. Çok sınıflı sınıflandırma problemlerinde, bu algoritmayla en küçüklenmeye çalışılan amaç fonksiyonu kategorik çapraz-entropi olarak adlandırılmaktadır.

Olasılıksal yaklaşımla düşünüldüğünde, kategorik çapraz-entropi amaç fonksiyonunun en küçüklenmesi, parametrelerin olabirlik fonksiyonunun en büyüklenmesiyle eşdeğerdir. Dolayısıyla, geri yayılım algoritmasıyla yapılan kestirim yöntemi ML kestirimidir. Parametreler üzerinde önsel olasılık dağılımlarının tanımlanması durumunda ise MAP kestirimi uygulanmaktadır.

Bu bölümde hiperspektral imgelerin izge-uzamsal sınıflandırılması için evrişimsel sinir ağlarının kullanımı ve ağ parametrelerinin kestirimi için önerilen yeni bir eğitim yöntemi tanıtılmaktadır. Önerilen eğitim yönteminde ağın sınıflandırma katmanı için Bölüm 3.3.2’de açıklanan yaklaşık seyrek ikinci derece parametre kestirim algoritması kullanılmakta, ağın önceki katmanları ise standart geri yayılım algoritması ve birinci derece optimizasyon algoritmalarıyla eğitilmektedir.

İzgesel vektörler üzerinde evrişim katmanlarıyla öznitelik vektörleri öğrenilmekte, çok sınıflı sınıflandırma problemlerinde kullanılan evrişimsel sinir ağlarının parametre kestirimi için, birinci ve ikinci derece optimizasyon yöntemlerinin birlikte kullanıldığı bir yöntem verilmektedir.

Bir çok sınıflı sınıflandırma probleminde, evrişimsel sinir ağı öznitelik çıkarma katmanları ve çokterimli lojistik bağlanım sınıflandırıcısı olarak düşünülebilir. Ağın sonundaki doğrusal

katman ve softmax fonksiyonu bir çokterimli lojistik bağlanım sınıflandırıcısı oluşturmaktadır. Önceki katmanlar ise evrişim fonksiyonları, doğrusal olmayan aktivasyon fonksiyonları ve doğrusal katmanlardan oluşan ve farklı yapılarda olabilen öznitelik çıkarma katmanları olarak düşünülebilir. Buna göre, bir evrişimsel sinir ağı şöyle ifade edilebilir:

$$p(\mathbf{z}_n | \mathbf{s}_n, \boldsymbol{\omega}_{1:K}) = \prod_{k=1}^K \left(\frac{e^{\boldsymbol{\omega}_k^T \boldsymbol{\phi}(\mathbf{s}_n)}}{\sum_{j=1}^K e^{\boldsymbol{\omega}_j^T \boldsymbol{\phi}(\mathbf{s}_n)}} \right)^{z_{n,k}} \quad (3.83)$$

Burada $\boldsymbol{\phi}(\mathbf{s}_n)$ fonksiyonu öznitelik çıkarma katmanlarını göstermektedir ve ağıın tasarımına göre farklı sayıda ve yapıda katmanlardan oluşmaktadır. Tezin bu bölümünde öznitelik çıkarma katmanlarının çıkışında $\mathbf{s}_n$ pikseline ilişkin elde edilen öznitelik vektörleri $\boldsymbol{\phi}(\mathbf{s}_n) = \boldsymbol{\phi}_n$ ile gösterilecektir.

Bu bölümde evrişimsel sinir ağlarının eğitimi için önerilen eğitim yönteminde Bölüm 3.3'te açıklanan üretici formülasyon ve yaklaşık parametre kestirim algoritması kullanılmaktadır. Buna göre denklem (3.83)'te tanımlanan yapay sinir ağı şöyle ifade edilebilir:

$$p(\boldsymbol{\phi}_n | \mathbf{z}_n, \boldsymbol{\omega}_{1:K}) = \prod_{k=1}^K \left(\frac{e^{\boldsymbol{\omega}_k^T \boldsymbol{\phi}_n}}{\sum_{j=1}^K e^{\boldsymbol{\omega}_j^T \boldsymbol{\phi}_n}} \right)^{z_{n,k}} \quad (3.84)$$

Önerilen eğitim yönteminde, son katman için Bölüm 3.3.2'de açıklanan yaklaşık ikinci derece optimizasyon algoritması kullanılmakta, öznitelik çıkarma katmanları ise geri yayılım algoritmasıyla eğitilmektedir. İkinci derece optimizasyon yöntemleri genellikle daha maliyetli olsa da bu yöntemle ikinci derece optimizasyon uygulanması nedeniyle gelen ek maliyet yalnızca son katmandan kaynaklanırken, ağıın yakınsama hızı artırılmıştır. Dolayısıyla bu ek maliyet ağıın derinliğinden bağımsız olmakta, sınıf sayısına ve çıkarılan özniteliklerin uzunluğuna bağlı olarak değişmektedir.

Bölüm 3.3'te çokterimli lojistik bağlanım modelinin Hessian matrisinin bir sabit matrisle alt sınırlı olduğu açıklanmıştır. Bu yöntem evrişimsel sinir ağlarının eğitimi için kullanıldığında, her iterasyonda evrişim katmanlarından gelen öznitelikler farklı olacağından, alt sınır matrisi de değişmektedir ve her iterasyonda tekrar hesaplanması gerekmektedir.

Sınıflandırma katmanının parametreleri üzerinde Laplace önsel dağılımı tanımlanmaktadır. Bu nedenle, bu parametrelerin kestiriminde Bölüm 3.3.2’de tanıtılan ayrımsayıcı modelin spektral bağlanım parametrelerinde de kullanılan yaklaşık ikinci derece optimizasyon algoritması kullanılmaktadır. Ancak, bu katmanın girişi her iterasyonda evrişim katmanlarının çıkışına göre değiştiğinden, Hessian matrisinin alt sınırının tekrar hesaplanması gerekmektedir.

Derin ağların eğitimi genellikle eğitim kümesi küçük parçalara (minibatch) bölünerek yapılmaktadır. Her yineleme adımında bir parça kullanılmakta, tüm parçaların tamamlanmasıyla eğitim kümesinin tamamının eğitimde kullanılmış olduğu bir tam döngüye devir (epoch) adı verilmektedir. Kayabol (2020) tarafından önerilen eğitim yönteminde eğitim kümesinin tamamı her iterasyonda kullanılmakta, dolayısıyla bir devir içinde yineleme yapılmamaktadır. Bu nedenle, evrişimsel sinir ağlarının eğitimi için önerilen yöntemde, parçalı eğitim uygulanması durumunda, denklem (3.66)’daki $\mathbf{S}^T \mathbf{S}$ yineleme adımında toplamsal olarak güncellenmekte ve buna bağlı olarak da alt sınır tekrar hesaplanmaktadır.

Önerilen eğitim algoritmasının adımları şöyledir:

```

Function ileri_yayilim():
    /* Verilen eğitim kümesi ile öznitelik katmanlarının ve ağın
    çıkışının hesaplanması */
End Function
Function maliyet_hesaplama():
    /* Ağın çıkışında çapraz-entropi maliyet fonksiyonunun
    hesaplanması */
End Function
Function oznitelik_katmanlari_guncelleme():
    /* Geri yayılım algoritmasıyla öznitelik çıkarma katmanlarının
    parametrelerinin güncellenmesi */
End Function
 $T_{max} :=$  devir sayısı
 $k = 1, \dots, K$  için  $\omega_k^{(0)}$  ve  $\mathbf{e}_k$  vektörlerini başlat
 $\lambda$  parametresini başlat
while  $t \leq T_{max}$  do
     $\mathbf{SS}$  matrisini başlat
    for each minibatch do
        ileri_yayilim()
         $\mathbf{SS} += \mathbf{S}_{mb}^T \mathbf{S}_{mb}$ 
         $B_{kj} = -\frac{1}{2} \left( \delta_{kj} - \frac{1}{K} \right) \mathbf{SS}$  her  $(k, j)$  için
    for  $k = 1:K$  do

```

```


$$\pi_{n,k} = \frac{e^{\left(\omega_k^{(t)}\right)^T \phi_n}}{\sum_{j=1}^K e^{\left(\omega_j^{(t)}\right)^T \phi_n}}$$


$$\mathbf{g}_k\left(\omega_k^{(t)}\right) = \sum_{n=1}^N \left(z_{n,k} - \pi_{n,k}\right) \phi_n - \lambda \text{sign}\left(\omega_k^{(t)}\right)$$

end
for  $k = 1:K$  do
     $\Lambda\left(\omega_k^{(t)}\right) = \text{diag}\left\{\left|\omega_{k,1}^{(t)}\right|^{-1}, \left|\omega_{k,2}^{(t)}\right|^{-1}, \dots, \left|\omega_{k,L}^{(t)}\right|^{-1}\right\}$ 
 $\mathbf{e}_k = -\left[\mathbf{B}_{kk} - \lambda \Lambda\left(\omega_k^{(t)}\right)\right]^{-1} \left[\mathbf{g}_k\left(\omega_k^{(t)}\right) + \sum_{j \neq k} \mathbf{B}_{kj} \mathbf{e}_j\right]$ 
 $\omega_k^{(t+1)} = \omega_k^{(t)} + \mathbf{e}_k$ 
end
ileri_yayilim()
maliyet_hesaplama()
oznitelik_katmanlari_guncelleme()
end
end

```

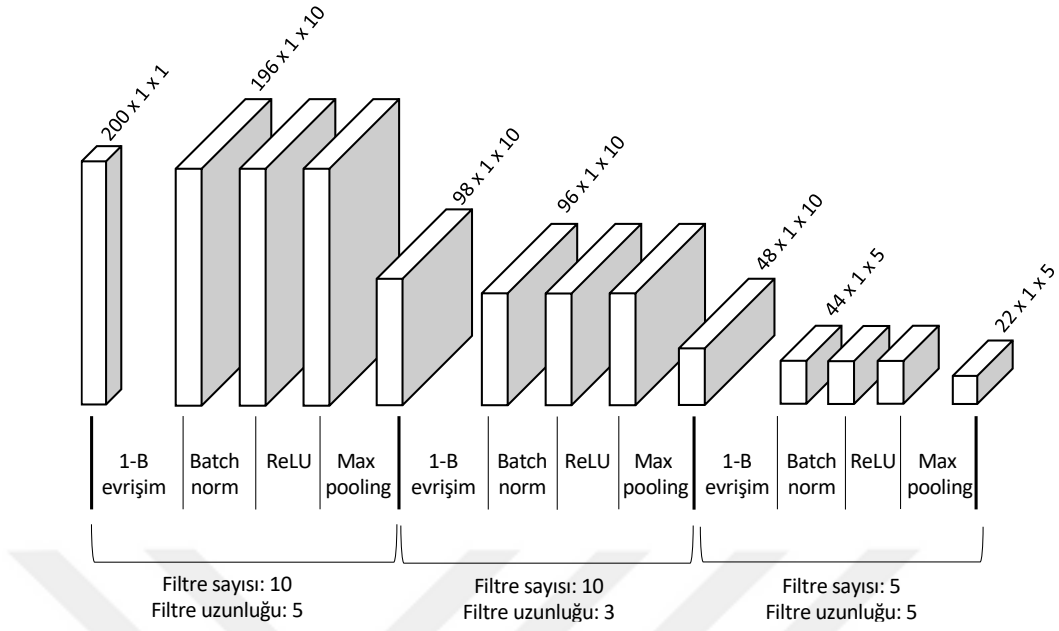
İzge-uzamsal sınıflandırma amacıyla, önerilen yöntemde sınıf etiketleri üzerinde çokterimli özlojistik bağlanım önselleri tanımlanmakta ve böylece uzamsal bilgiden de yararlanılarak daha iyi sınıflandırma haritaları elde edilmektedir.

3.4.1. Ağ Mimarisi

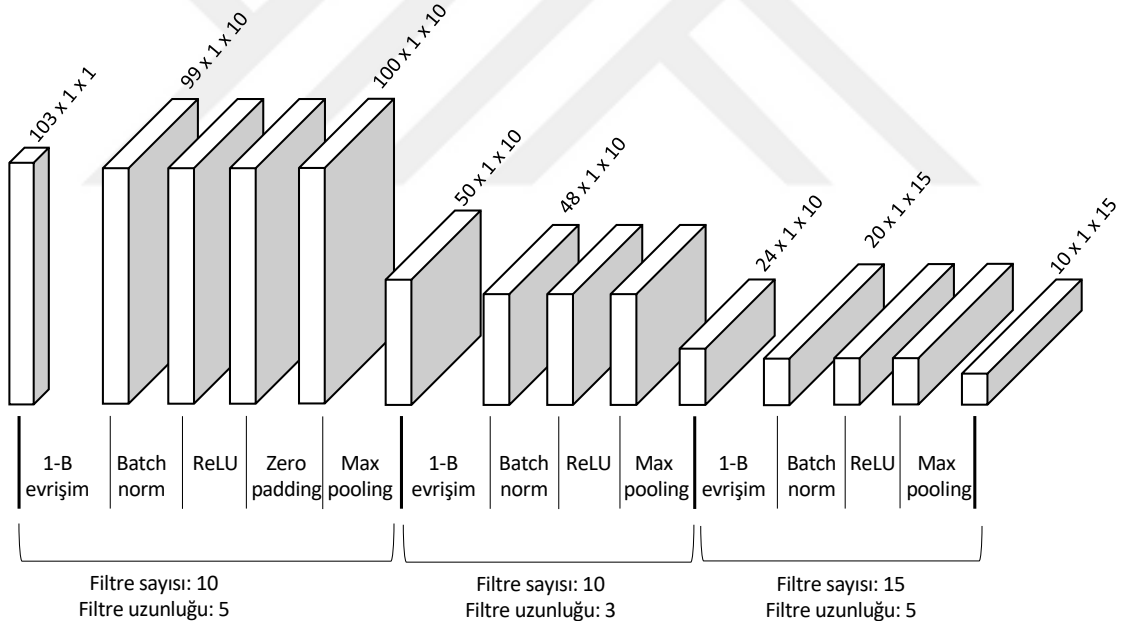
Evrişimsel sinir ağlarının tasarımında genellikle çok katmanlı bir yapıda evrişim, normalizasyon ve havuz işlemleri kullanılmaktadır.

Tezde kullanılan evrişimsel sinir ağlarında bir boyutlu evrişim işlemi uygulanmış, her piksel için spektral vektör ağın girişine uygulanmıştır. Aynı imge üzerinde hem eğitim hem de test yapıldığından, eğitim ve test kümeleri üzerinde bilgi geçişi olmaması için iki boyutlu ve üç boyutlu ağlar kullanılmamıştır. Ayrıca bir boyutlu evrişimsel ağlar kullanılarak ağın basit olması sağlanmış, uzamsal bilgi etiket değişkenleri üzerinde tanımlanan önsel dağılımlarla modele dahil edilmiştir.

Şekil 3.5 ve Şekil 3.6’da sırayla 200 ve 103 spektral bant sayısına sahip iki farklı imge için tasarlanmış ağlardaki evrişim katmanlarının yapısı görülmektedir. Bu katmanlar öznitelik çıkarımı katmanları olup, elde edilen özniteliklerin boyutları sırayla 110 ve 150’dir.



Şekil 3.5: Spektral bant sayısı 200 olan bir imge için tasarlanan ağdaki evrişim katmanlarının yapısı.



Şekil 3.6: Spektral bant sayısı 103 olan bir imge için tasarlanan ağdaki evrişim katmanlarının yapısı.

3.4.2. Sınıflandırma

Yapay sinir ağları ile sınıflandırma yapılması için test pikselinin ağın girişine verilmesi ile ileri yayılım algoritmasının çalıştırılması gerekmektedir. Ağın son katmanı bir çokterimli lojistik bağlantı sınıflandırıcısı olduğundan, verilen pikselin her sınıfa ait olma olasılığı ağ çıkışında hesaplanmaktadır ve sınıflandırma en büyük olasılığı veren sınıfın seçilmesiyle yapılmaktadır.

Önerilen yöntemde, hiperspektral imgelerin izge-uzamsal sınıflandırılması için piksel etiketleri üzerinde çokterimli özlojistik bağlanım önselleri tanımlanmaktadır. Etiket önseliyle birlikte sınıflandırma Bölüm 3.3.1’de açıklanan yöntem gibi, yinelemeli koşullu mod algoritmasıyla (Besag, 1974) piksel etiketlerinin sonsal olasılığı en büyüklenerek yapılmaktadır. Bu algoritmayla en büyüklenecek sonsal dağılım şöyle yazılabilir:

$$p(\mathbf{z}_n | \mathbf{z}_{\bar{n}}, \boldsymbol{\phi}_B, \hat{\boldsymbol{\omega}}_{1:K}, \beta) \propto p(\boldsymbol{\phi}_n | \mathbf{z}_n, \hat{\boldsymbol{\omega}}_{1:K}) p(\mathbf{z}_n | \mathbf{z}_{\bar{n}}, \beta)$$

$$= \prod_{k=1}^K \left(\frac{e^{\hat{\omega}_k^T \boldsymbol{\phi}_n} e^{\beta v_{n,k}}}{\sum_{j=1}^K e^{\hat{\omega}_j^T \boldsymbol{\phi}_n} \sum_{j=1}^K e^{\beta v_{n,j}}} \right)^{z_{n,k}} \quad (3.85)$$

Burada $\bar{n} = \{1, 2, \dots, N\} \setminus \{n\}$ ve $v_{n,k} = 1 + \sum_{m \in \bar{n}} z_{m,k}$ olup B test piksellerinin indisleri kümesi, $\hat{\boldsymbol{\omega}}_k$ ise k . sınıfa ilişkin kestirilmiş parametre vektörüdür.

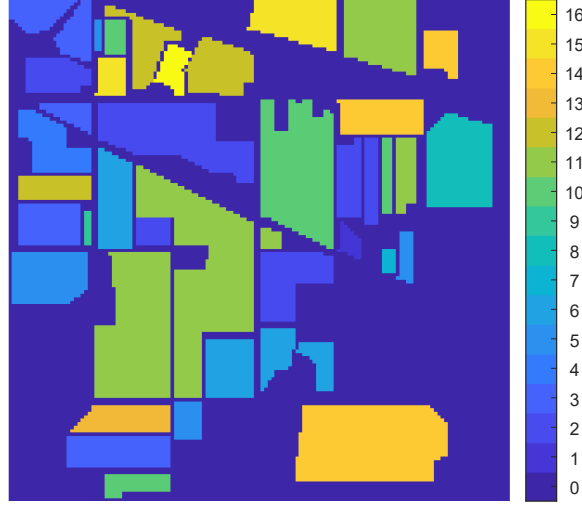
3.5. HİPERSPEKTRAL İMGE VERİ KÜMELERİ

Bu bölümde hiperspektral imgelerle yeryüzü bölgelerinin sınıflandırılması çalışmalarında kullanılan ve tez kapsamında geliştirilen yöntemlerin değerlendirilmesinde kullanılan imgeler tanıtılacaktır. Hiperspektral imge sınıflandırma literatüründe kullanılan veri kümeleri genellikle tek bir imgeden oluşmakta, dolayısıyla eğitim ve sınıflandırma testleri tek bir imge üzerinde yapılmaktadır.

3.5.1. Indian Pines İmgesi

Indian Pines imgesi AVIRIS sensörüyle elde edilmiş bir hiperspektral imgedir. Amerika Birleşik Devletleri’nin Indiana eyaletinde elde edilen bu imgede farklı tarım arazileri bulunmaktadır. İmgenin piksel sayısı 145×145 olup 224 spektral bant bulunmaktadır. Deneylerde 24 gürültülü bant hariç tutularak, 200 bant kullanılmıştır. Bu imgede 16 sınıf bulunmaktadır.

Indian Pines imgesinin gerçek sınıflandırma haritası Şekil 3.7’de görülmektedir.



Şekil 3.7: Indian Pines imgesinin gerçek sınıflandırma haritası.

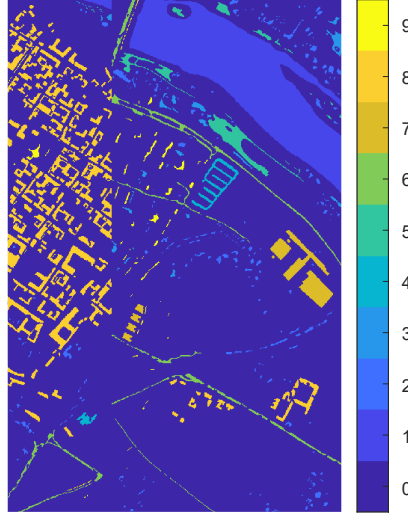
Indian Pines imgesinin sınıf isimleri ile her sınıftaki piksel sayısı Tablo 3.1’de görülmektedir.

Tablo 3.1: Indian Pines imgesindeki sınıflar ve piksel sayıları.

Sınıf no	Sınıf	Piksel sayısı
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-mintill	830
4	Corn	237
5	Grass-pasture	483
6	Grass-trees	730
7	Grass-pasture-mowed	28
8	Hay-windrowed	478
9	Oats	20
10	Soybean-notill	972
11	Soybean-mintill	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Buildings-Grass-Trees-Drives	386
16	Stone-Steel-Towers	93

3.5.2. Pavia Centre İmgesi

Pavia Centre imgesi İtalya’nın Pavia şehrinde elde edilmiş bir hiperspektral imgedir. Bu imgede 102 spektral bant, 9 sınıf ve 1096×715 piksel bulunmaktadır. İşaretlenen pikseller çeşitli yapı malzemeleri ve bitki türlerinden oluşan sınıflara aittir. Pavia Centre imgesinin gerçek sınıflandırma haritası Şekil 3.8’de görülmektedir.



Şekil 3.8: Pavia Centre imgesinin gerçek sınıflandırma haritası.

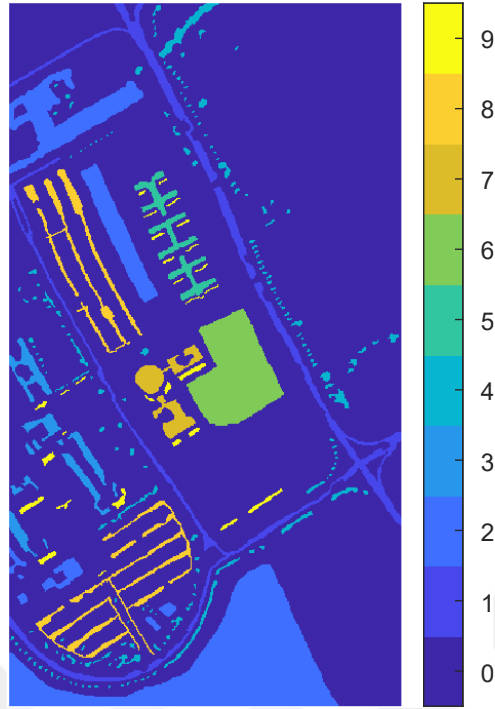
Tablo 3.2’de Pavia Centre imgesindeki sınıfların adları ve her sınıftaki piksel sayısı görülmektedir.

Tablo 3.2: Pavia Centre imgesindeki sınıflar ve piksel sayıları.

Sınıf no	Sınıf	Piksel sayısı
1	Water	824
2	Trees	820
3	Asphalt	816
4	Self-Blocking Bricks	808
5	Bitumen	808
6	Tiles	1260
7	Shadows	476
8	Meadows	824
9	Bare soil	820

3.5.3. Pavia University İmgesi

Pavia University imgesi İtalya’nın Pavia şehrinde elde edilmiş bir hiperspektral imge olup, 9 sınıf, 103 spektral bant ve 610×340 pikselden oluşmaktadır. Şekil 3.9’da Pavia University imgesinin gerçek sınıflandırma haritası görülmektedir.



Şekil 3.9: Pavia University imgesinin gerçek sınıflandırma haritası.

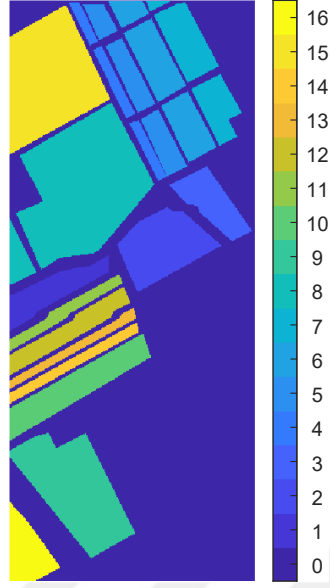
Pavia University imgesindeki sınıf isimleri ve her sınıftaki piksel sayıları Tablo 3.3'te görülmektedir.

Tablo 3.3: Pavia University imgesindeki sınıflar ve piksel sayıları.

Sınıf no	Sınıf	Piksel sayısı
1	Asphalt	6631
2	Meadows	18649
3	Gravel	2099
4	Trees	3064
5	Painted metal sheets	1345
6	Bare soil	5029
7	Bitumen	1330
8	Self-blocking bricks	3682
9	Shadows	947

3.5.4. Salinas İmgesi

Salinas imgesi Amerika Birleşik Devletleri'nin Kaliforniya eyaletindeki Salinas Vadisi'nde AVIRIS sensörüyle elde edilmiş olup, 224 spektral bant, 16 sınıf ve 512×217 pikselden oluşmaktadır. Testlerde 20 adet gürültülü bant hariç tutularak 204 spektral bant kullanılmaktadır. Şekil 3.10'da Salinas imgesinin gerçek sınıflandırma haritası görülmektedir.



Şekil 3.10: Salinas imgesinin gerçek sınıflandırma haritası.

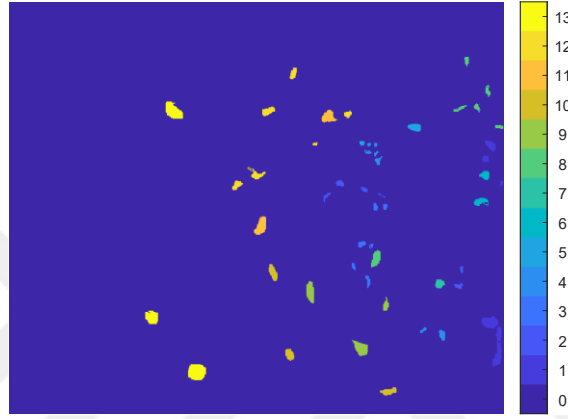
Salinas imgesindeki sınıfların isimleri ve her sınıftaki piksel sayıları Tablo 3.4'te görülmektedir.

Tablo 3.4: Salinas imgesindeki sınıflar ve piksel sayıları.

Sınıf no	Sınıf	Piksel sayısı
1	Brocoli_green_weeds_1	2009
2	Brocoli_green_weed_2	3726
3	Fallow	1976
4	Fallow_rough_plow	1394
5	Fallow_smooth	2678
6	Stubble	3959
7	Celery	3579
8	Grapes_untrained	11271
9	Soil_vinyard_develop	6203
10	Corn_senesced_green_weeds	3278
11	Lettuce_romaine_4wk	1068
12	Lettuce_romaine_5wk	1927
13	Lettuce_romaine_6wk	916
14	Lettuce_romaine_7wk	1070
15	Vinyard_untrained	7268
16	Vinyard_vertical_trellis	1807

3.5.5. Kennedy Space Center İmgesi

Kennedy Space Center imgesi AVIRIS sensörüyle Amerika Birleşik Devletleri'nin Florida eyaletindeki Kennedy Uzay İstasyonu üzerinde elde edilmiştir. Gürültülü bantların çıkarılmasıyla, 176 spektral bant kullanılmakta olup, imgede 13 sınıf ve 512×614 piksel bulunmaktadır. Şekil 3.11'de Kennedy Space Center imgesinin gerçek sınıflandırma haritası görülmektedir.



Şekil 3.11: Kennedy Space Center imgesinin gerçek sınıflandırma haritası.

Kennedy Space Center imgesindeki sınıfların isimleri ve piksel sayıları Tablo 3.5'te görülmektedir.

Tablo 3.5: Kennedy Space Center imgesindeki sınıflar ve piksel sayıları.

Sınıf no	Sınıf	Piksel sayısı
1	Scrub	761
2	Willow swamp	243
3	CP hammock	256
4	Slash pine	252
5	Oak/broadleaf	161
6	Hardwood	229
7	Swamp	105
8	Graminoid marsh	431
9	Spartina marsh	520
10	Cattail marsh	404
11	Salt marsh	419
12	Mud flats	503
13	Water	927

4. BULGULAR

4.1. AZ SAYIDA ÖRNEKLE EĞİTİM YAPILAN MODELLERLE ELDE EDİLEN SONUÇLAR

4.1.1. Uzamsal Olarak Değişen Student-t Karışım Modeli ile Elde Edilen Sonuçlar

Önerilen yöntemin başarımının değerlendirilmesi amacıyla Indian Pines, Pavia Centre, Pavia University ve Salinas imgeleriyle testler yapılmış, sonuçlar Bayesçi Gauss karışım modeli (BGMM: Bayesian Gaussian mixture model) (Kayabol, 2015), sözde-doğrusal ayırtaç analizi (p-LDA: Pseudo-linear discriminant analysis) ve köşegen-karesel ayırtaç analizi (d-QDA: Diagonal quadratic discriminant analysis) yöntemleriyle karşılaştırılmıştır. Tüm yöntemlerde piksel etiketleri üzerinde tanımlanan çokterimli özlojistik bağlanım önselleriyle izge-uzamsal sınıflandırma yapılmıştır.

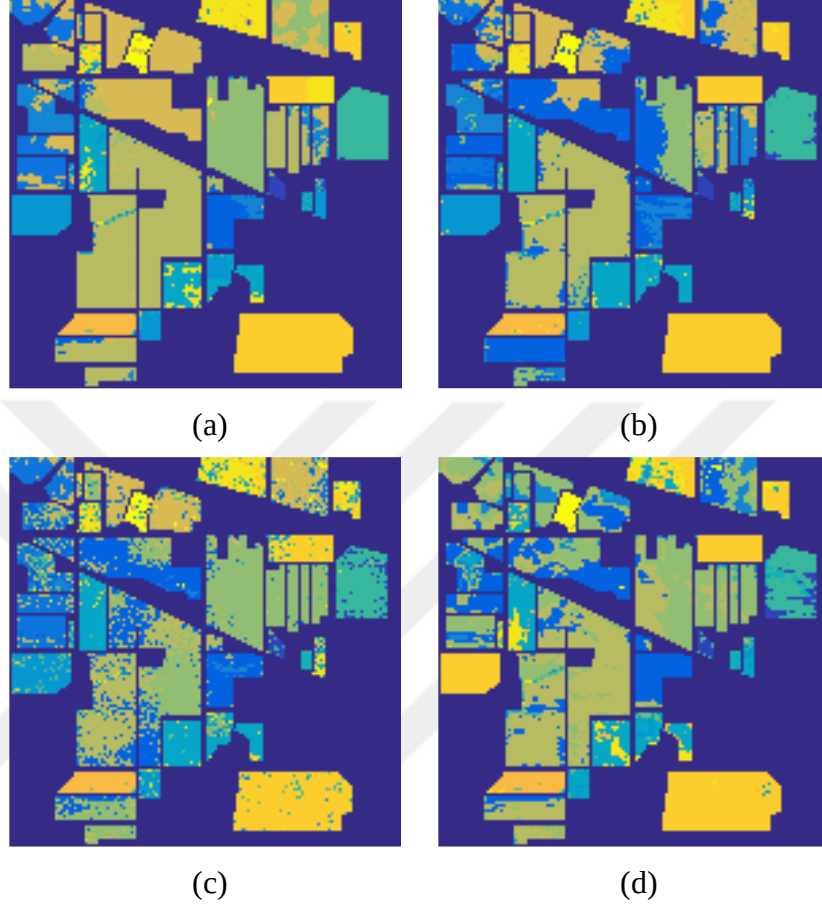
Eğitim kümesinin küçük seçilmesi nedeniyle kovaryans matrisinin pozitif tanımlı olmamasından dolayı LDA yerine p-LDA, QDA yerine de d-QDA algoritmaları kullanılmıştır. p-LDA algoritmasında matris tersi alma işlemi tekil matris ayrıştırma yöntemiyle yapılmakta olup d-QDA algoritmasında ise kovaryans matrisinin yalnız köşegen üzerindeki elemanları kullanılmaktadır.

Eğitim ve test kümeleri piksellerin rastgele seçilmesiyle oluşturulmuştur ve her sınıftaki piksellerin %10'u eğitim kümesi için ayrılmıştır. Piksel sayısı 20 veya daha az olan sınıflarda piksellerin yarısı eğitim için seçilmiş, geri kalanı test için ayrılmıştır. Testler rastgele ayrılan eğitim ve test kümeleri kullanılarak 20 kez tekrarlanmış ve ortalama sınıflandırma doğruluğu raporlanmıştır. Tablo 4.1'de elde edilen ortalama doğruluk değerleri görülmektedir. Bu tabloda önerilen yöntem olan Bayesçi Student-t karışım modeli BStMM olarak gösterilmiştir.

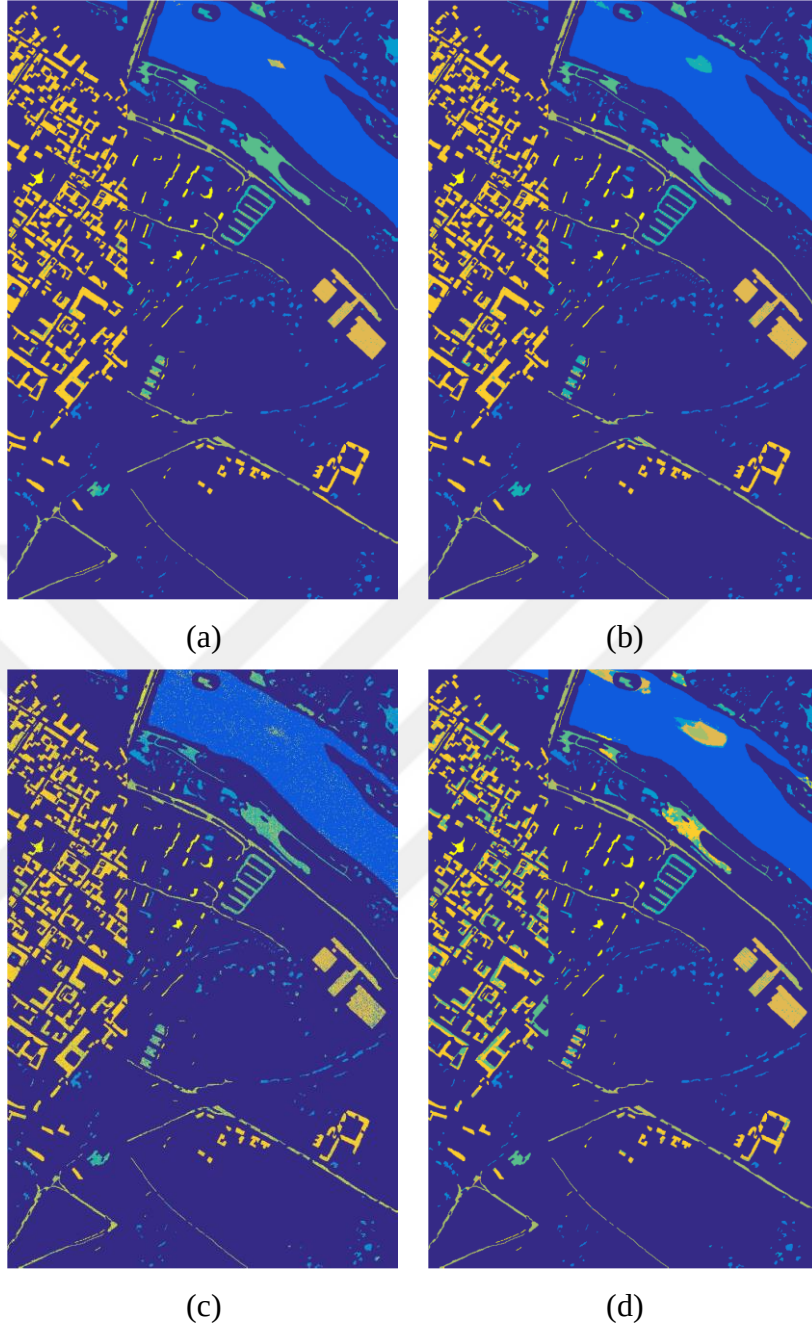
Tablo 4.1: Student-t karışım modeliyle elde edilen ortalama yüzde doğruluk değerleri.

	BGMM	BStMM	p-LDA	d-QDA
Indian Pines	65.04	65.30	61.16	47.65
Pavia Centre	95.03	95.26	81.19	84.77
Pavia University	68.11	73.62	56.27	65.05
Salinas	82.74	82.57	81.74	76.83

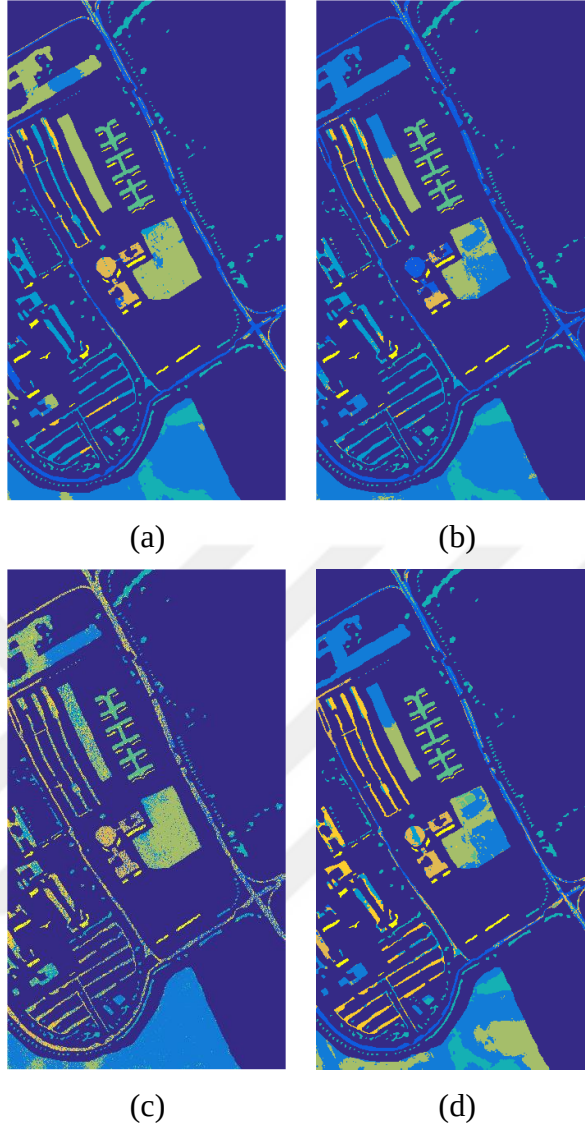
Indian Pines, Pavia Centre, Pavia University ve Salinas imgeleriyle elde edilmiş örnek sınıflandırma haritaları sırayla Şekil 4.1, Şekil 4.2, Şekil 4.3 ve Şekil 4.4'te görülmektedir.



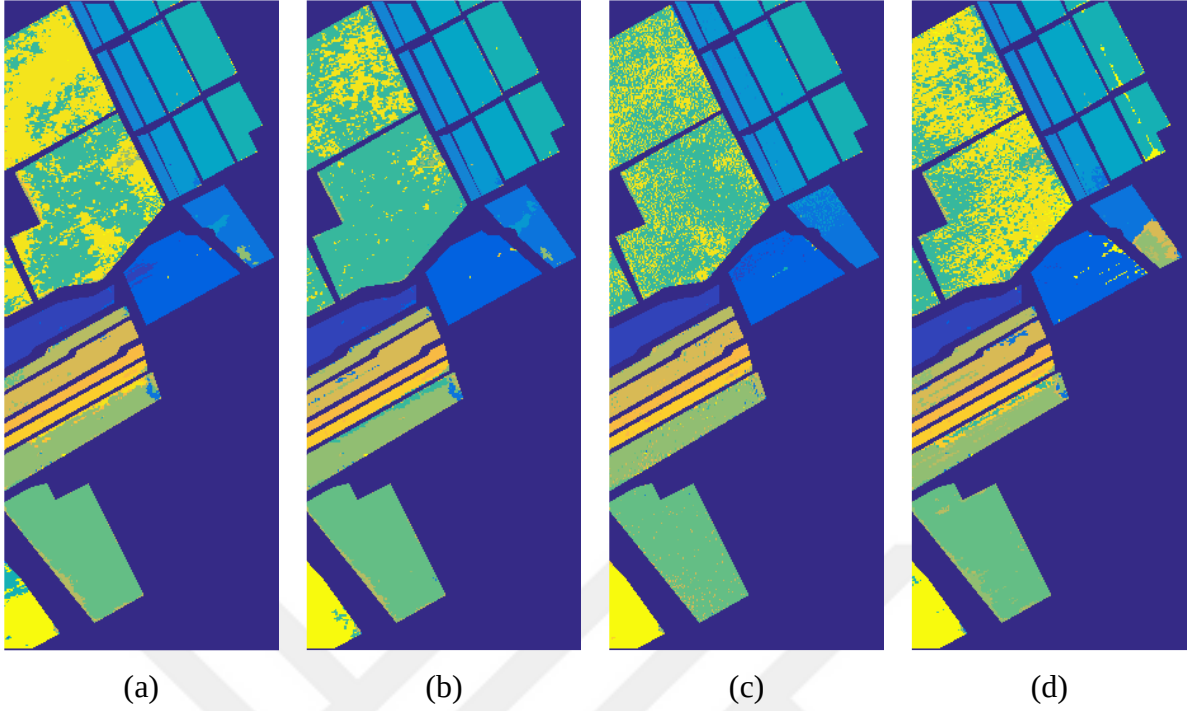
Şekil 4.1: Indian Pines imgesi üzerinde a) BGMM b) BStMM c) p-LDA d) d-QDA ile elde edilen örnek sınıflandırma haritaları.



Şekil 4.2: Pavia Centre imgesi üzerinde a) BGMM b) BStMM c) p-LDA ve d-QDA ile elde edilen örnek sınıflandırma haritaları.



Şekil 4.3: Pavia University imgesi üzerinde a) BGMM b) BStMM c) p-LDA d) d-QDA ile elde edilen örnek sınıflandırma haritaları.



Şekil 4.4: Salinas imgesi üzerinde a) BGMM b) BStMM c) p-LDA d) d-QDA ile elde edilen örnek sınıflandırma haritaları.

Elde edilen test sonuçlarına göre, Indian Pines, Pavia Centre ve Pavia University imgelerinde önerilen yöntemle diğer yöntemlerden daha iyi sonuçlar elde edilmiş, özellikle Pavia University imgesindeki sınıflandırma doğruluğunda önemli oranda bir iyileşme gözlenmiştir. Salinas imgesinde ise BGMM yöntemi az bir farkla daha iyi bir başarımla göstermiştir.

4.1.2. Uzamsal Olarak Değişen Gauss Karışım Modeli ile Elde Edilen Sonuçlar

Uzamsal olarak değişen Gauss karışım modelinin (SV-GMM: Spatially varying Gaussian mixture model) başarımlarının ölçülmesi için Indian Pines, Pavia Centre, Pavia University ve Salinas imgeleriyle testler yapılmıştır. Gauss karışım modeli, K-en yakın komşu (KNN: k-nearest neighbor), p-LDA, d-QDA, düzenlileştirilmiş LDA (RegLDA: Regularized LDA) ve destek vektör makinesi (SVM: Support vector machine) yöntemleriyle de testler yapılmıştır. Karşılaştırma amacıyla kullanılan algoritmalarda piksel etiketleri üzerinde Markov rastgele alan (MRF) (Geman ve Geman, 1984; Besag, 1986) önselleri tanımlanmıştır.

Eğitim kümesinin küçük seçilmesi nedeniyle kovaryans matrisinin tersi alınamadığından LDA yerine p-LDA ve RegLDA, QDA yerine de d-QDA algoritmaları kullanılmıştır. p-LDA’de kovaryans matrisinin tersinin bulunması için tekil matris ayrıştırma algoritması kullanılırken,

RegLDA’de kovaryans matrisinin köşegenindeki elemanlar büyütülerek tersinin alınabilmesi sağlanmaktadır. d-QDA’de ise yalnızca köşegen üstündeki elemanlar kullanılmaktadır.

GMM+MRF algoritmasında bir Gauss karışım modeli ile sınıflandırma yapılmakta, sonra da MRF ile düzeltirme yapılarak daha iyi sınıflandırma haritaları elde edilmektedir (Trianni ve Gamba, 2005; Plaza ve diğ., 2009). Kovaryans matrisi için Tadjudin ve Landgrebe (1999) tarafından önerilmiş olan bLOOC kestirim yöntemi kullanılmıştır.

SVM+MRF modeli Tarabalka ve diğ. (2010) tarafından önerilmiştir. SVM bir olasılıksal sınıflandırıcı değildir, ancak Platt (1999) tarafından geliştirilen yöntemle olasılık değerleri elde edilebilmektedir.

Eğitim ve test kümelerinin ayrılması için pikseller rastgele seçilmiş, sınıf başına düşen eğitim piksel sayısı ise $N_e \in \{5,10,15,20,25\}$ olarak seçilmiştir. Piksel sayısı az olan sınıflarda eğitime yeterli sayıda piksel ayrılabilmesi amacıyla, N_k k . sınıftaki piksel sayısını ve N_e^k k . sınıf için seçilen eğitim piksel sayısını göstermek üzere, eğitim piksel sayısı $N_e^k = \min(N_k/2, N_e)$ kuralıyla belirlenmiştir. Eğitim iterasyon sayısı 30 olarak seçilmiş ve testler 20 kez rastgele ayrılan eğitim ve test kümeleriyle tekrarlanmıştır.

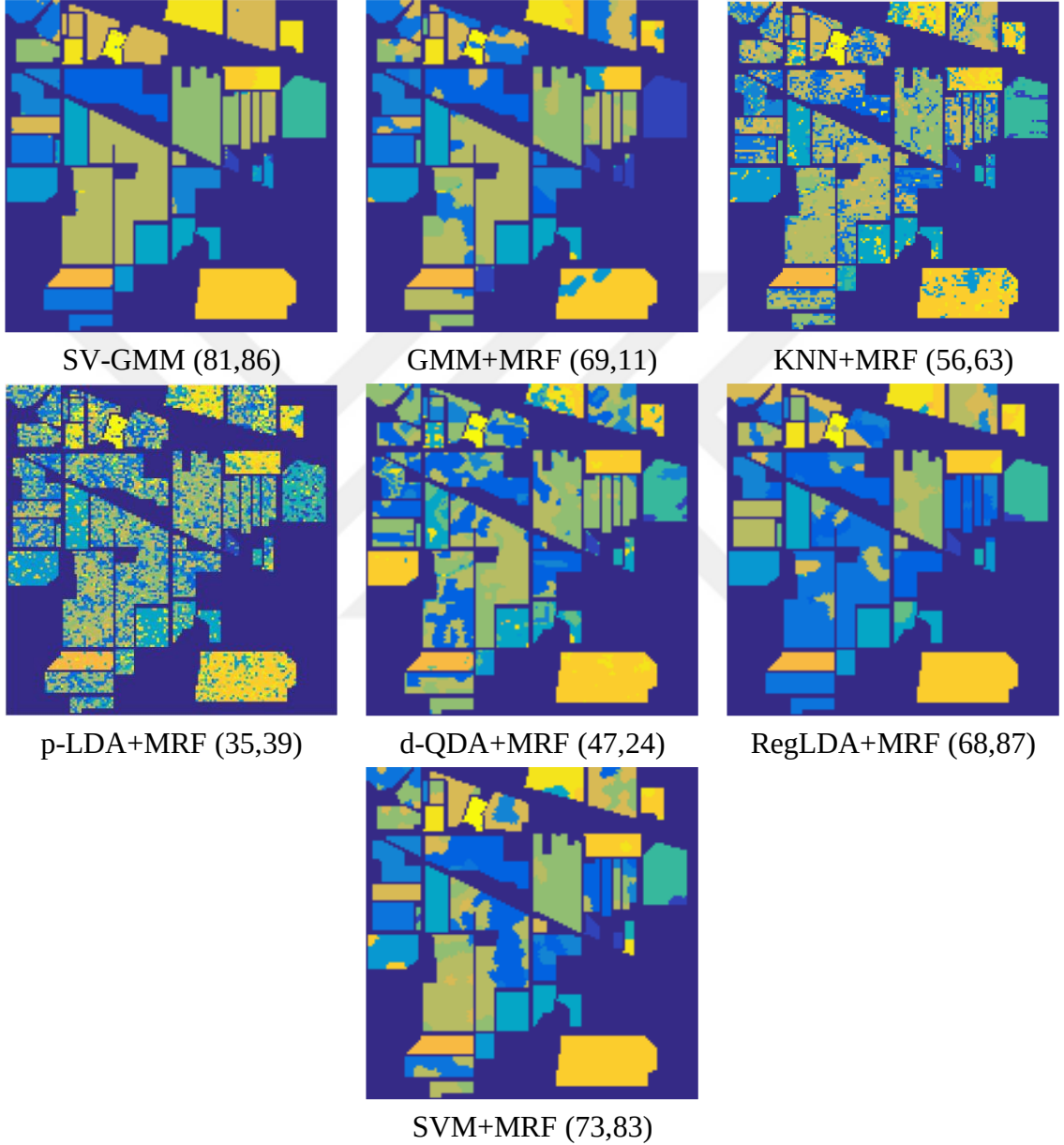
$N_e = 15$ durumu için Indian Pines, Pavia Centre, Pavia University ve Salinas imgeleriyle elde edilen sonuçlar Tablo 4.2’de görülmektedir. Bu tabloda önerilen yöntem SV-GMM olarak gösterilmiştir.

Tablo 4.2: $N_e = 15$ için elde edilen ortalama yüzde doğruluk ve standart sapma değerleri.

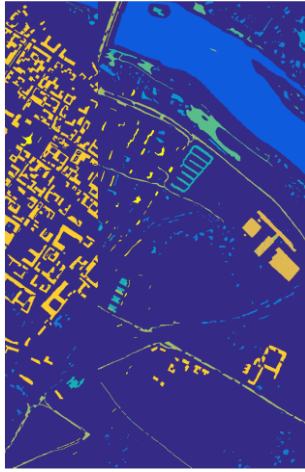
	Indian Pines	Pavia Centre	Pavia University	Salinas
SV-GMM	81,86 ± 3,30	97,78 ± 0,26	81,75 ± 4,84	91,41 ± 2,63
GMM+MRF	69,11 ± 4,40	85,00 ± 6,30	44,17 ± 6,67	86,84 ± 3,57
KNN+MRF	56,63 ± 1,69	93,36 ± 1,22	65,90 ± 3,25	80,91 ± 2,31
p-LDA+MRF	35,39 ± 3,68	90,40 ± 2,32	59,11 ± 6,83	74,95 ± 2,92
d-QDA+MRF	47,24 ± 4,02	85,52 ± 1,74	68,11 ± 6,65	78,87 ± 2,66
RegLDA+MRF	68,87 ± 4,94	96,59 ± 0,77	75,67 ± 5,30	89,82 ± 2,89
SVM+MRF	73,83 ± 5,18	96,70 ± 1,53	70,81 ± 6,16	87,55 ± 4,57

Indian Pines, Pavia Centre, Pavia University ve Salinas imgeleri için $N_e = 15$ durumunda elde edilen örnek sınıflandırma haritaları sırayla Şekil 4.5, Şekil 4.6, Şekil 4.7 ve Şekil 4.8’de görülmektedir.

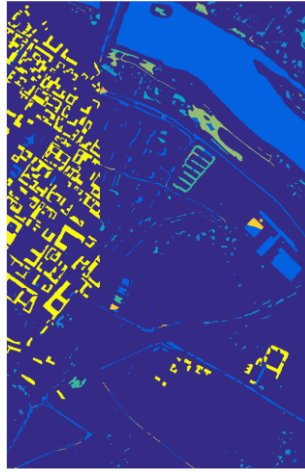
Eğitim kümesindeki örnek sayısına göre doğruluğun değişiminin gözlenmesi amacıyla, $N_e \in \{5,10,15,20,25\}$ durumları için $N_e^k = \min(N_k/2, N_e)$ kuralıyla belirlenen boyutlardaki eğitim kümeleri ile testler tekrarlanmıştır. Şekil 4.9’da bu beş farklı durum için sınıflandırma doğruluğunun eğitim kümesindeki toplam örnek sayısına göre değişimi görülmektedir.



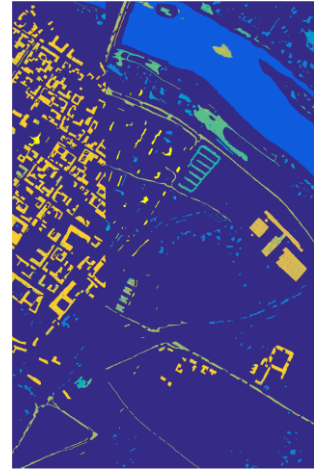
Şekil 4.5: Indian Pines imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kayabol ve Kutluk, 2016). Parantez içindeki değerler bu haritalar için hesaplanmış yüzde doğruluk değerleridir.



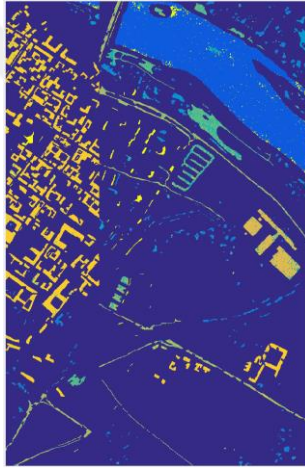
SV-GMM (97,78)



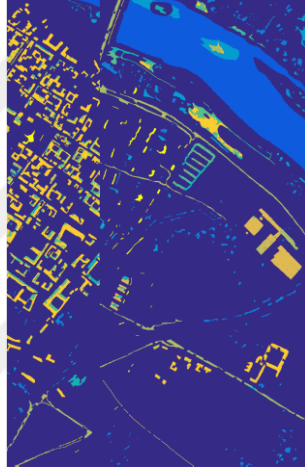
GMM+MRF (85,00)



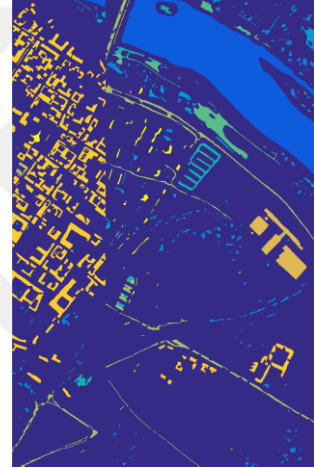
KNN+MRF (93,36)



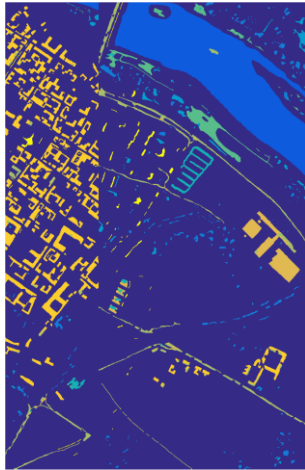
p-LDA+MRF (90,40)



d-QDA+MRF (85,52)

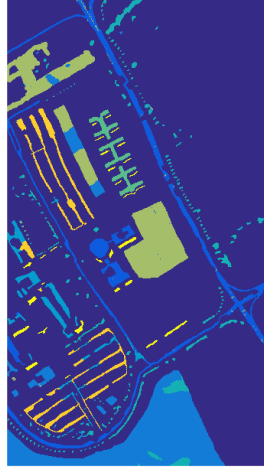


RegLDA+MRF (96,59)

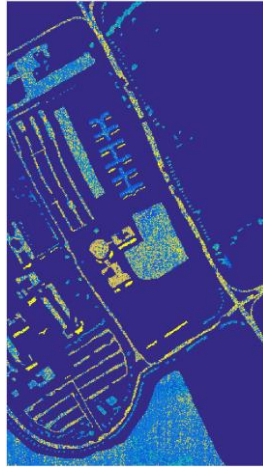


SVM+MRF (96,70)

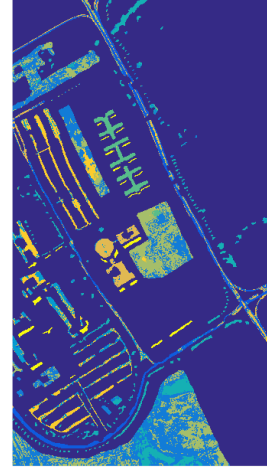
Şekil 4.6: Pavia Centre imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kayabol ve Kutluk, 2016). Parantez içindeki değerler bu haritalar için hesaplanmış yüzde doğruluk değerleridir.



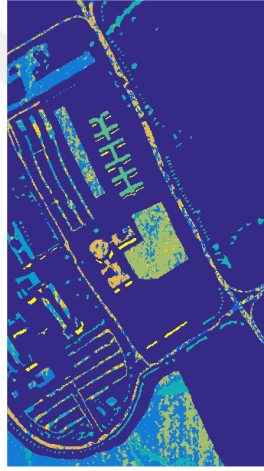
SV-GMM (81,75)



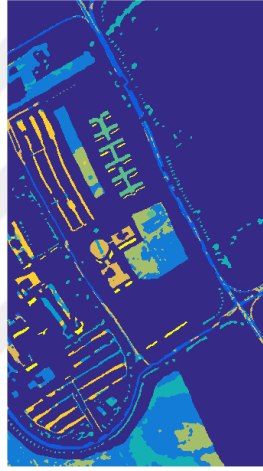
GMM+MRF (44,17)



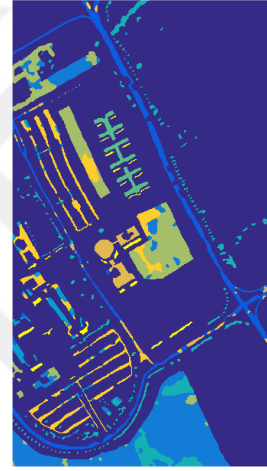
KNN+MRF (65,90)



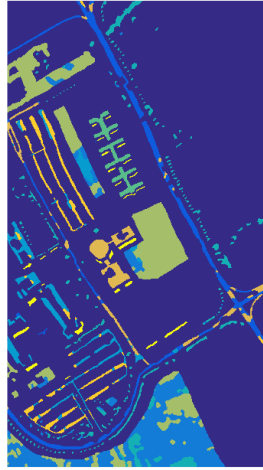
p-LDA+MRF (59,11)



d-QDA+MRF (68,11)

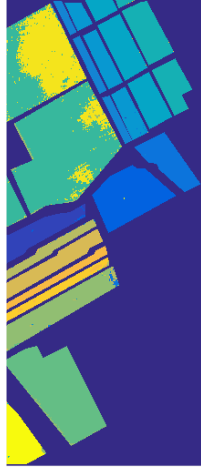


RegLDA+MRF (75,67)

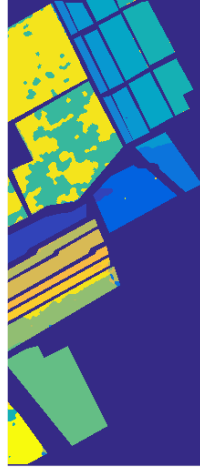


SVM+MRF (70,81)

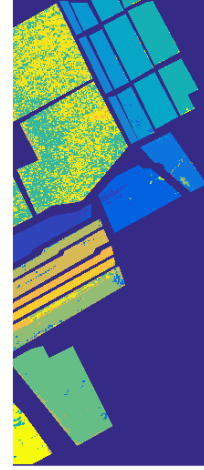
Şekil 4.7: Pavia University imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kayabol ve Kutluk, 2016). Parantez içindeki değerler bu haritalar için hesaplanmış yüzde doğruluk değerleridir.



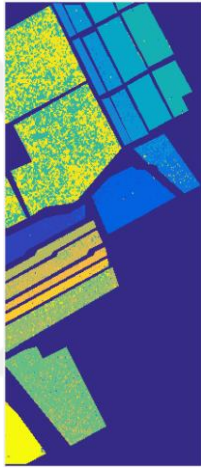
SV-GMM (91,41)



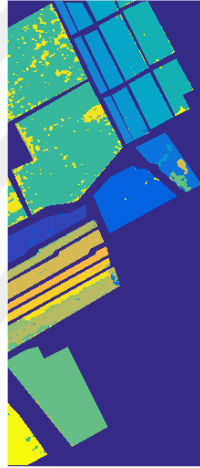
GMM+MRF (86,84)



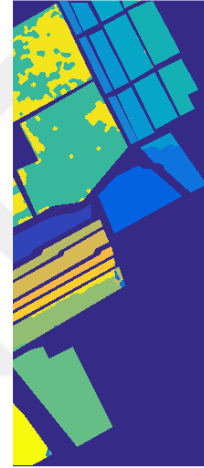
KNN+MRF (80,91)



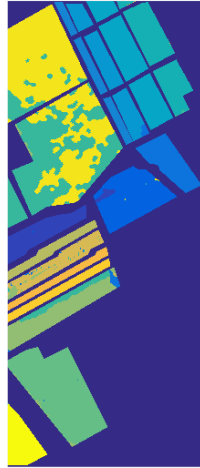
p-LDA+MRF (74,95)



d-QDA+MRF (78,87)

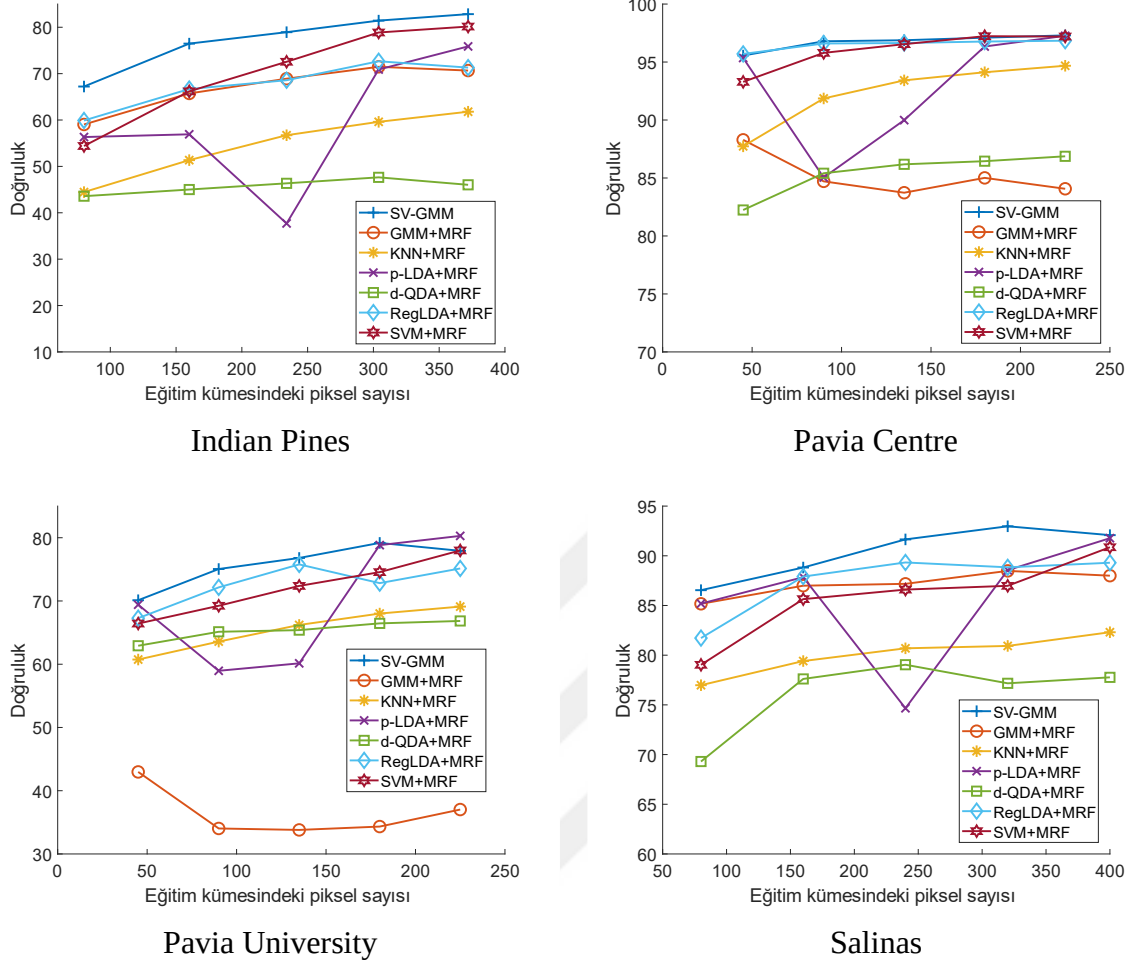


RegLDA+MRF (89,82)



SVM+MRF (87,55)

Şekil 4.8: Salinas imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kayabol ve Kutluk, 2016). Parantez içindeki değerler bu haritalar için hesaplanmış yüzde doğruluk değerleridir.



Şekil 4.9: Eğitim kümesi boyutuna göre sınıflandırma doğruluğunun değişimi (Kayabol ve Kutluk, 2016).

Tablolardan görüldüğü gibi, her dört imge için $N_e = 15$ durumunda diğer yöntemlerden daha iyi sınıflandırma sonuçları elde edilmiştir. Şekil 4.9'dan da görüldüğü gibi, önerilen yöntem az örnekle eğitim yapılması durumunda diğer yöntemlere göre daha iyi sınıflandırma başarımı göstermektedir.

4.2. ÜRETİCİ ALT-UZAY MODELİ İLE ELDE EDİLEN SONUÇLAR

Bölüm 3.2'de açıklanan olasılıksal temel bileşen analizi (PPCA: Probabilistic principal component analysis) ile durağan olmayan karışım modelinin başarımının ölçülmesi için Indian Pines, Pavia Centre, Pavia University ve Salinas imgeleri ile testler yapılmıştır.

Bu modelde spektral vektörler olasılıksal temel bileşen analizi ile modellenmekte olup, sınıf-koşullu dağılımlar normal dağılım olarak elde edilmektedir. Çokterimli özlojistik bağlanım

modeliyle sınıf etiketleri üzerinde önsel dağılım tanımlanmakta, böylece durağan olmayan Gauss karışım modeli elde edilmektedir (Kutluk ve diğ. 2016a; 2016b).

Önerilen modelin başarımının ölçülmesi için temel bileşen sayısı 5, 10 ve 15 seçilerek testler yapılmıştır. Elde edilen sonuçlar iki farklı yöntemle karşılaştırılmıştır. Bunlardan ilki Kayabol (2015) tarafından önerilen Bayesçi Gauss karışım modelidir (BGMM: Bayesian Gaussian mixture model), diğeri ise veriye temel bileşen analizi uygulandıktan sonra elde edilen Gauss karışım modelidir (PCA+GMM: Principal component analysis + Gaussian mixture model).

Testler rastgele ayrılan eğitim ve test kümeleri üzerinde 20 kez tekrarlanmış ve ortalama doğruluk değerleri hesaplanmıştır. Tablo 4.3'te 5, 10 ve 15 temel bileşen için olasılıksal temel bileşen analizi (sırayla PPCA-5, PPCA-10 ve PPCA-15) ile BGMM ve PCA+GMM yöntemleriyle elde edilen sonuçlar görülmektedir.

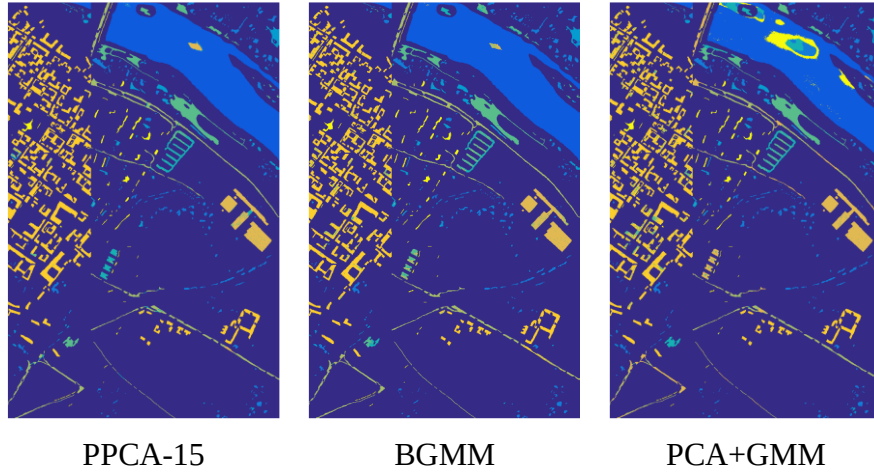
Tablo 4.3: Olasılıksal temel bileşen analizi karışım modeli ile elde edilen ortalama yüzde doğruluk değerleri.

	PPCA-5	PPCA-10	PPCA-15	BGMM	PCA+GMM
Indian Pines	73,18	69,17	71,62	65,04	57,40
Pavia Centre	96,04	96,13	96,16	95,03	88,67
Pavia University	81,26	82,86	82,10	68,11	71,97
Salinas	87,00	88,58	87,22	82,74	81,56

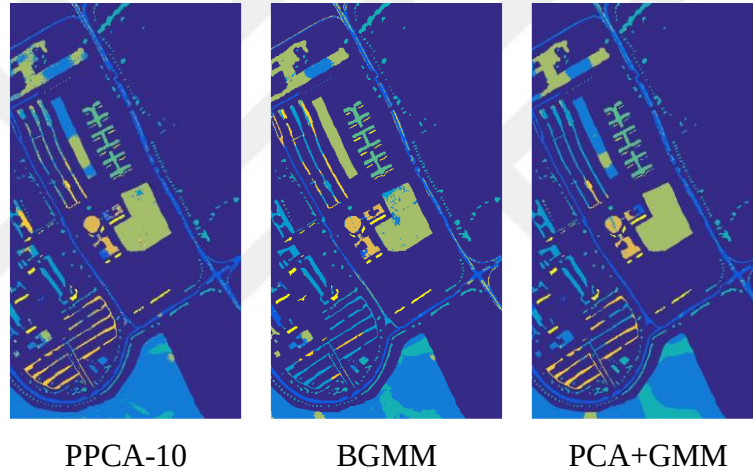
Indian Pines, Pavia Centre, Pavia University ve Salinas imgeleri üzerinde PPCA, BGMM ve PCA+GMM yöntemleriyle elde edilmiş örnek sınıflandırma haritaları sırayla Şekil 4.10, Şekil 4.11, Şekil 4.12 ve Şekil 4.13'te görülmektedir. Bu şekillerde önerilen yöntemde en iyi sonuç veren temel bileşen sayısı ile elde edilen sınıflandırma haritaları görülmektedir.



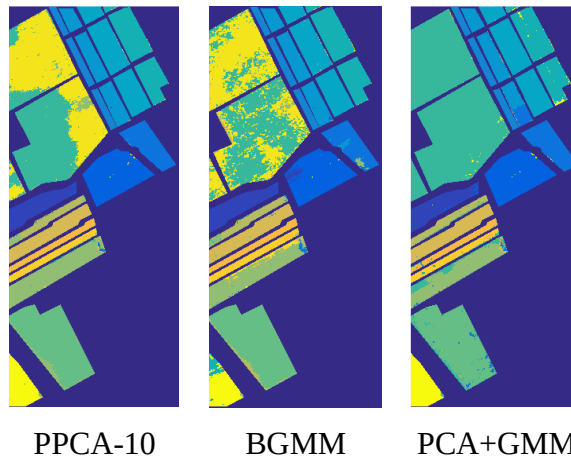
Şekil 4.10: Indian Pines imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kutluk ve diğ., 2016a).



Şekil 4.11: Pavia Centre imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kutluk ve diğ., 2016a).



Şekil 4.12: Pavia University imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kutluk ve diğ., 2016a).



Şekil 4.13: Salinas imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kutluk ve diğ., 2016a).

Tablo 4.3'ten görüldüğü gibi, her dört imge için önerilen yöntemle daha iyi sınıflandırma sonuçları elde edilmiştir. Dolayısıyla önerilen yöntemde boyut indirgeme ile hem işlemsel karmaşıklık azaltılmış, hem parametre kestirimi kolaylaştırılmış, hem de daha yüksek sınıflandırma başarımı elde edilmiştir. Olasılıksal temel bileşen analizi ile, standart temel bileşen analizine göre önemli oranda yüksek sınıflandırma doğruluğu elde edildiği gözlenmiştir.

4.3. ÇOKTERİMLİ LOJİSTİK BAĞLANIM MODELLERİ İLE ELDE EDİLEN SONUÇLAR

4.3.1. Üretici Model ile Elde Edilen Sonuçlar

İmgedeki spektral vektörlerin çokterimli lojistik bağlanım modelleri ile modellenmesi ve sınıf etiketleri üzerinde çokterimli özlojistik bağlanım önsellerinin tanımlanmasıyla elde edilen karışım modeli ile Indian pines ve Pavia University imgeleri kullanılarak testler yapılmıştır. Bu yöntemde Kayabol (2020) tarafından önerilen yöntemden farklı olarak uzamsal etiket önselleri tanımlanmış, spektral parametreler üzerinde ise bir önsel dağılım tanımlanmamıştır.

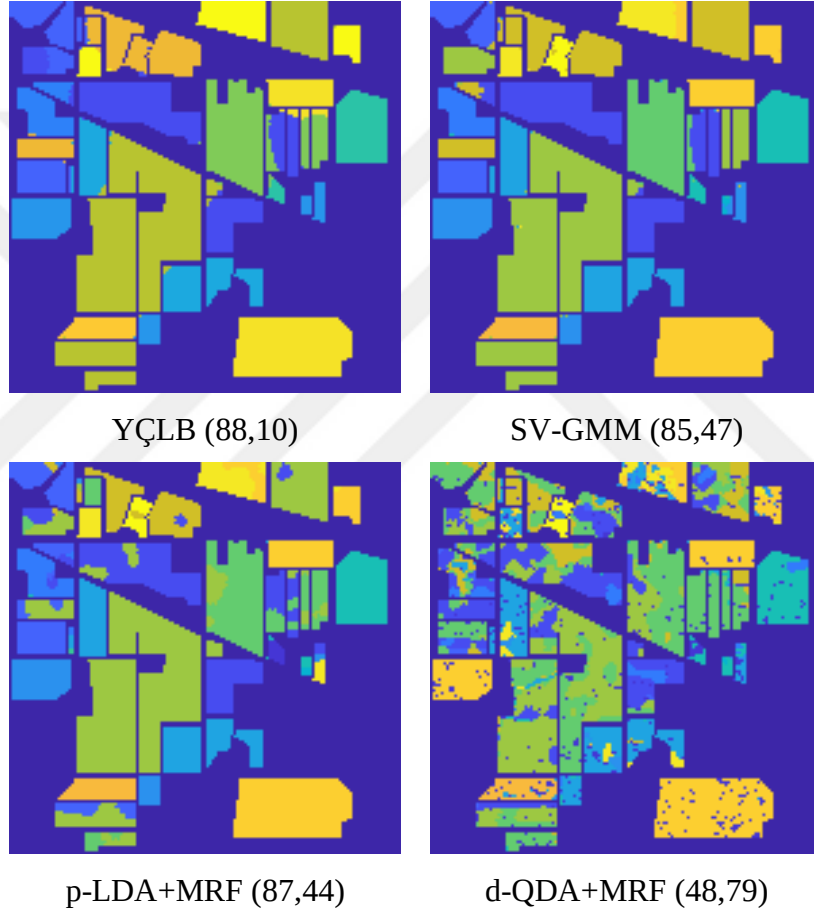
Başarımın değerlendirilmesi amacıyla bazı yöntemlerle karşılaştırma yapılmıştır. Bunlardan ilki Kayabol ve Kutluk (2016) tarafından önerilen uzamsal olarak değişen Gauss karışım modelidir (SV-GMM). Bu modelde Bölüm 3.1.2'de tanıtıldığı gibi, sınıf-koşullu dağılımlar Gauss olarak modellenmekte, piksel etiketleri üzerinde önsel dağılımların tanımlanması amacıyla da çokterimli özlojistik bağlanım modeli kullanılmaktadır. Böylece uzamsal olarak değişen bir Gauss karışım modeli elde edilmektedir. Karşılaştırma yapılan diğer yöntemlerde ise spektral vektörler p-LDA ve d-QDA ile modellenmiş, piksel etiketleri üzerinde ise Markov rastgele alan önselleri (Besag, 1986; Geman ve Geman, 1984) tanımlanarak izge-uzamsal sınıflandırma yapılmıştır.

İmgelerdeki her sınıfın %10'u rastgele seçilerek eğitim kümesi oluşturulmuştur. Rastgele seçilen eğitim ve test kümeleri üzerinde testler 20 kez tekrarlanarak ortalama doğruluk değerleri bulunmuştur. Tablo 4.4'te elde edilen ortalama doğruluk değerleri görülmektedir.

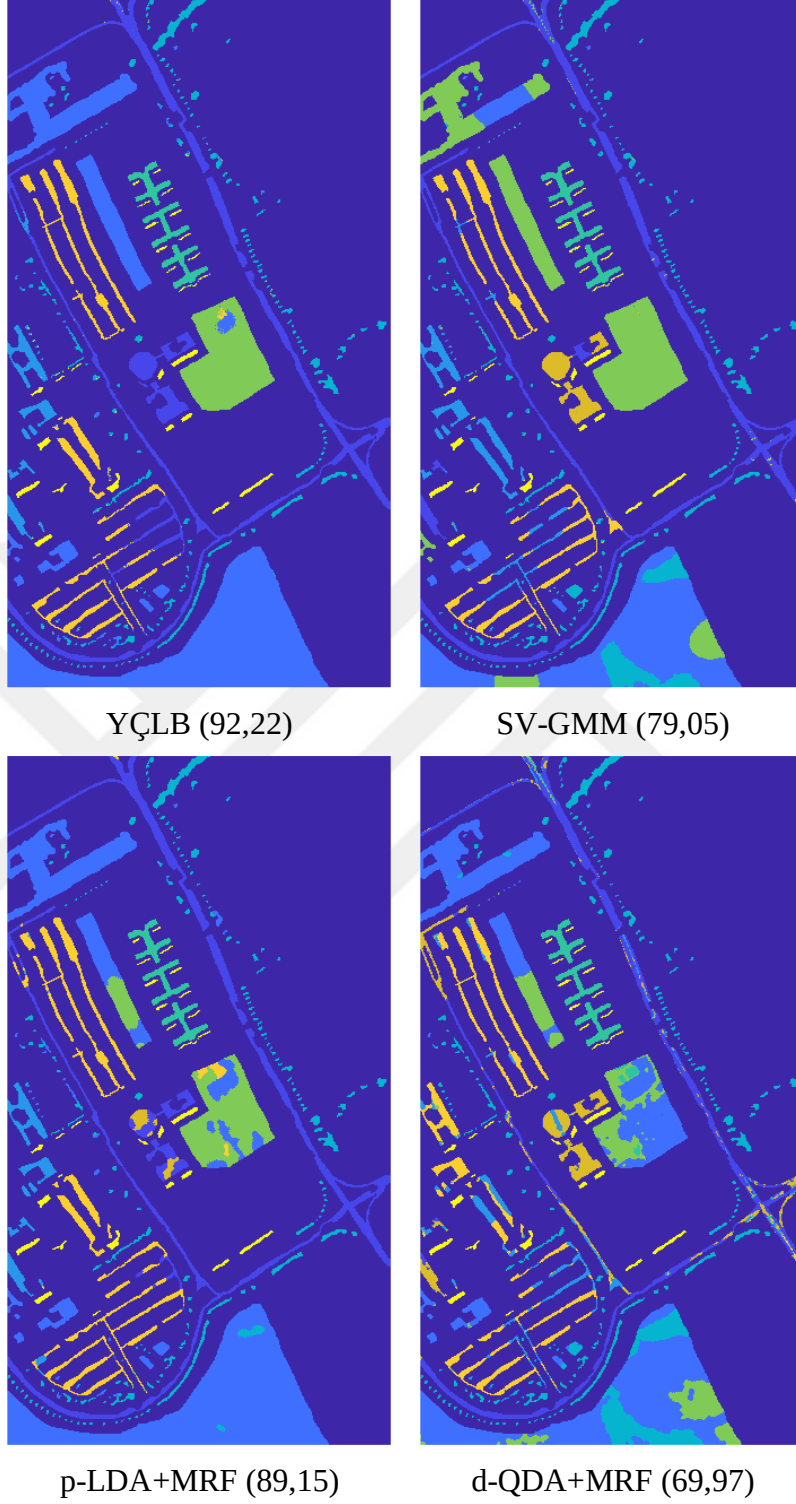
Tablo 4.4: Yaklaşık çokterimli lojistik bağlanım modeli ile elde edilen ortalama yüzde doğruluk değerleri.

	Önerilen yöntem (YÇLB)	SV-GMM	p-LDA+MRF	d-QDA+MRF
Indian Pines	88,12	85,41	87,37	48,92
Pavia University	92,00	78,96	89,19	69,88

Indian Pines ve Pavia University imgeleri ile elde edilen örnek sınıflandırma haritaları sırayla Şekil 4.14 ve Şekil 4.15'te görülmektedir.



Şekil 4.14: Indian Pines imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kutluk ve diğ., 2018). Parantez içinde verilen değerler bu haritalarda hesaplanan yüzde doğruluk değerleridir.



Şekil 4.15: Pavia University imgesi üzerinde elde edilen örnek sınıflandırma haritaları (Kutluk ve diğ., 2018). Parantez içinde verilen değerler bu haritalarda hesaplanan yüzde doğruluk değerleridir.

Test sonuçlarından görüldüğü gibi, önerilen izge-uzamsal sınıflandırma yöntemiyle her iki imge için SV-GMM, p-LDA+MRF ve d-QDA+MRF yöntemlerinden daha yüksek

sınıflandırma doğruluğu elde edilmiştir. Şekillerdeki örneklerden görüldüğü gibi daha iyi sınıflandırma haritaları da elde edilmiştir.

4.3.2. Ayrımsayıcı Model ile Elde Edilen Sonuçlar

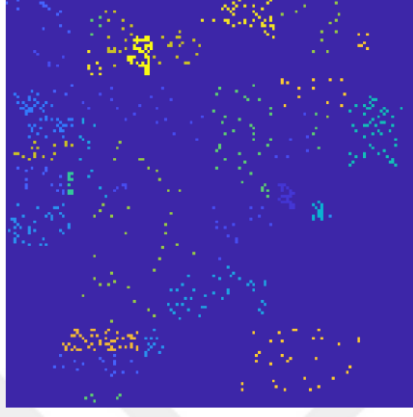
Bölüm 3.3.2’de açıklanan ayrımsayıcı modelde hem spektral hem uzamsal bilgi kullanılmaktadır. Bu modelde sınıf etiketlerinin koşullu olasılıkları doğrudan modellenmekte ve sınıflandırmada kullanılmaktadır. Indian Pines, Pavia University ve Salinas imgeleri üzerinde test yapılmış olup, imgeler eğitim ve test kümelerine pikseller rastgele seçilerek ayrılmıştır. Her imgede, bir sınıfa ilişkin eğitim piksellerinin sayısı $\min(N_k/2, 50)$ kuralıyla belirlenmiş olup burada N_k k . sınıftaki piksel sayısını göstermektedir. Toplam eğitim yineleme sayısı 60 test yineleme sayısı ise 10 olarak seçilmiştir. Tüm testler rastgele ayrılan eğitim ve test kümeleriyle 10 kez tekrarlanmış olup, ortalama doğruluk ve standart sapma değerleri kaydedilmiştir. Uzamsal modelleme için pencere boyutu 9×9 olarak, spektral parametrelerin seyreklik katsayısı ise 0,1 olarak seçilmiştir.

Uzamsal bilginin katkısının gözlenmesi amacıyla, uzamsal parametreler 0 seçilerek 30 eğitim iterasyonu boyunca yalnızca spektral parametre kestirimi yapılmış, daha sonra iki farklı durum için eğitime devam edilmiş ve 60 iterasyona tamamlanmıştır. İlk durumda uzamsal parametreler yine sıfır tutularak 30 iterasyon daha yalnızca spektral parametre kestirimine devam edilmiştir. İkinci durumda ise, spektral parametreler sabit tutulmuş ve 30 iterasyon boyunca uzamsal parametrelerin kestirimi gerçekleştirilerek toplam 60 iterasyon tamamlanmıştır. Testler 10 kez tekrarlanarak ortalama doğruluk ve standart sapma değerleri kaydedilmiştir. Tablo 4.5’te Indian Pines, Pavia University ve Salinas imgeleri için, yalnızca spektral parametre kestirimi yapılan 30 ve 60 iterasyon sonuçları ile, 30 iterasyon spektral parametre kestirimi yapıldıktan sonra, 30 iterasyon boyunca bu parametreler sabit tutularak yalnızca uzamsal parametre kestirimi yapıldığında elde edilen sonuçlar görülmektedir.

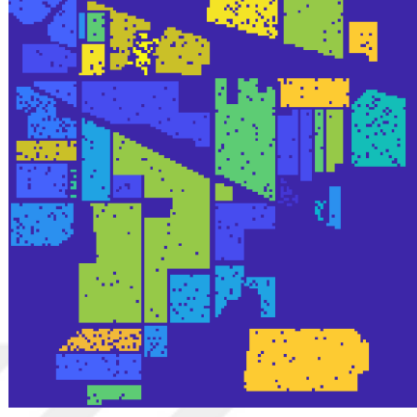
Tablo 4.5: Ayrımsayıcı model ile elde edilen sınıflandırma sonuçları (ortalama yüzde doğruluk $\pm$ standart sapma).

	Spektral sınıflandırma (30 iterasyon eğitim)	Spektral sınıflandırma (60 iterasyon eğitim)	İzge-uzamsal sınıflandırma (30 iterasyon spektral + 30 iterasyon uzamsal eğitim)
Indian Pines	64,28 $\pm$ 1,43	64,48 $\pm$ 1,42	72,28 $\pm$ 1,80
Pavia University	69,99 $\pm$ 2,71	71,49 $\pm$ 2,95	73,14 $\pm$ 3,80
Salinas	86,83 $\pm$ 1,23	86,74 $\pm$ 1,06	89,64 $\pm$ 1,29

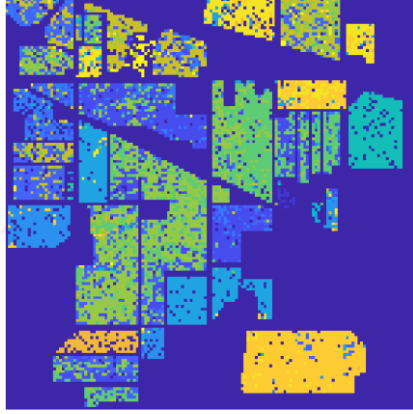
Indian Pines, Pavia University ve Salinas imgeleri için tekrarlanan testlerde kullanılan eğitim ve test kümeleri ile elde edilen sınıflandırma haritalarına örnekler sırayla Şekil 4.16, Şekil 4.17 ve Şekil 4.18’de görülmektedir.



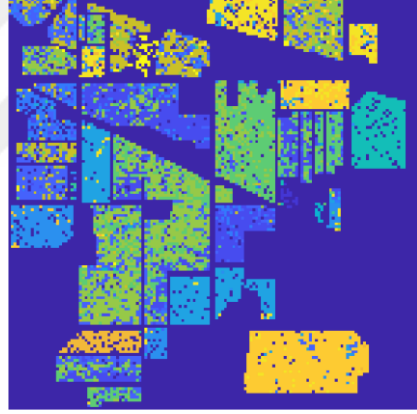
a) Eğitim imgesi



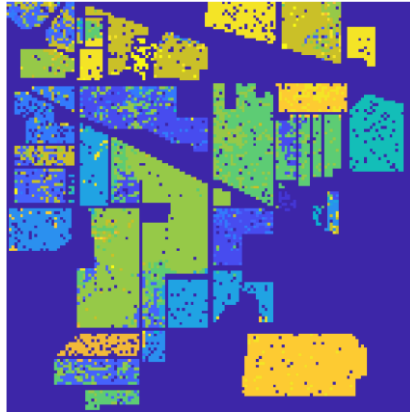
b) Test imgesi



c) Spektral sınıflandırma (30 iterasyon)

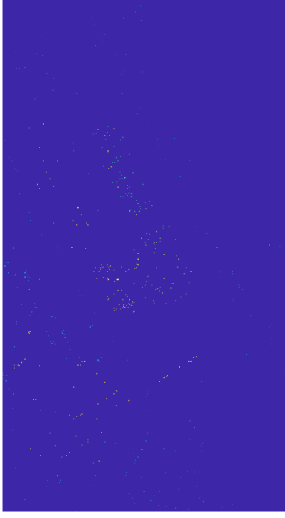


d) Spektral sınıflandırma (60 iterasyon)

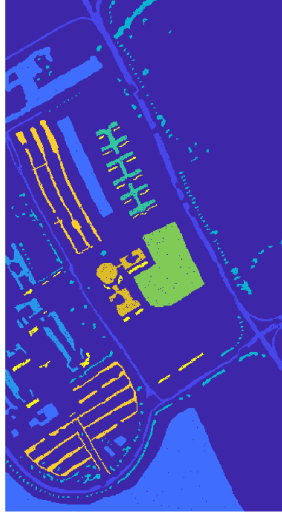


e) İzge-uzamsal sınıflandırma (30 iterasyon spektral, 30 iterasyon uzamsal)

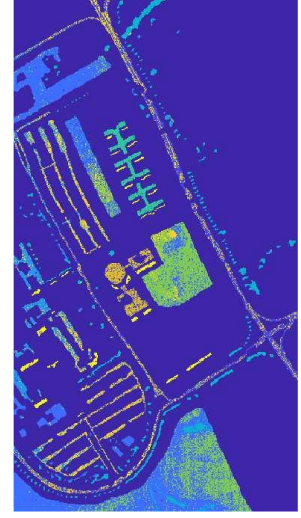
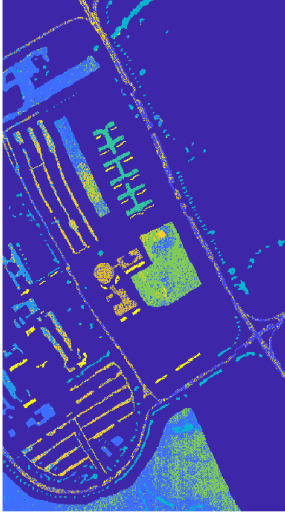
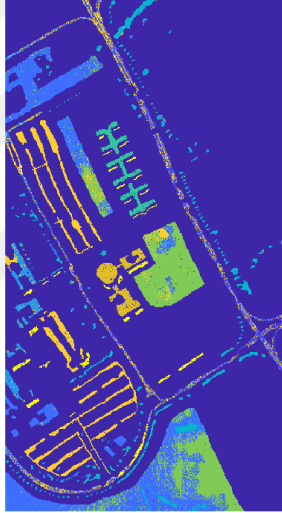
Şekil 4.16: Ayrımsayıcı model ile Indian Pines imgesinde elde edilen örnek sınıflandırma haritaları.



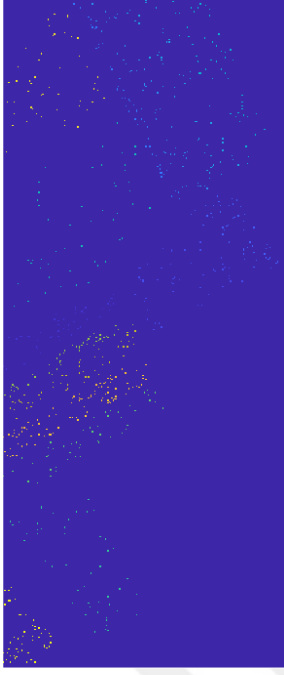
a) Eğitim imgesi



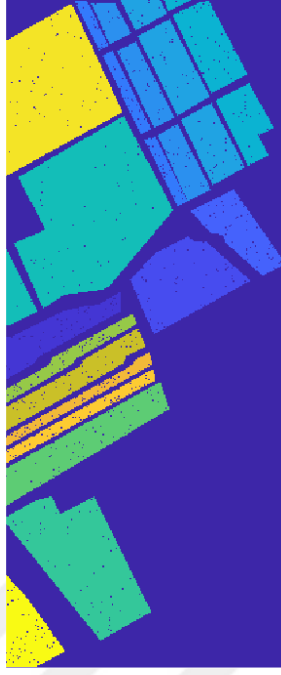
b) Test imgesi

c) Spektral sınıflandırma
(30 iterasyon)d) Spektral sınıflandırma
(60 iterasyon)e) İzge-uzamsal sınıflandırma (30
iterasyon spektral, 30 iterasyon
uzamsal)

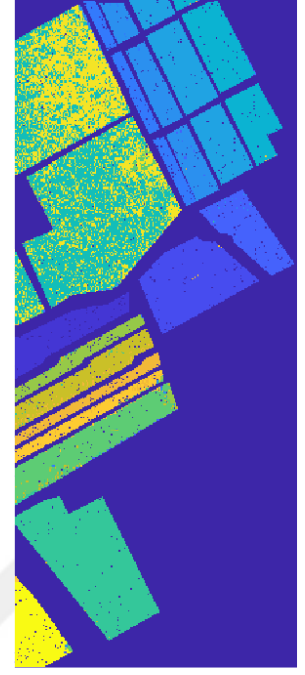
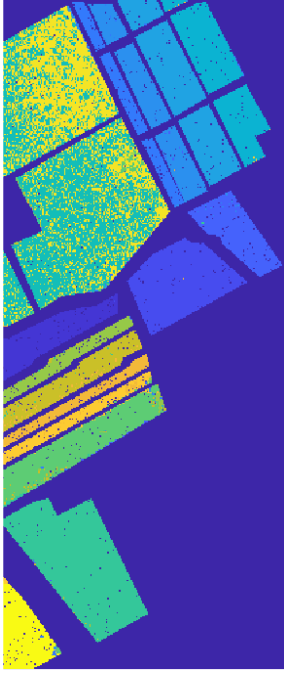
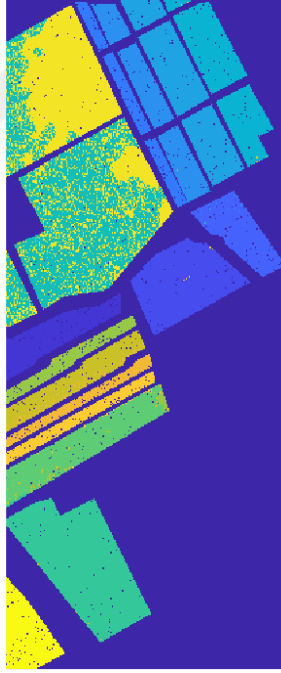
Şekil 4.17: Ayrımsayıcı model ile Pavia University imgesinde elde edilen örnek sınıflandırma haritaları.



a) Eğitim imgesi



b) Test imgesi

c) Spektral sınıflandırma
(30 iterasyon)d) Spektral sınıflandırma
(60 iterasyon)e) İzge-uzamsal sınıflandırma (30
iterasyon spektral, 30 iterasyon
uzamsal)

Şekil 4.18: Ayrımsayıcı model ile Salinas imgesinde elde edilen örnek sınıflandırma haritaları.

Tablo 4.5'te görüldüğü gibi her üç imge için de önerilen izge-uzamsal yöntemle, yalnızca izgesel sınıflandırma yapılması durumuna göre daha iyi sonuçlar elde edilmiştir. 30

iterasyondan sonra izgesel eğitime devam edilmesi yerine, izgesel parametrelerin sabit tutulmasıyla uzamsal parametrelerin eğitilmesinin sınıflandırma doğruluğunu önemli miktarda iyileştirdiği gözlenmiştir.

4.4. EVRİŞİMSEL SİNİR AĞI MODELİ İLE ELDE EDİLEN SONUÇLAR

Bu bölümde evrişimsel sinir ağlarının eğitimi için önerilmiş olan yeni yöntem ve bu modelin çokterimli özlojistik bağlanım modeliyle birlikte hiperspektral imgelerin izge-uzamsal sınıflandırılması için uygulanmasıyla elde edilen sonuçlar açıklanmaktadır. Önerilen yöntemin (CNN+ASMLR) başarımının ölçülmesi için Indian Pines ve Pavia University imgeleri üzerinde testler yapılmıştır. Karşılaştırma amacıyla, Kayabol (2020) tarafından önerilen yaklaşık seyrek çokterimli lojistik bağlanım modeli (ASMLR: Approximate sparse multinomial logistic regression), evrişimsel sinir ağı (CNN: Convolutional neural network) ve destek vektör makinesi (SVM: Support vector machine) ile de testler yapılmıştır. Tüm yöntemlerde uzamsal sınıflandırma için çokterimli özlojistik bağlanım modeli kullanılmıştır.

Destek vektör makinesi bir olasılıksal sınıflandırıcı değildir. Özlojistik bağlanım modeliyle birlikte kullanılabilmesi için, Platt (1999) tarafından önerilen ölçekleme yöntemi uygulanarak, destek vektör makinesi çıkışında olasılık değerlerinin elde edilmesi sağlanmıştır. Böylece uzamsal modelle de birleştirilmesi mümkün olmuştur.

Indian Pines imgesinde eğitim kümesi her sınıftan rastgele 50 piksel seçilmesiyle oluşturulmuştur. Piksel sayısı 100'den az olan sınıflarda piksellerin yarısı eğitim için kullanılmıştır. Pavia University imgesinde aynı işlem uygulanmış, ancak her sınıftan 50 yerine 100 piksel seçilerek eğitim kümesi oluşturulmuştur. Eğitimden önce eğitim kümesi standart hâle getirilerek sıfır ortalama ve birim standart sapmalı duruma getirilmiştir. Eğitim kümesinden hesaplanan ortalama ve standart sapma test kümesine de uygulanmıştır.

Indian Pines ve Pavia University imgeleri için kullanılan ağlardaki evrişim katmanlarının yapısı sırayla Şekil 3.5 ve Şekil 3.6'da görülmektedir. Bu katmanlar öznitelik çıkarımı katmanları olup, Indian Pines imgesi için elde edilen özniteliklerin boyutu 110, Pavia University imgesi içinse 150'dir.

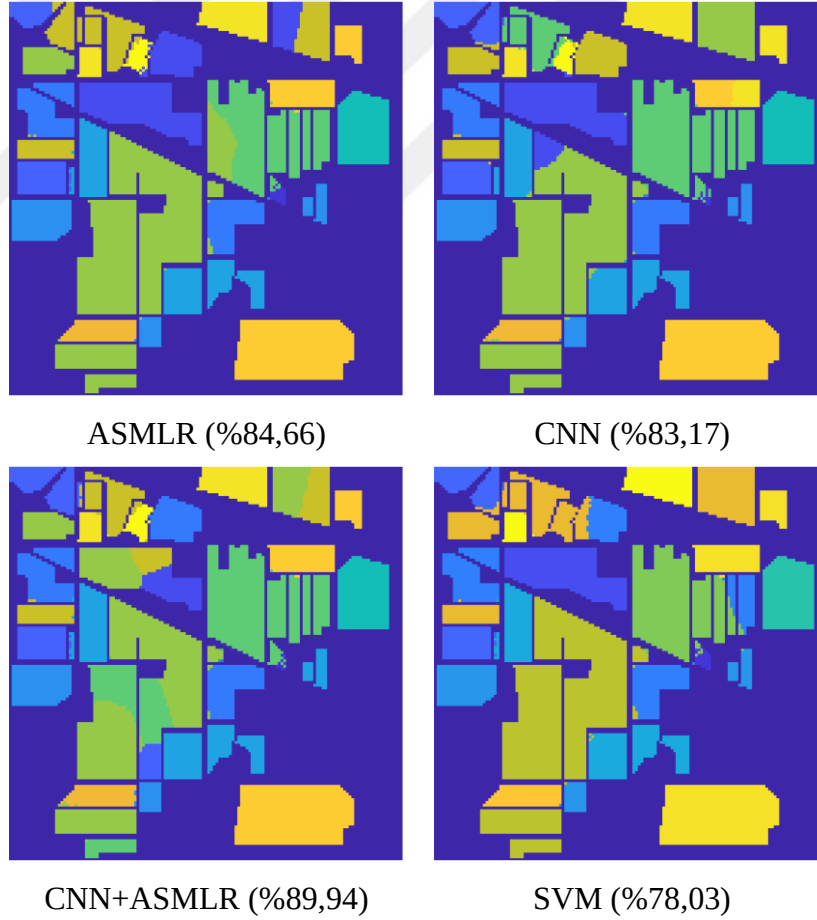
ASMLR, CNN ve CNN+ASMLR algoritmalarının eğitiminde iterasyon sayısı Indian Pines imgesi için 50, Pavia University içinse 100 olarak seçilmiştir. Her iterasyonda eğitim verisinin tamamı kullanıldığından bir devir bir iterasyona denktir.

Testler 20 kez tekrarlanmış, ortalama doğruluk ve standart sapma değerleri hesaplanmıştır. Tablo 4.6’da elde edilen ortalama doğruluk ve standart sapma değerleri görülmektedir.

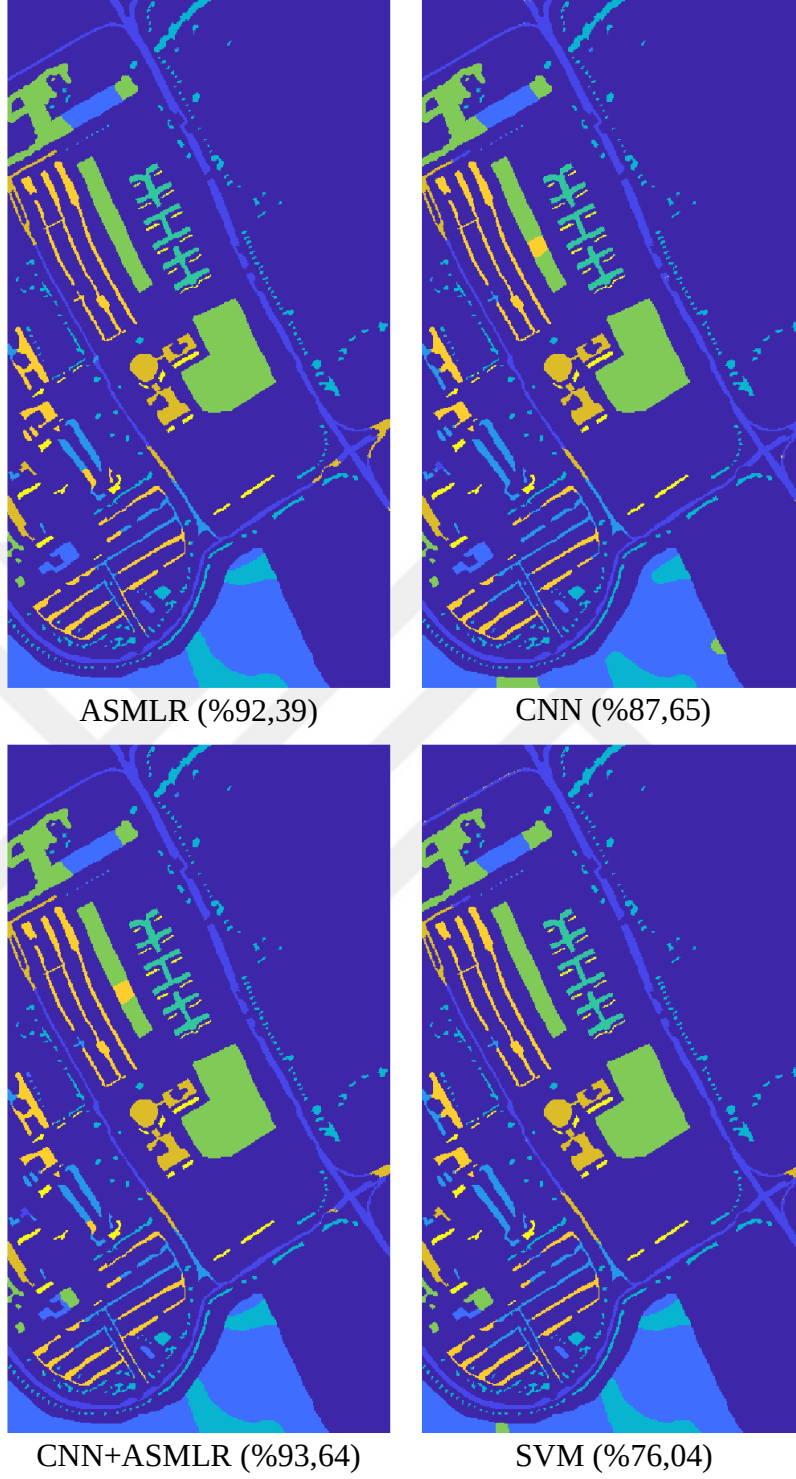
Tablo 4.6: Evrimsel sinir ağı ile elde edilen sonuçlar (% ortalama doğruluk $\pm$ standart sapma).

	ASMLR	CNN	CNN+ASMLR	SVM
Indian Pines	83,74 $\pm$ 2,99	83,30 $\pm$ 5,70	89,19 $\pm$ 2,82	77,88 $\pm$ 2,62
Pavia University	92,14 $\pm$ 2,13	87,33 $\pm$ 10,62	92,77 $\pm$ 5,29	76,31 $\pm$ 1,79

Indian Pines ve Pavia University imgeleri ile yapılan testlerde elde edilen örnek sınıflandırma haritaları ise sırayla Şekil 4.19 ve Şekil 4.20’de görülmektedir.



Şekil 4.19: Indian Pines imgesi ile elde edilen örnek sınıflandırma haritaları. Parantez içindeki değerler sınıflandırma doğruluğunu göstermektedir.



Şekil 4.20: Pavia University imgesi ile elde edilen örnek sınıflandırma haritaları. Parantez içindeki değerler sınıflandırma doğruluğunu göstermektedir.

Tablo 4.6'dan görüldüğü gibi, her iki imge için de ASMLR, CNN ve olasılıksal SVM sınıflandırıcılarına göre, önerilen CNN+ASMLR yöntemiyle daha iyi sonuçlar elde edilmiştir. Bu yöntemlerin hepsinde aynı uzamsal düzeltme önseli olan çokterimli özlojistik bağlanım

modeli kullanıldığından, bu fark CNN+ASMLR yöntemi için geliştirilmiş olan eğitim yöntemine bağlı olarak ortaya çıkmıştır.

Bu çalışmanın devamı olarak, parametre kestirimi yapılırken bir iterasyonda eğitim kümesinin tamamı yerine küçük bir grup eğitim örneği kullanılarak yöntemin başarımı ölçülmüştür. Başarım ölçümü için Indian Pines, Pavia University, Salinas ve Kennedy Space Center imgeleri ile testler yapılmıştır. Karşılaştırma amacıyla ADAM (Kingma ve Ba, 2015), DiffGrad (Dubey ve diğ., 2020), RAdam (Liu ve diğ., 2020) ve Limited memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algoritmalarıyla da testler yapılmıştır. LBFGS algoritması haricindeki algoritmalarda her iterasyonda eğitim kümesinin bir bölümü (minibatch) ile test yapılmış, LBFGS algoritmasında ise her iterasyonda eğitim kümesinin tamamı kullanılmıştır.

İmgeler eğitim ve test kümelerine ayrıldıktan sonra eğitim verisi standart hâle getirilmiş, aynı ortalama ve standart sapma değerleriyle test kümeleri de standart hâle getirilmiştir. Her sınıftan seçilen eğitim piksel sayısı, N_k k . sınıftaki etiketli piksel sayısını göstermek üzere, $\min(100, N_k/2)$ kuralına göre belirlenmiştir.

Eğitimde devir sayısı 100 olarak seçilmiş ve her 10 devirde test verisi üzerinde sınıflandırma doğruluğu ölçülmüştür. Testler 30 kez tekrarlanmış olup ortalama doğruluk ve standart sapma değerleri raporlanmıştır.

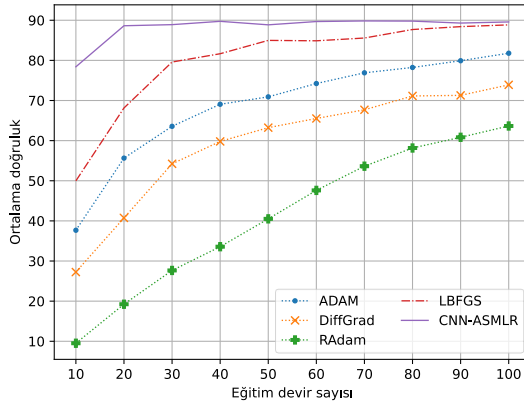
Indian Pines, Pavia University, Salinas ve Kennedy Space Center imgeleri için kullanılan ağların öznetelik çıkarma katmanlarının yapısı Tablo 4.7’de görülmektedir. Tüm ağlarda üç evrişim katmanı bulunmaktadır. Son katman ise doğrusal katman olup softmax fonksiyonu ile birlikte bir çokterimli lojistik bağlanım sınıflandırıcısı olarak düşünülebilir. Öznetelik çıkarma katmanları bir boyutlu evrişim, parametrik ReLU (He ve diğ., 2015), max pooling, grup normalizasyon (Ioffe ve Szegedy, 2015) ve sıfır ekleme (zero padding) bileşenlerinden oluşmaktadır. Tüm imgeler için evrişim katmanlarında sırayla 10, 10 ve 3 adet filtre kullanılmıştır ve filtre boyutu 5 olarak seçilmiştir.

ADAM, DiffGrad, RAdam algoritmaları ile CNN-ASMLR algoritmasındaki öznetelik katmanlarının eğitimi için öğrenme adım boyu 10^{-3} olarak seçilmiştir. Eğitim kümesi her iterasyonda biri kullanılmak üzere küçük parçalara ayrılmış, bu parçalardaki örnek sayısı (minibatch size) ise 512 olarak seçilmiştir.

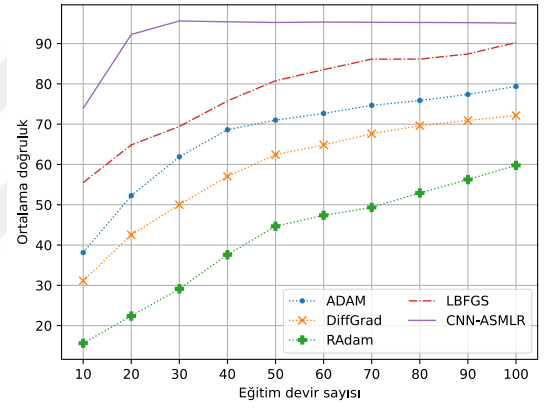
Tablo 4.7: Indian Pines, Pavia University, Salinas ve Kennedy Space Center imgeleri için kullanılan ağlardaki öznetelik çıkarma katmanlarının yapısı.

	Indian Pines	Pavia University	Salinas	Kennedy Space Center
1. katman	1-B evrişim PReLU Max pooling	1-B evrişim PReLU Padding	1-B evrişim PReLU Max pooling	1-B evrişim PReLU Max pooling
2. katman	1-B evrişim PReLU Max pooling	1-B evrişim PReLU Max pooling	1-B evrişim PReLU Max pooling	1-B evrişim PReLU Max pooling
3. katman	1-B evrişim PReLU Max pooling Grup normalizasyon	1-B evrişim PReLU Max pooling Grup normalizasyon	1-B evrişim PReLU Max pooling Grup normalizasyon	1-B evrişim PReLU Max pooling Grup normalizasyon

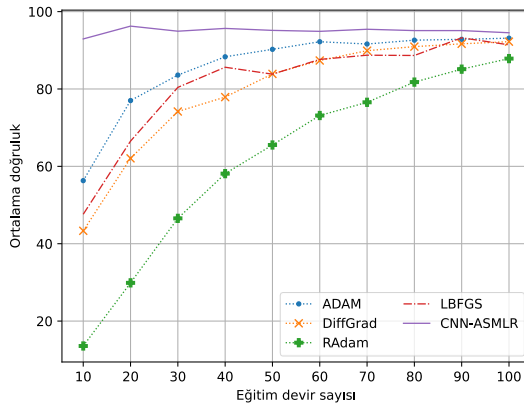
Şekil 4.21’de eğitim devir sayısına göre ortalama sınıflandırma doğruluğunun değişimi görülmektedir.



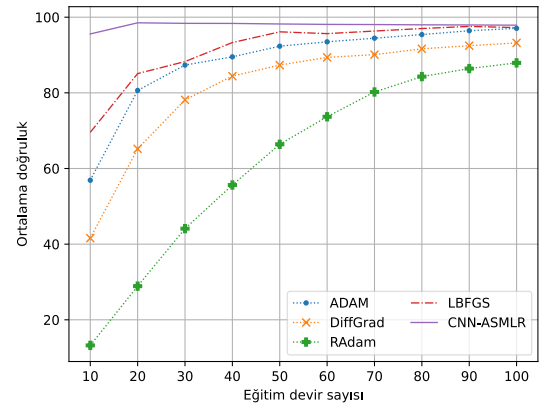
(a) Indian Pines



(b) Pavia University



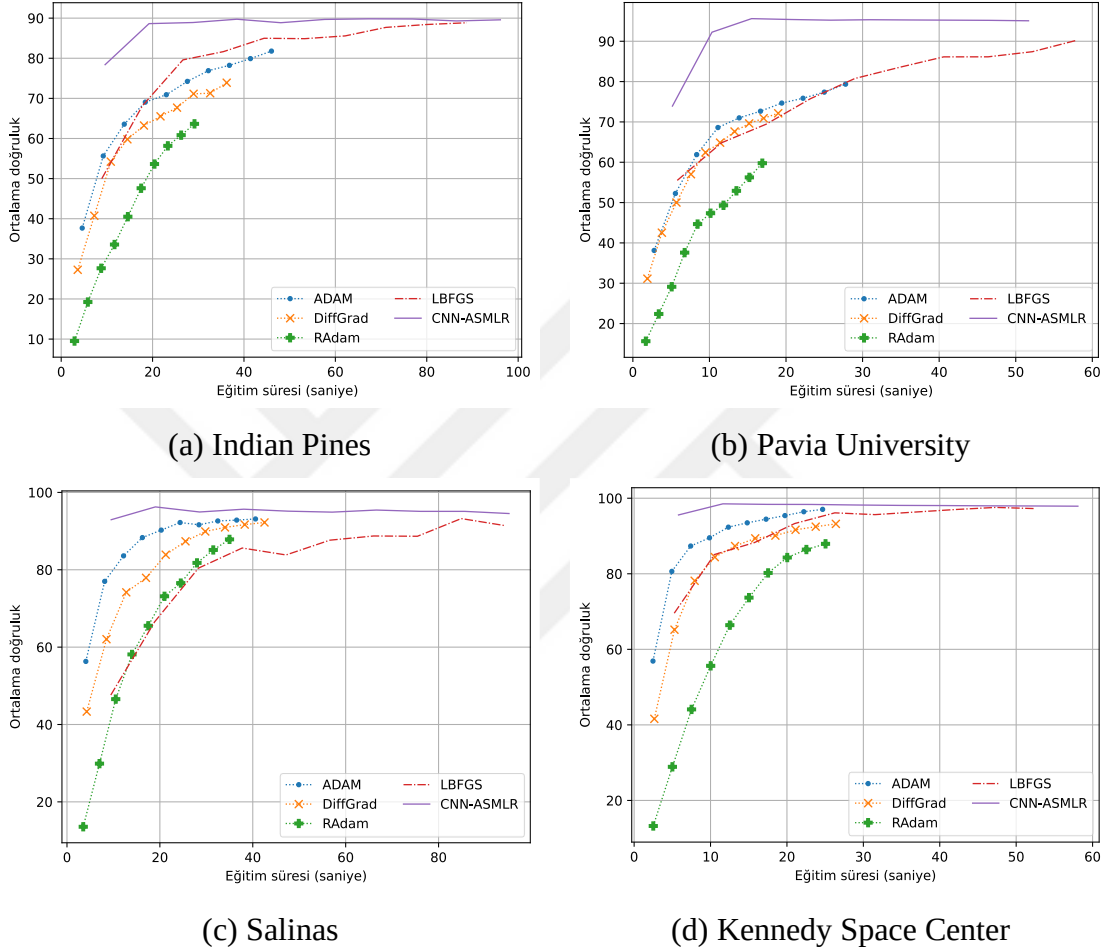
(c) Salinas



(d) Kennedy Space Center

Şekil 4.21: Devir sayısı ile ortalama sınıflandırma doğruluğunun değişimi.

Şekil 4.22’de saniye cinsinden ortalama eğitim süresi ile ortalama sınıflandırma doğruluğunun değişimi görülmektedir. Ortalama eğitim süreleri 100 devir boyunca ölçülen eğitim sürelerinin 10 tekrar için ortalaması olarak hesaplanmış, bir devir süresi ise toplam ortalama eğitim süresinin 1/100’ü olarak hesaplanmıştır.



Şekil 4.22: Eğitim süresi ile sınıflandırma doğruluğunun değişimi.

Tablo 4.8’de yüzüncü iterasyonda elde edilen ortalama sınıflandırma doğruluğu ve standart sapma değerleri görülmektedir.

Testler sonucunda her dört imge için de önerilen yöntemle diğer yöntemlere göre yüzüncü iterasyon sonucunda elde edilen sınıflandırma doğruluklarının daha yüksek olduğu gözlenmiştir. Ayrıca önerilen yöntemle, diğer algoritmalarla göre daha az sayıda eğitim iterasyonu ile daha yüksek sınıflandırma doğruluğu elde edildiği gözlenmiştir. İkinci derece optimizasyon kullanılması nedeniyle oluşan daha yüksek maliyete rağmen, önerilen yöntemin

eğitim süresi açısından da avantajlı olduğu, dolayısıyla daha kısa süre eğitilerek daha yüksek sınıflandırma başarımına ulaşabildiği de gözlenmiştir.

Tablo 4.8: Evrişimsel sinir ağı ile 100. devirde elde edilen ortalama sınıflandırma doğruluğu ve standart sapma değerleri.

	Indian Pines	Pavia University	Salinas	Kennedy Space Center
ADAM	81,78 ± 3,40	79,36 ± 3,54	93,14 ± 4,69	97,05 ± 1,43
DiffGrad	73,90 ± 5,06	72,15 ± 3,56	92,23 ± 3,70	93,20 ± 2,02
RAdam	63,65 ± 6,33	59,77 ± 9,76	87,86 ± 5,94	87,93 ± 2,43
LBFGS	88,85 ± 3,67	90,25 ± 4,72	91,41 ± 10,06	97,25 ± 2,65
CNN-ASMLR	89,58 ± 3,32	95,08 ± 1,09	94,54 ± 2,22	97,89 ± 0,39

Bu yöntemle elde edilen sonuçların istatistiksel olarak anlamlılığının değerlendirilmesi için tek yönlü ANOVA ve Tukey post-hoc analizleri yapılmıştır. Tablo 4.9’da bu analiz sonuçları görülmektedir.

Tablo 4.9: Evrişimsel sinir ağı modeli için istatistiksel analiz sonuçları.

	ANOVA	η^2 etki büyüklüğü	Post-hoc	
Indian Pines	$p < 0,05$	0,83	CNN-ASMLR – ADAM	$p < 0,05$
			CNN-ASMLR – DiffGrad	$p < 0,05$
			CNN-ASMLR – Radam	$p < 0,05$
			CNN-ASMLR – LBFGS	$p = 0,97$
Pavia University	$p < 0,05$	0,85	CNN-ASMLR – ADAM	$p < 0,05$
			CNN-ASMLR – DiffGrad	$p < 0,05$
			CNN-ASMLR – Radam	$p < 0,05$
			CNN-ASMLR – LBFGS	$p < 0,05$
Salinas	$p < 0,05$	0,13	CNN-ASMLR – ADAM	$p = 0,89$
			CNN-ASMLR – DiffGrad	$p = 0,56$
			CNN-ASMLR – Radam	$p < 0,05$
			CNN-ASMLR – LBFGS	$p = 0,26$
Kennedy Space Center	$p < 0,05$	0,79	CNN-ASMLR – ADAM	$p = 0,46$
			CNN-ASMLR – DiffGrad	$p < 0,05$
			CNN-ASMLR – Radam	$p < 0,05$
			CNN-ASMLR – LBFGS	$p = 0,71$

Tablo 4.9’deki sonuçlara göre dört imge için de ANOVA sonuçlarına göre yöntemlerden en az biri diğerlerinden farklı bir ortalama değere sahiptir. CNN-ASMLR yönteminin ortalama doğruluğu tüm imgeler için diğer yöntemlerden daha yüksek olsa da post-hoc testlere göre CNN-ASMLR ile Indian Pines imgesinde LBFGS yönteminin, Salinas imgesinde ADAM, DiffGrad ve LBFGS yöntemlerinin, Kennedy Space Center imgesinde ise ADAM ve LBFGS yöntemlerinin ortalama doğruluk değerleri arasında istatistiksel olarak anlamlı bir fark görülmemektedir. Buna karşın, önerilen yöntemin diğer yöntemlere göre devir sayısı ve eğitim süresi açısından daha hızlı yakınsadığı ve daha kısa sürede daha iyi sınıflandırma başarımı

sağladığı görülmektedir. Dolayısıyla evrişimsel sinir ağlarının eğitimi için önerilen yöntemin diğer yöntemlere göre daha etkin bir öğrenme sağladığı sonucuna ulaşılmaktadır.



5. TARTIŞMA VE SONUÇ

Bu tezde farklı yaklaşımlarla hiperspektral imgelerde yeryüzü bölgelerinin sınıflandırılması için çeşitli olasılıksal modeller ile öğrenme ve çıkarsama algoritmaları geliştirilmiştir.

Problemin diğer imge işleme problemlerinden bazı farkları bulunmaktadır. Bunlardan biri deneylerde kullanılabilecek verinin sınırlı olmasıdır. Güncel literatürde hâlen Indian Pines ve Pavia University gibi imgeler kullanılmaktadır. Bunun temel nedeni, imgelerin elde edildikten sonra veri etiketleme sürecinin zorluğu ve yüksek maliyetli olmasıdır. Bir pikselin yeryüzünde görece büyük bir bölgeyi temsil etmesi, etiketleme yapılırken bu bölgeleri bir uzmanın gezerek incelemesinin gerekliliği ve bazı durumlarda yeryüzü şeklinden dolayı uzmanın pikselin temsil ettiği bölgeye ulaşamaması nedeniyle etiketli veri sınırlı kalmaktadır. Gürültü kaynakları ve atmosferdeki değişim nedeniyle aynı bölgede, aynı sensörle, farklı zamanlarda yapılan görüntülemelerin de farklı sonuçlanacak olması, yapay öğrenme modelinin eğitimi ve başarımının ölçülmesinin genellikle aynı imgeden ayrılan eğitim ve test kümeleri üzerinde yapılmasını zorunlu kılmakta, literatürdeki çalışmalarda da bu yaklaşım uygulanmaktadır. Eğitim kümesinin küçük olmasıyla birlikte hiperspektral imgelerdeki spektral bant sayısının fazla olması sınıflandırma modellerinin eğitimi zorlaştırmaktadır.

Tezde yapılan çalışmalarda yanıt aranan sorulardan biri, sınırlı eğitim kümesiyle modelin başarılı bir şekilde eğitilebilmesidir. Bunun için az örnekle eğitim yapılabilmesi amacıyla spektral verinin farklı dağılımlarla modellenmesi ve parametreler üzerinde Bayesçi yaklaşımla önsel dağılımlar tanımlanarak parametre kestiriminin kolaylaştırılması sağlanmıştır. Diğer bir yöntemde ise bir üretici alt-uzay yöntemiyle veri boyutu indirgenerek parametre kestirimi kolaylaştırılmış, ayrıca sınıflandırma başarımının da iyileştiği gözlenmiştir.

Lojistik bağlanım modeli bir ayırmsayıcı model olup, doğrudan sınıf etiketinin koşullu olasılığı, bir diğer deyişle, sınıf etiketinin sonsal dağılımı modellenmektedir. Bu modelin başka çalışmalarda önerilmiş üretici formülasyonu, etiket sonsal dağılımının sınıf-koşullu olasılıklara ve sınıf etiket önsel dağılımlarına bağlı olarak tanımlanabilmesini sağlamış, böylece sınıf etiket değişkenleri üzerinde tanımlanan önsel dağılımlarla uzamsal bilgi de modele dahil edilebilmiştir. Tezde çokterimli lojistik bağlanım modeli üretici ve ayırmsayıcı olarak iki farklı şekilde kullanılmıştır. Üretici modelde sınıf etiket değişkenleri üzerinde çokterimli özlojistik bağlanım önseli tanımlanarak, piksel etiketlerinin komşu piksel etiketlerine bağlı olması

sağlanmıştır. Sınıf etiketlerinin lokal Markov özelliğinden dolayı komşu piksellerin aynı etikete sahip olması ihtimali yüksek olduğundan, bu izge-uzamsal modellerle daha iyi sınıflandırma haritaları elde edilmektedir. Ayrımsayıcı modelde ise, sınıf etiketleri için bir önsel dağılım tanımlanmamış, ancak sınıf etiket vektörleri de komşu piksellere bağlı olacak şekilde çokterimli lojistik bağlanım modeline dahil edilmiştir. Bu modelde parametre kestirimi hem spektral hem de uzamsal parametreler için yapılmaktadır. Hem üretici modelde hem de ayrımsayıcı modelde yaklaşık ikinci derece optimizasyon yöntemi kullanılmıştır. Bu yöntem, lojistik bağlanım modelinin log-olabilirlik fonksiyonunun Hessian matrisinin alt sınırlı olması nedeniyle, sabit bir yaklaşık alt sınır matrisi kullanılmasına dayanmaktadır.

Evrişimsel sinir ağı modeli için önerilen parametre kestirimi yöntemi, basit ancak etkili bir yöntem olup, ağ derinliğine bağlı olarak değişmemektedir. Bu parametre kestirimi yönteminde bir iterasyonun maliyeti standart geri yayılım algoritmasına göre daha fazla olsa da, eğitim için gereken sürenin sadece geri yayılım algoritmasına göre daha kısa olduğu da deneylerle gösterilmiştir. Bu yöntem yalnız bu problemde değil, son katmanlı bir lojistik bağlanım modeli olan tüm ağlarda kullanılabilir. Tezde hiperspektral imgeler için tek boyutlu evrişimsel katmanlar kullanılmış olmasına karşın, ayrık eğitim ve test kümeleri bulunabilen başka tür imgelerde iki veya üç boyutlu evrişimsel katmanların kullanılması da mümkündür. Ağın son katmanını yine bir çokterimli lojistik bağlanım sınıflandırıcısı olduğundan, evrişim işleminin tek veya çok boyutlu olması önerilen eğitim yönteminin uygulanabilirliğini etkilememektedir.

Yapılan deneylerde hiperspektral imge sınıflandırma literatüründe standart hâle gelmiş ve hâlen kullanılan hiperspektral imgeler kullanılmıştır. İmgelerin eğitim ve test kümelerine rastgele ayrılması nedeniyle testler tekrarlanarak ortalama değerleri raporlanmıştır. Tezde tanıtılan yöntemlerin hepsinde, belli koşullar altında diğer yöntemlerden daha iyi sınıflandırma başarımı elde edildiği gözlenmiştir. Ayrıca evrişimsel sinir ağı modeli için tanıtılmış olan eğitim yöntemi ile, literatürde standart olarak kabul edilen, stokastik en dik iniş tabanlı birinci derece optimizasyon yöntemleri kullanılan geri yayılım algoritmasıyla eğitim yöntemine göre hem iterasyon sayısı, hem de eğitim süresi açısından daha hızlı eğitim yapılabildiği gözlenmiştir.

Tezde yapılan çalışmalar sonucunda, veri kümesinin özelliklerine göre farklı yöntemlerin kullanılmasının daha uygun olduğu sonucuna ulaşılmıştır. Sınırlı sayıda etiketli piksel içeren imgelerde, az sayıda örnekle eğitim yapılabilmesi için Bayesçi karışım modellerinin kullanılması iyi sonuç vermektedir. Verideki karmaşık dağılım karışım modelleri ile

öğrenilebilmekte, parametre kestirimi de tanımlanan önsel dağılımlarla mümkün olabilmektedir. Az örnekle eğitimde parametre kestirimi yapılabilmesi için boyut indirgeme yöntemleri de faydalı olmaktadır.

Literatürde sıkça kullanılmasına karşın, iki ve üç boyutlu evrişimsel sinir ağlarının ve diğer eğiticili uzamsal modellerin kullanılmasında eğitim ve test kümelerinin ayrılmasına büyük özen gösterilmesi model başarımının ölçülmesi açısından büyük önem taşımaktadır. Eğitim ve test kümelerinin örtüşme olmadan ayrılması ve bu kümelerde her sınıftan yeterli sayıda piksel bulunabilmesi durumunda iki ve üç boyutlu evrişimsel sinir ağları ile eğitim aşamasında parametre kestirimi yapılan koşullu rastgele alan gibi yöntemlerin iyi sonuç vermesi beklenmektedir.

Tez kapsamında yapılan çalışmalardan, Bölüm 3.1.2’de tanıtılan uzamsal olarak değişen Gauss karışım modeli (Kayabol ve Kutluk, 2016), Bölüm 3.2’de tanıtılan üretici alt-uzay yöntemi (Kutluk ve diğ., 2016a; 2016b), Bölüm 3.3.1’de tanıtılan yaklaşık çokterimli lojistik bağlanım karışım modeli (Kutluk ve diğ., 2018) ve Bölüm 3.4’te tanıtılan evrişimsel sinir ağ modeli (Kutluk ve diğ., 2019; 2021) konferans bildirisi ve dergi makalesi olarak yayımlanmıştır.

Bu alanda yapılacak çalışmalarda, tezde geliştirilen yöntemlerine devamı niteliğinde, gizli değişken modeli formülasyonu ile başka alt-uzay yöntemlerinin geliştirilmesi, farklı dağılımlarla karışım modellerinin geliştirilmesi, Bölüm 3.3.2’de tanıtılan ayımsayıcı modelin evrişimsel sinir ağıyla birleştirilerek farklı bir izge-uzamsal model elde edilmesi ve evrişimsel sinir ağlarının eğitimi için önerilen yöntemin sınıflara ilişkin katsayıların eğitim kümesinde her sınıftaki örnek sayısı dikkate alınarak ağırlıklandırılacak şekilde geliştirilmesi mümkündür. Uygun veri kümelerinde iki ve üç boyutlu evrişimsel sinir ağları temel alınarak, U-Net ve SegNet gibi mimarilerin kullanılması, bunlarla birlikte Markov rastgele alan, koşullu rastgele alan ve özlojistik bağlanım yöntemlerinin kullanılması ile de iyi sonuçlar elde edilebileceği düşünülmektedir.

KAYNAKLAR

- Agarwal, A., El-Ghazawi, T., El-Askary, H. and Le-Moigne, J., 2007, Efficient hierarchical-PCA dimension reduction for hyperspectral imagery, *IEEE International Symposium on Signal Processing and Information Technology*, 15-18 December 2007 Giza, Egypt, IEEE, 353-356.
- Alam, F. I., Zhou, J., Liew, A. W. and Jia, X., 2016, CRF learning with CNN features for hyperspectral image segmentation, *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 10-15 July 2016 Beijing, China, IEEE, 6890–6893.
- Alam, F.I., Zhou, J., Liew, A. W., Jia, X., Chanussot, J. and Gao, Y., 2019, Conditional random field and deep feature learning for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.*, 57(3), 1612–1628.
- Alipourfard, T., Arefi, H. and Mahmoudi, S., 2018, A novel deep learning framework by combination of subspace-based feature extraction and convolutional neural networks for hyperspectral images classification, *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 22-27 July 2018 Valencia, Spain, IEEE, 4780–4783.
- Alpaydın E., 2013, *Yapay Öğrenme*, Boğaziçi Üniversitesi Yayınevi, İstanbul.
- Audebert, N., Le Saux, B. and Lefevre, S., 2019, Deep learning for classification of hyperspectral data: A comparative review, *IEEE Geoscience and Remote Sensing Magazine*, 7(2), 159–173.
- Bali, N., Mohammad-Djafari, A., 2008, Markov modeling and mean field approximation for hyperspectral data analysis, *IEEE Trans. Image Process.* 17(2), 217–225.
- Bandos, T. V., Bruzzone, L. and Camps-Valls, G., 2009, Classification of hyperspectral images with regularized linear discriminant analysis, *IEEE Trans. Geosci. Remote Sens.*, 47(3), 862-873.
- Banerjee, A., 2007, An Analysis of Logistic Models: Exponential Family Connections and Online Performance, *SIAM International Conference on Data Mining (SDM)*, 26-28 April 2007 Minneapolis, Minesota, USA, 204-215.
- Barber, D., 2012, *Bayesian Reasoning and Machine Learning*, Cambridge University Press.
- Bellman, R., 1961, *Adaptive Control Processes: A Guided Tour*, Princeton University Press.
- Berge A. and Solberg, A. H. S., 2006, Structured Gaussian components for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.*, 44(11), 3386–3396.
- Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N. and Chanussot, J., 2013, Hyperspectral remote sensing data analysis and future challenges, *IEEE Geoscience and Remote Sensing Magazine*, 1 (2), 6-36.

- Bishop, Christopher M., 2006, *Pattern Recognition and Machine Learning*, Springer-Verlag New York, Singapore.
- Bishop, C. M. and Lasserre, J., 2007, Generative or discriminative? Getting the best of both worlds, *Bayesian Statistics* 8, 3-24.
- Blekas, K., Likas, A., Galatsanos, N. P. and Lagaris, I. E., 2005, A spatially constrained mixture model for image segmentation, *IEEE Trans. Neural Netw.*, 16(2), 494–498.
- Bohning, D., 1992, Multinomial logistic regression algorithm, *Ann. Instit. Stat. Math.*, 44, 197–200.
- Borges, J., Bioucas-Dias, J., and Marçal, A., 2006, Fast sparse multinomial regression applied to hyperspectral data, *Int. Conf. Image Analysis and Recognition (ICIAR)*, 18-20 September 2006 Povia de Varzim, Portugal, Springer Berlin, Heidelberg, 700-709.
- Borges, J., Bioucas-Dias, J., and Marçal, A., 2011, Bayesian hyperspectral image segmentation with discriminative class learning, *IEEE Trans. Geosci. Remote Sens.*, 49(6), 2151–2164.
- Botev, A., Ritter, H. and Barber, D., 2017, Practical Gauss-Newton optimisation for deep learning, *34th International Conference on Machine Learning (ICML)*, 06-11 Aug 2017 Sydney, Australia, PMLR 70, 557–565.
- Cao, X., Zhou, F., Xu, L., Meng, D., Xu, Z. and Paisley, J., 2018, Hyperspectral image classification with Markov random fields and a convolutional neural network, *IEEE Transactions on Image Processing*, 27(5), 2354–2367.
- Chen, Y., Jiang, H., Li, C., Jia, X. and Ghamisi, P., 2016, Deep feature extraction and classification of hyperspectral images based on convolutional neural networks, *IEEE Transactions on Geoscience and Remote Sensing*, 54(10), 6232–6251.
- Chen, C., Ma, Y. and Ren, G., 2019, A convolutional neural network with Fletcher–Reeves algorithm for hyperspectral image classification, *Remote Sens.* 11(11), 1325.
- Dempster, A. P., Laird, N. M. and Rubin, D. B., 1977, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, B 39(1), 1–38.
- Diplaros, A., Vlassis, N. A. and Gevers, T., 2007, A spatially constrained generative model and an EM algorithm for image segmentation, *IEEE Trans. Neural Netw.*, 18(3), 798–808.
- Dubey, S.R., Chakraborty, S., Roy, S.K., Mukherjee, S., Singh, S.K. and Chaudhuri, B.B., 2020, Diffgrad: An optimization method for convolutional neural networks, *IEEE Trans. Neural Netw. Learn. Syst.* 31(11), 4500–4511.
- Duchi, J., Hazan, E., Singer, Y., 2011, Adaptive subgradient methods for online learning and stochastic optimization, *J. Mach. Learn. Res.* 12, 2121–2159.
- Eches, O., Dobigeon, N. and Tournieret, J.-Y., 2011, Enhancing hyperspectral image unmixing with spatial correlations, *IEEE Trans. Geosci. Remote Sens.* 49(11), 4239–4247.

- Farrell, M. D. and Mersereau, R. M., 2005, Hyperspectral parameter estimation of elliptically contoured t mixture models using expectation-maximisation, *International Journal of Remote Sensing*, 26(22), 5071–5097.
- Fernandez C. and Green, P. J., 2002, Modelling spatially correlated data via mixtures: A Bayesian approach, *J. R. Stat. Soc. B*, 64(4), 805–826.
- Figueiredo, M. A. T., 2005, Bayesian image segmentation using Gaussian field priors, *Energy Minimization Methods Computer Vision and Pattern Recognition*, LNCS 3757, 74–89.
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B., 2014, *Bayesian Data Analysis*, 3rd edition, CRC Press, Boca Raton, FL.
- Geman, S. and Geman, D., 1984, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.*, 6(6), 721–741.
- Ghahramani, Z., 2015, Probabilistic machine learning and artificial intelligence, *Nature*, 521, 452–459.
- Goodfellow, I., Bengio, Y. and Courville, A., 2016, *Deep Learning*, MIT Press.
- Hagan, M.T. and Menhaj, M.B., 1994, Training feedforward networks with the Marquardt algorithm, *IEEE Trans. Neural Netw.*, 5(6), 989–993.
- Hastie, T., Tibshirani, R. and Friedman, J., 2009, *The Elements of Statistical Learning*, 2nd ed., Springer.
- Haut, J.M., Paoletti, M.E., Plaza, J., Li, J. and Plaza, A., 2018, Active learning with convolutional neural networks for hyperspectral image classification using a new Bayesian approach, *IEEE Trans. Geosci. Remote Sens.*, 56(11), 6440–6461.
- He, K., Zhang, X., Ren, S. and Sun, J., 2015, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, *IEEE International Conference on Computer Vision (ICCV)*, 7–13 December 2015 Santiago, Chile, IEEE, 1026–1034.
- Hotelling, H., 1933, Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology*, 24, 417–441.
- Hu, W., Huang, Y., Wei, L., Zhang, F. and Li, H., 2015, Deep convolutional neural networks for hyperspectral image classification, *J. Sens.*, 2015.
- Hughes, G., 1968, On the mean accuracy of statistical pattern recognizers, *IEEE Trans. Inf. Theory*, 14(1), 55–63.
- Ioffe, S. and Szegedy, C., 2015, Batch normalization: Accelerating deep network training by reducing internal covariate shift, *32nd International Conference on International Conference on Machine Learning (ICML)*, 6–11 July 2015 Lille, France, PMLR 37, 448–456.

- Jia, P., Zhang, M., Yu, W., Shen, F. and Shen, Y., 2016, Convolutional neural network based classification for hyperspectral data, *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 10-15 July 2016 Beijing, China, IEEE, 5075–5078.
- Jordan, M. I., 1995, Why the logistic function? A tutorial discussion on probabilities and neural networks, *Computational Cognitive Science Report 9503*, MIT.
- Kayabol, K., Voisin, A. and Zerubia, J., 2011, SAR image classification with non-stationary multinomial logistic mixture of amplitude and texture densities, *Int. Conf. Image Process. (ICIP)*, 11-14 September 2011 Brussels, IEEE, 173–176.
- Kayabol, K. and Zerubia, J., 2013, Unsupervised amplitude and texture classification of SAR images with multinomial latent model, *IEEE Trans. Image Process.*, 22(2), 561–572.
- Kayabol, K., 2015, Bayesian Gaussian mixture model for spatial-spectral classification of hyperspectral images, *23rd European Signal Processing Conference (EUSIPCO)*, 31 August-4 September 2015 Nice, IEEE, 1805-1809.
- Kayabol, K. and Kutluk, S., 2016, Bayesian classification of hyperspectral images using spatially-varying Gaussian mixture model, *Digital Signal Processing*, 59, 106–114.
- Kayabol, K., 2020, Approximate sparse multinomial logistic regression for classification, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 490-493.
- Kingma, D.P. and Ba, J., 2015, ADAM: A method for stochastic optimization, *3rd International Conference on Learning Representations (ICLR 2015)*, 7–9 May 2015 San Diego, CA, USA.
- Krishnapuram, B., Carin, L., Figueiredo, M. A. T. and Hartemink, A. J., 2005, Sparse multinomial logistic regression: Fast algorithms and generalization bounds, *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(6), 957–968.
- Krizhevsky, A., Sutskever, I. and Hinton, G. E., 2012, ImageNet classification with deep convolutional neural networks, *Advances in neural information processing systems 25 (NIPS 2012)*, 3-8 December 2012 Lake Tahoe, Nevada, Curran Associates, Inc., 1097-1105.
- Kutluk, S., Kayabol, K. and Akan, A., 2016a, Classification of hyperspectral images using mixture of probabilistic PCA models, *24th European Signal Processing Conference (EUSIPCO)*, 29 August-2 September 2016 Budapest, IEEE, 1568-1572.
- Kutluk, S., Kayabol, K. and Akan, A., 2016b, A probabilistic method for the classification of hyperspectral images, *24th Signal Processing and Communication Application Conference (SIU)*, 16-19 May 2016 Zonguldak, IEEE, 905-908.
- Kutluk, S., Kayabol, K. and Akan, A., 2018, A discriminative model for contextual classification of hyperspectral images, *26th Signal Processing and Communications Applications Conference (SIU)*, 2-5 May 2018 Izmir, IEEE, 1-4.

- Kutluk, S., Kayabol, K. and Akan, A., 2019, Spectral-spatial classification of hyperspectral images using CNNs and approximate sparse multinomial logistic regression, *27th European Signal Processing Conference (EUSIPCO)*, 2-6 September 2019 A Coruna, IEEE, 1-5.
- Kutluk, S., Kayabol, K. and Akan, A., 2021, A new CNN training approach with application to hyperspectral image classification, *Digital Signal Processing*, 113, 1-9.
- Lafferty, J., McCallum, A. and Pereira, F. C. N., 2001, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, *Eighteenth International Conference on Machine Learning (ICML)*, 28 June-1 July 2001 Williamstown, MA, USA, Morgan Kaufmann Publishers Inc., 282-289.
- Le, Q.V., Ngiam, J., Coates, A., Lahiri, A., Prochnow, B. and Ng, A.Y., 2011, On optimization methods for deep learning, *28th International Conference on Machine Learning (ICML '11)*, 28 June-2 July 2011 Bellevue, Washington, USA, Omnipress, Madison, WI, USA, 265–272.
- LeCun, Y., 1989, Generalization and network design strategies, *Technical Report CRG-TR-89-4*, University of Toronto.
- Lee H. and Kwon, H., 2016, Contextual deep CNN based hyperspectral classification, *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 10-15 July 2016 Beijing, China, IEEE, 3322–3325.
- Li, W., Prasad, S. and Fowler, J. E., 2014, Hyperspectral image classification using Gaussian mixture models and Markov random fields, *IEEE Geosci. Remote Sens. Lett.*, 11(1), 153-157.
- Li, J., Zhao, X., Li, Y., Du, Q., Xi, B. and Hu, J., 2018, Classification of hyperspectral imagery using a new fully convolutional neural network, *IEEE Geoscience and Remote Sensing Letters*, 15(2), 292–296.
- Liu, L., Jiang, H., He, P., Chen, W., Liu, X., Gao, J. and Han, J., 2020, On the variance of the adaptive learning rate and beyond, *Eight International Conference on Learning Representations (ICLR 2020)*, 26 April-1 May 2020.
- Martens, J., 2010, Deep learning via Hessian-free optimization, *27th International Conference on Machine Learning (ICML '10)*, 21-24 June 2010 Haifa, Israel, Omnipress, Madison, WI, USA, 735–742.
- Martens, J., Grosse, R., 2015, Optimizing neural networks with Kronecker-factored approximate curvature, *32nd International Conference on Machine Learning (ICML)*, 06–11 Jul 2015 Lille, France, PMLR 37, 2408–2417.
- McLachlan, G. J. and Krishnan, T., 2008, *The EM Algorithm and Extensions*, Wiley-Interscience, Hoboken, NJ.
- Møller, M.F., 1993, A scaled conjugate gradient algorithm for fast supervised learning, *Neural Netw.*, 6(4), 525–533.

- Murphy, K. P., 2012, *Machine Learning: A Probabilistic Perspective*, MIT Press.
- Nalepa, J., Myller, M. And Kawulok, M., 2019, Validating hyperspectral image segmentation, *IEEE Geoscience and Remote Sensing Letters*, 16(8), 1264–1268.
- Neal, R. M., 1996, *Bayesian Learning for Neural Networks*, Springer, New York, NY, ISBN: 978-1-4612-0745-0.
- Ng, A. and Jordan, M., 2002, On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes, *Advances in neural information processing systems*, 14, MIT Press, 841-848.
- Pearson, K., 1901, On lines and planes of closest fit to systems of points in space, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Series 6, 2, 559–572.
- Platt, J. C., 1999, Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, *Advances in Large Margin Classifiers*, 10(3), 61–74.
- Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., Marconcini, M., Tilton, J. C., and Trianni, G., 2009, Recent advances in techniques for hyperspectral image processing, *Remote Sensing of Environment*, 113(Supplement 1), 110–122.
- Quarteroni, A., Sacco, R., and Saleri, F., 2000, *Numerical Mathematics*, Springer-Verlag, New York.
- Rodarmel, C. and Shan, J., 2002, Principal component analysis for hyperspectral image classification, *Surveying and Land Information Systems*, 62(2), 115-123.
- Rumelhart, D. E., Hinton, G. E. and Williams, R. J., 1986. Learning representations by backpropagating errors, *Nature*, 323, 533–536.
- Sanjay-Gopal, S. and Hebert, T. J., 1998, Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm, *IEEE Trans. Image Process.*, 7(7), 1014–1028.
- Sfikas, G., Nikou, C. and Galatsanos, N., 2007, Robust image segmentation with mixtures of Student's t-distributions, *IEEE International Conference on Image Processing (ICIP)*, 16 September-19 October 2007 San Antonio, TX, USA, IEEE, I-273–I-276.
- Tadjudin, S. and Landgrebe, D., 1999, Covariance estimation with limited training samples, *IEEE Trans. Geosci. Remote Sens.*, 37(4), 2113–2118.
- Tan, H.H., Lim, K.H. and Harno, H.G., 2017, Stochastic diagonal approximate greatest descent in convolutional neural networks, *IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 12-14 September 2017 Kuching, Malaysia, IEEE, 451–454.

- Tan, H.H., and Lim, K.H., 2019, Review of second-order optimization techniques in artificial neural networks backpropagation, *IOP Conf. Ser.: Mater. Sci. Eng.* 495.
- Tarabalka, Y., Fauvel, M., Chanussot, J. and Benediktsson, J.A., 2010, SVM and MRF-based method for accurate classification of hyperspectral images, *IEEE Geoscience and Remote Sensing Letters*, 7(4), 736–740.
- Tipping, M. E. and Bishop, C. M., 1999, Probabilistic principal component analysis, *Journal of the Royal Statistical Society*, B21(3), 611–622.
- Trianni, G. and Gamba, P., 2005, A novel Markov random field model for multi source data fusion in urban areas, *XXVIIIth URSI General Assembly*, October 2005 New Delhi.
- Veracini, T., Matteoli, S., Diani, M. and Corsini, G., 2010, Robust hyperspectral image segmentation based on a non-Gaussian model, *2nd International Workshop on Cognitive Information Processing (CIP)*, 14-16 June 2010 Elba, Italy, 192–197.
- Wang, C., Zhang, L., Wei, W. and Zhang, Y., 2019, Hyperspectral image classification with data augmentation and classifier fusion, *IEEE Geosci. Remote Sens. Lett.*, 17(8), 1420–1424.
- Wilamowski, B. M. and Yu, H., 2010a, Improved computation for Levenberg–Marquardt training, *IEEE Trans. Neural Netw.*, 21(6), 930–937.
- Wilamowski, B. M. and Yu, H., 2010b, Neural network learning without backpropagation, *IEEE Trans. Neural Netw.*, 21(11), 1793–1803.
- Xia, J., Chanussot, J., Du, P. and He, X., 2014, (Semi-) supervised probabilistic principal component analysis for hyperspectral remote sensing image classification, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6), 2225–2237.
- Yue, J., Zhao, W., Mao, S. and Liu, H., 2015, Spectral–spatial classification of hyperspectral images using deep convolutional neural networks, *Remote Sens. Lett.* 6(6), 468–477.
- Zeiler, M.D., 2012, Adadelata: An adaptive learning rate method, *arXiv preprint*, arXiv:1212.5701.
- Zhang, Y., Brady, M. and Smith, S., 2001, Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm, *IEEE Trans. Medical Imaging*, 20(1), 45–57.
- Zhang, H., Li, Y., Jiang, Y., Wang, P., Shen, Q. and Shen, C., 2019, Hyperspectral classification based on lightweight 3-D-CNN with transfer learning, *IEEE Trans. Geosci. Remote Sens.*, 57(8), 5813–5828.
- Zhong, Z., Li, J., Clausi, D. A. and Wong, A., 2020, Generative adversarial networks and conditional random fields for hyperspectral image classification, *IEEE Trans. Cybern.* 50(7), 3318–3329.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F. and Fraundorfer, F., 2017, Deep learning in remote sensing: A comprehensive review and list of resources, *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36.



EKLER

EK-1. ÇOKTERİMLİ STUDENT-T DAĞILIMININ PARAMETRE KESTİRİMİ

Bir pikselin sınıf-koşullu olasılık dağılımı şöyle yazılabilir:

$$p(\mathbf{s}_n | z_{n,k} = 1, \mathbf{m}_k, \mathbf{\Sigma}_k, \eta_k) = \int_0^\infty \mathcal{N}(\mathbf{s}_n | \mathbf{m}_k, \frac{\mathbf{\Sigma}_k}{\tau_n}) \text{Gamma}(\tau_n | \frac{\eta_k}{2}, \frac{\eta_k}{2}) d\tau_n \quad (\text{E.1})$$

Tam olabilirlik fonksiyonu şöyle yazılabilir:

$$\begin{aligned} \prod_{n=1}^{N_k} p(\mathbf{s}_n, \tau_n | z_{n,k} = 1, \mathbf{m}_k, \mathbf{\Sigma}_k, \eta_k) &= \prod_{n=1}^{N_k} \mathcal{N}(\mathbf{s}_n | \mathbf{m}_k, \frac{\mathbf{\Sigma}_k}{\tau_n}) \text{Gamma}(\tau_n | \frac{\eta_k}{2}, \frac{\eta_k}{2}) \\ &= \prod_{n=1}^{N_k} \frac{1}{(2\pi)^{\frac{L}{2}} |\frac{\mathbf{\Sigma}_k}{\tau_n}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{s}_n - \mathbf{m}_k)^T \left(\frac{\mathbf{\Sigma}_k}{\tau_n} \right)^{-1} (\mathbf{s}_n - \mathbf{m}_k) \right\} \\ &\quad \times \frac{1}{\Gamma(\frac{\eta_k}{2})} \left(\frac{\eta_k}{2} \right)^{\frac{\eta_k}{2}} \tau_n^{\frac{\eta_k}{2}-1} \exp \left\{ -\frac{\eta_k}{2} \tau_n \right\} \end{aligned} \quad (\text{E.2})$$

Tam log-olabilirlik fonksiyonu şöyle yazılabilir:

$$\begin{aligned} \sum_{n=1}^{N_k} \ln p(\mathbf{s}_n, \tau_n | z_{n,k} = 1, \mathbf{m}_k, \mathbf{\Sigma}_k, \eta_k) &= \sum_{n=1}^{N_k} -\frac{L}{2} \ln 2\pi - \frac{1}{2} \ln |\mathbf{\Sigma}_k| + \frac{L}{2} \ln \tau_n \\ &\quad - \frac{1}{2} (\mathbf{s}_n - \mathbf{m}_k)^T \left(\frac{\mathbf{\Sigma}_k}{\tau_n} \right)^{-1} (\mathbf{s}_n - \mathbf{m}_k) - \ln \Gamma\left(\frac{\eta_k}{2}\right) + \frac{\eta_k}{2} \ln \frac{\eta_k}{2} \\ &\quad + \left(\frac{\eta_k}{2} - 1 \right) \ln \tau_n - \frac{\eta_k}{2} \tau_n \end{aligned} \quad (\text{E.3})$$

Beklenen değerin hesaplanması için gizli değişkenin sonsal dağılımının hesaplanması gerekmektedir. Bu sonsal dağılım, Bayes kuralından yararlanılarak şöyle yazılabilir:

$$\prod_{n=1}^{N_k} p(\tau_n | \mathbf{s}_n, z_{n,k} = 1, \mathbf{m}_k, \mathbf{\Sigma}_k, \eta_k) \propto \prod_{n=1}^{N_k} p(\mathbf{s}_n, \tau_n | z_{n,k} = 1, \mathbf{m}_k, \mathbf{\Sigma}_k, \eta_k) \quad (\text{E.4})$$

Gamma dağılımı, normal dağılım için eşlenik önsel dağılım olduğundan, gizli değişkenin sonsal dağılımı da Gamma dağılımıdır:

$$\prod_{n=1}^{N_k} p(\tau_n | \mathbf{s}_n, z_{n,k} = 1, \mathbf{m}_k, \mathbf{\Sigma}_k, \eta_k) \propto \prod_{n=1}^{N_k} \text{Gamma}(\tau_n | \alpha_n, \beta_n) \quad (\text{E.5})$$

Dağılımların açık ifadeleriyle, bu oransallık ilişkileri şöyle yazılabilir:

$$\begin{aligned} & \prod_{n=1}^{N_k} \frac{\beta_n^{\alpha_n} \tau_n^{\alpha_n-1} e^{-\beta_n \tau_n}}{\Gamma(\alpha_n)} \\ & \propto \prod_{n=1}^{N_k} \frac{1}{(2\pi)^{\frac{L}{2}} \left| \frac{\mathbf{\Sigma}_k}{\tau_n} \right|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{s}_n - \mathbf{m}_k)^T \left(\frac{\mathbf{\Sigma}_k}{\tau_n} \right)^{-1} (\mathbf{s}_n - \mathbf{m}_k) \right\} \\ & \quad \times \frac{1}{\Gamma\left(\frac{\eta_k}{2}\right)} \left(\frac{\eta_k}{2}\right)^{\frac{\eta_k}{2}} \tau_n^{\frac{\eta_k}{2}-1} \exp \left\{ -\frac{\eta_k}{2} \tau_n \right\} \end{aligned} \quad (\text{E.6})$$

τ_n 'e bağlı olan terimlerden α_n ve β_n parametreleri çözümlürse, gizli değişken sonsalı şöyle elde edilir:

$$\begin{aligned} & p(\tau_n | \mathbf{s}_n, z_{n,k} = 1, \mathbf{m}_k, \mathbf{\Sigma}_k, \eta_k) \\ & = \text{Gamma} \left(\tau_n \left| \frac{L + \eta_k}{2}, \frac{(\mathbf{s}_n - \mathbf{m}_k)^T \mathbf{\Sigma}_k (\mathbf{s}_n - \mathbf{m}_k) + \eta_k}{2} \right. \right) \end{aligned} \quad (\text{E.7})$$

Beklenti adımımda, bir önceki adımda bulunan parametrelerle gizli değişken sonsal dağılımı hesaplanır ve bu dağılıma göre tam log-olabilirlik fonksiyonunun beklenen değeri hesaplanır. Bu şöyle ifade edilebilir:

$$\begin{aligned}
& Q(\theta, \theta^{t-1}) \\
&= \int_{\tau_n} p(\tau_n | \mathbf{s}_n, z_{n,k} = 1, \mathbf{m}_k^{t-1}, \mathbf{\Sigma}_k^{t-1}, \eta_k^{t-1}) \sum_{n=1}^{N_k} \ln p(\mathbf{s}_n, \tau_n | z_{n,k} \\
&= 1, \mathbf{m}_k, \mathbf{\Sigma}_k, \eta_k) \\
&= \int_{\tau_n} \text{Gamma}\left(\tau_n \middle| \frac{L + \eta_k}{2}, \frac{(\mathbf{s}_n - \mathbf{m}_k)^T \mathbf{\Sigma}_k (\mathbf{s}_n - \mathbf{m}_k) + \eta_k}{2}\right) \sum_{n=1}^{N_k} -\frac{L}{2} \ln 2\pi \\
&\quad - \frac{1}{2} \ln |\mathbf{\Sigma}_k| + \frac{L}{2} \ln |\tau_n| - \frac{1}{2} (\mathbf{s}_n - \mathbf{m}_k)^T \tau_n \mathbf{\Sigma}_k^{-1} (\mathbf{s}_n - \mathbf{m}_k) - \ln \Gamma\left(\frac{\eta_k}{2}\right) + \frac{\eta_k}{2} \ln \frac{\eta_k}{2} \quad (\text{E.8}) \\
&\quad + \left(\frac{\eta_k}{2} - 1\right) \ln \tau_n - \frac{\eta_k}{2} \tau_n \\
&= -\frac{N_k L}{2} \ln 2\pi - \frac{N_k}{2} \ln |\mathbf{\Sigma}_k| + \frac{L}{2} \sum_{n=1}^{N_k} \mathbb{E}[\tau_n] \\
&\quad - \frac{1}{2} \sum_{n=1}^{N_k} \mathbb{E}[\tau_n] (\mathbf{s}_n - \mathbf{m}_k)^T \mathbf{\Sigma}_k^{-1} (\mathbf{s}_n - \mathbf{m}_k) - N_k \ln \Gamma\left(\frac{\eta_k}{2}\right) + \frac{N \eta_k}{2} \ln \frac{\eta_k}{2} \\
&\quad + \left(\frac{\eta_k}{2} - 1\right) \sum_{n=1}^{N_k} \mathbb{E}[\ln \tau_n] - \frac{\eta_k}{2} \sum_{n=1}^{N_k} \mathbb{E}[\tau_n]
\end{aligned}$$

Burada θ^{t-1} bir önceki adımda bulunan parametre kümesini göstermektedir, dolayısıyla $\theta^{t-1} = \{\mathbf{m}_k^{t-1}, \mathbf{\Sigma}_k^{t-1}, \eta_k^{t-1}\}$ yazılabilir.

En büyükleme adımında $Q(\theta, \theta^{t-1})$ fonksiyonunu en büyükleyen parametreler bulunmaktadır. Parametreler için önsel dağılımlar tanımlanması durumunda $Q(\theta, \theta^{t-1}) + \ln p(\theta)$ toplamı en büyüklenmektedir (Bishop, 2006).

Bu çalışmada ortalama ve kovaryans matrisi için normal-ters Wishart önsel dağılımı kullanılmıştır. Bu önsel dağılım şöyle yazılabilir:

$$p(\mathbf{m}_k, \mathbf{\Sigma}_k) = \mathcal{N}\left(\mathbf{m}_k \middle| \mathbf{m}_0, \frac{\mathbf{\Sigma}_k}{\lambda}\right) W^{-1}(\mathbf{\Psi}, \nu) \quad (\text{E.9})$$

Log-önsel dağılım şöyle yazılabilir:

$$\begin{aligned}
\ln p(\mathbf{m}_k, \mathbf{\Sigma}_k) = & -\frac{L}{2} \ln 2\pi + \frac{L}{2} \ln \lambda - \frac{1}{2} \ln |\mathbf{\Sigma}_k| \\
& - \frac{1}{2} (\mathbf{m}_k - \mathbf{m}_0)^T \lambda \mathbf{\Sigma}_k^{-1} (\mathbf{m}_k - \mathbf{m}_0) + \frac{\nu}{2} \ln |\mathbf{\Psi}| - \frac{\nu L}{2} \ln 2 \\
& - \ln \Gamma\left(\frac{\nu}{2}\right) - \frac{\nu + L + 1}{2} \ln |\mathbf{\Sigma}_k| - \frac{1}{2} \text{tr}(\mathbf{\Psi} \mathbf{\Sigma}_k^{-1})
\end{aligned} \tag{E.10}$$

$Q(\theta, \theta^{t-1}) + \ln p(\theta)$ toplamının $\mathbf{m}_k$ ve $\mathbf{\Sigma}_k$ parametrelerine göre türevi alınarak, bu parametrelerin kestirimi şöyle elde edilebilir:

$$\begin{aligned}
\hat{\mathbf{m}}_k &= \frac{\lambda \mathbf{m}_0 + \sum_{n=1}^N \mathbb{E}[\tau_n] \mathbf{s}_n}{\lambda + \sum_{n=1}^N \mathbb{E}[\tau_n]} \\
\hat{\mathbf{\Sigma}}_k &= \frac{\lambda (\mathbf{m}_k - \mathbf{m}_0)(\mathbf{m}_k - \mathbf{m}_0)^T + \mathbf{\Psi}^T + \sum_{n=1}^N \mathbb{E}[\tau_n] (\mathbf{s}_n - \mathbf{m}_k)(\mathbf{s}_n - \mathbf{m}_k)^T}{N_k + \eta_k + L + 2}
\end{aligned} \tag{E.11}$$

τ_n 'in beklenen değeri ise şöyle hesaplanır:

$$\mathbb{E}[\tau_n] = \frac{L + \eta_0}{\eta_0 + (\mathbf{s}_n - \mathbf{m}_0)^T \mathbf{\Sigma}_k^{-1} (\mathbf{s}_n - \mathbf{m}_k)} \tag{E.12}$$

En büyükleme adımında kestirilen parametreler sonraki iterasyondaki beklenti hesaplama adımında kullanılır ve algoritma yakınsayana kadar tekrarlanır.

EK-2. SINIF-KOŞULLU DAĞILIMLARIN NORMALİZE EDİLMİŞ ÜSTEL DAĞILIMLAR ŞEKLİNDE TANIMLANMASI İLE ÇOKTERİMLİ LOJİSTİK BAĞLANIM SINIFLANDIRICISININ DENKLİĞİ

Çokterimli lojistik bağlanım modeli bir ayımsayıcı model olup, bir $\mathbf{s}_n$ pikseli ve bu pikselin etiket vektörü $\mathbf{z}_n$ için, doğrudan $p(\mathbf{z}_n | \mathbf{s}_n, \boldsymbol{\omega}_{1:K})$ dağılımını modellemektedir. $\mathbf{s}_n$ pikselinin k . sınıftan olması olasılığı şöyle hesaplanmaktadır:

$$p(z_{n,k} = 1 | \mathbf{s}_n, \boldsymbol{\omega}_k) = \frac{e^{\boldsymbol{\omega}_k^T \mathbf{s}_n}}{\sum_{j=1}^K e^{\boldsymbol{\omega}_j^T \mathbf{s}_n}} \tag{E.13}$$

Kayabol (2020) çokterimli lojistik bağlanım modelini sınıf-köşullu olasılıkları ve etiket önsellerini açık şekilde yazarak, üretici bir model olarak tanımlamıştır. Bu modelde sınıf-köşullu olasılıklar normalize edilmiş üstel dağılımlar ile tanımlanmış olup şöyle yazılabilir:

$$p(\mathbf{s}_n | z_{n,k} = 1, \boldsymbol{\omega}_k) = \frac{e^{\boldsymbol{\omega}_k^T \mathbf{s}_n}}{\sum_{j=1}^K e^{\boldsymbol{\omega}_j^T \mathbf{s}_n}} \quad (\text{E.14})$$

Sınıf etiketleri üzerinde kategorik önsel dağılım tanımlanmıştır ve sınıf etiketleri bağımsız kabul edildiğinden, bu etiketlerin bileşik dağılımı şöyle yazılabilir:

$$p(z_{1:N} | \mathbf{q}) = \prod_{n=1}^N \prod_{k=1}^K q_k^{z_{n,k}} \quad (\text{E.15})$$

Bu dağılımın parametreleri $\mathbf{q} = \{q_1, \dots, q_K\}$ olup, bu parametreler her sınıfın olasılığını göstermektedir:

$$p(z_{n,k} = 1 | \mathbf{q}) = q_k \quad (\text{E.16})$$

Denklem (E.14) ve (E.15)'teki sınıf-koşullu olasılık ile önsel dağılım kullanılarak, sınıf etiketlerinin sonsal dağılımı şöyle yazılabilir:

$$\begin{aligned} p(z_{n,k} = 1 | \mathbf{s}_n, \boldsymbol{\omega}_{1:K}, \mathbf{q}) &= \frac{p(\mathbf{s}_n | z_{n,k} = 1, \boldsymbol{\omega}_{1:K}) p(z_{n,k} = 1 | \mathbf{q})}{p(\mathbf{s}_n | \boldsymbol{\omega}_{1:K})} \\ &= \frac{p(\mathbf{s}_n | z_{n,k} = 1, \boldsymbol{\omega}_{1:K}) p(z_{n,k} = 1 | \mathbf{q})}{\sum_{l=1}^K [p(\mathbf{s}_n | z_{n,l} = 1, \boldsymbol{\omega}_{1:K}) p(z_{n,l} = 1 | \mathbf{q})]} \\ &= \frac{\frac{e^{\boldsymbol{\omega}_k^T \mathbf{s}_n}}{\sum_{j=1}^K e^{\boldsymbol{\omega}_j^T \mathbf{s}_n}} q_k}{\sum_{l=1}^K \left[\frac{e^{\boldsymbol{\omega}_l^T \mathbf{s}_n}}{\sum_{j=1}^K e^{\boldsymbol{\omega}_j^T \mathbf{s}_n}} q_l \right]} = \frac{\frac{e^{\boldsymbol{\omega}_k^T \mathbf{s}_n}}{\sum_{j=1}^K e^{\boldsymbol{\omega}_j^T \mathbf{s}_n}} q_k}{\frac{1}{\sum_{j=1}^K e^{\boldsymbol{\omega}_j^T \mathbf{s}_n}} \sum_{l=1}^K e^{\boldsymbol{\omega}_l^T \mathbf{s}_n} q_l} \\ &= \frac{e^{\boldsymbol{\omega}_k^T \mathbf{s}_n} q_k}{\sum_{l=1}^K e^{\boldsymbol{\omega}_l^T \mathbf{s}_n} q_l} \end{aligned} \quad (\text{E.17})$$

Sınıfların eşit olasılıklı olması durumunda her k için $q_k = q$ yazılırsa,

$$\begin{aligned} p(z_{n,k} = 1 | \mathbf{s}_n, \boldsymbol{\omega}_{1:K}, \mathbf{q}) &= \frac{e^{\boldsymbol{\omega}_k^T \mathbf{s}_n} q_k}{\sum_{l=1}^K e^{\boldsymbol{\omega}_l^T \mathbf{s}_n} q_l} = \frac{e^{\boldsymbol{\omega}_k^T \mathbf{s}_n} q}{\sum_{l=1}^K e^{\boldsymbol{\omega}_l^T \mathbf{s}_n} q} = \frac{e^{\boldsymbol{\omega}_k^T \mathbf{s}_n} q}{q \sum_{l=1}^K e^{\boldsymbol{\omega}_l^T \mathbf{s}_n}} \\ &= \frac{e^{\boldsymbol{\omega}_k^T \mathbf{s}_n}}{\sum_{l=1}^K e^{\boldsymbol{\omega}_l^T \mathbf{s}_n}} \end{aligned} \quad (\text{E.18})$$

Denklem (E.18)'de görüldüğü gibi, sınıf-koşullu olasılıkların normalize edilmiş üstel dağılıma sahip olduğu, sınıf etiketlerinin ise eşit olasılıklı kategorik dağılıma sahip olduğu durumda çokterimli lojistik bağlanım modeli elde edilmektedir

EK-3. SÖZLÜK

İngilizce	Türkçe
Batch normalization	Grup normalizasyon
Clustering	Öbekleme
Discriminant	Ayrıtaç
Discriminative	Ayrımsayıcı
Generative	Üretici
Likelihood	Olabilirlik
Machine learning	Yapay öğrenme
Maximization	En büyükleme
Outlier	Aykırı
Pooling layer	Havuz katmanı
Posterior	Sonsal
Prior	Önsel
Pseudo likelihood	Sözde olabilirlik
Regularization	Düzenlileştirme
Robust	Gürbüz
Singular value decomposition	Tekil matris ayrıştırma
Smoothing	Düzleştirme
Spatial	Uzamsal
Spectral	İzgesel
Subgradient	Alt-gradyan