

## ÖDEV 5

Selen Erdoğan  
selenerdogan2019@gtu.edu.tr  
Elektronik Mühendisliği Bölümü, GTÜ

### I. GİRİŞ

Bu çalışmada, spektral veri analizi için Ana Bileşen Analizi'nin (PCA) önemi ve uygulanmasının sonuçları üzerine odaklanılmıştır. PCA, veri madenciliği ve görüntü işleme alanlarında boyut azaltma ve kümeleme gibi birçok uygulamada kullanılan güçlü bir yöntemdir. Bu ödevde, PCA'nın spektral veriler üzerinde nasıl kullanılabileceğini ve bu uygulamaların sonuçlarını incelemektedir.

Spektral veriler, uzaktan algılama ve çevresel bilimler gibi birçok alanda önemlidir. Bu tür veriler, bir nesnenin farklı dalga boylarındaki yansımalarını kaydeder ve her dalga boyunda bir değer içerir. Ancak, bu çok boyutlu veri setleri genellikle karmaşık ve işlenmesi zor olabilir. Bu nedenle, veriyi daha yönetilebilir bir hale getirmek ve önemli bilgileri çıkarmak için boyut azaltma teknikleri gereklidir.

PCA, yani Temel Bileşen Analizi (Principal Component Analysis), çok değişkenli verilerdeki desenleri tespit etmek ve verileri daha az sayıda temel bileşene indirgemek için kullanılan bir yöntemdir. Temel amaç, çok sayıda özelliğe sahip veri setlerini, verinin asıl yapısını olabildiğince koruyarak daha az sayıda yapay değişkenle ifade etmektir. Bu yapay değişkenlere "temel bileşenler" denir ve her biri, veri setindeki maksimum varyansı yakalamaya çalışır.

PCA'nın temel adımları şunlardır

- Standartlaştırma: Veri setindeki her bir değişkenin ortalama değerini çıkarıp, standart sapmaya böleriz. Böylece her bir değişkenin ortalaması 0 ve standart sapması 1 olur. Bu işlem, farklı ölçeklere sahip değişkenler arasında karşılaştırılabilirlik sağlar.
- Kovaryans Matrisinin Hesaplanması: Standartlaştırılmış veri setinin kovaryans matrisi hesaplanır. Bu matris, değişkenler arasındaki varyans ve kovaryans değerlerini içerir.
- Özdeğerlerin ve Özvektörlerin Hesaplanması: Kovaryans matrisinin özdeğerleri ve özvektörleri hesaplanır. Özdeğerler, her temel bileşenin veride ne kadar varyans taşıdığını gösterir, özvektörler ise bu bileşenlerin yönünü belirtir.
- Temel Bileşenlerin Sıralanması: Özdeğerler büyükten küçüğe sıralanır ve özvektörler buna göre düzenlenir. Bu sıralama, temel

bileşenlerin veri setindeki önem sırasını gösterir.

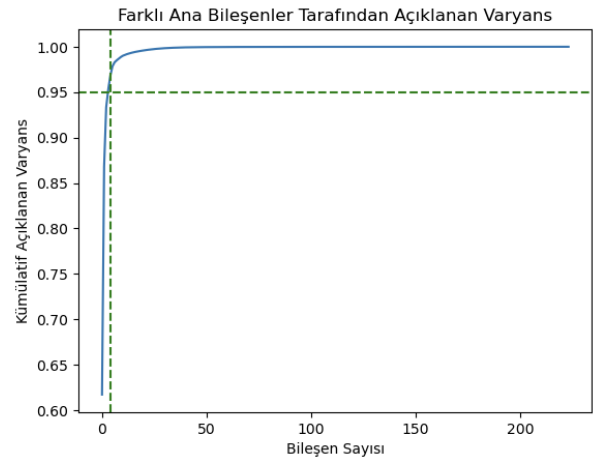
### II. KULLANIM

PCA'nın uygulandığı örnekte, veri seti bir hiper-spektral görüntüdür ve her bir piksel farklı spektral bantlarda ölçülen yoğunlukları içerir. PCA bu veri setindeki boyut sayısını azaltarak, verinin büyük kısmını oluşturan varyansı korurken, hesaplama ve analiz kolaylığı sağlamak için veriyi daha az sayıda temel bileşene indirger. Çıktı grafiği, her bir bileşenin kümülatif olarak ne kadar varyans açıkladığını gösterir ve bu sayede hangi bileşen sayısının seçilmesi gerektiğine karar vermemize yardımcı oldu.

### III. ALGORITMANIN UYGULANMASI

Yazılan kod ile bir hiperspektral görüntü veri seti olan 'Salinas' üzerinde ana bileşen analizi (PCA) gerçekleştirildi ve kümülatif varyansın bir grafik üzerinde görselleştirilmesini sağlandı. İşlemin temel adımları veri yükleme, yeniden şekillendirme, standartlaştırma, kovaryans matrisi hesaplama, özdeğerlerin ve özvektörlerin bulunması, kümülatif varyans hesaplama ve varyansın belirli bir yüzdeyi geçtiği bileşen sayısını bulma şeklindedir. Amaç, veri boyutunu azaltırken bilgi kaybını minimize etmektir.

224 boyuttan 4 boyuta indirildi ve %95 varyans korundu.



Şekil 1

Şekil 1, PCA (Ana Bileşen Analizi) kullanılarak gerçekleştirilen boyut azaltmanın sonuçlarını gösteren bir

# ELM472 Makine Öğrenmesinin Temelleri

kümülatif açıklanan varyans grafiğini temsil etmekte. Grafik, verinin farklı ana bileşen sayılarına göre ne kadar varyans açıkladığını göstermektedir.

Salinas adlı bir spektral veri seti (veri) yüklendi.

Veri, her pikselin ve her spektral bandın bir sütun olarak temsil edildiği bir 2D diziye yeniden şekillendirildi.

Veri, standartlaştırma işlemi uygulanarak ölçeklendirildi. Bu, her bandın ortalama değerinin 0 ve standart sapmasının 1 olmasını sağladı.

Standartlaştırılmış veri üzerinde bir kovaryans matrisi hesaplandı.

Kovaryans matrisinin özdeğerleri ve özvektörleri hesaplandı ve özdeğerler azalan sırada sıralandı.

Her bileşenin açıkladığı kümülatif varyans hesaplandı.

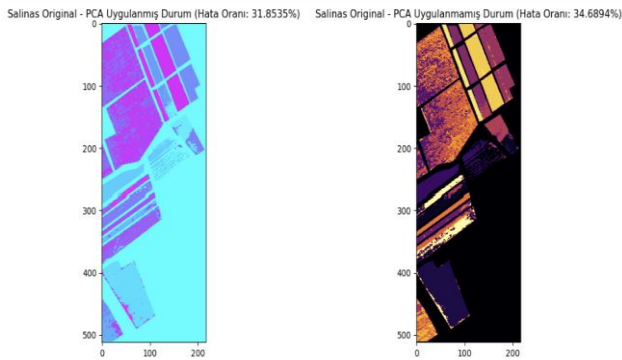
Belirli bir varyans eşiğini karşılayacak bileşen sayısı belirlendi.

Belirlenen bileşen sayısı kadar en üstteki ana bileşenler seçildi.

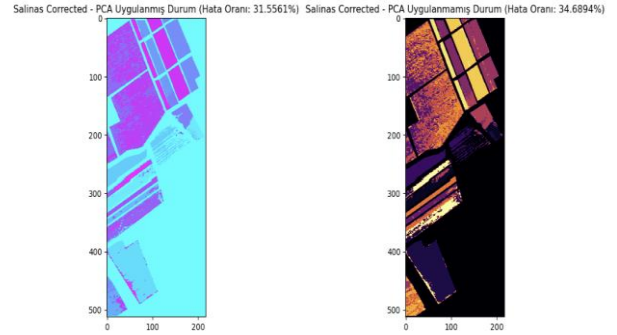
Veri, bu seçilen ana bileşenlere göre yeni bir alt uzaya dönüştürüldü.

Son olarak, boyut azaltmanın ne kadar yapıldığı ve neden yapıldığı yazıldı ve isteğe bağlı olarak kümülatif açıklanan varyans bir grafik olarak çizildi.

Bu kod, spektral veri setinin boyutunu azaltarak verinin önemli bilgilerini korumayı amaçlayan bir veri madenciliği veya görüntü işleme uygulaması için kullanılabilir.



Şekil 2



Şekil 3

Hata oranının düştüğü gözlemlenmiştir.

## IV. SONUÇLAR

Bu çalışmada, PCA'nın Salinas spektral veri seti üzerinde nasıl uygulandığını ve sonuçlarını ele aldık. Özellikle, PCA'nın hata oranları üzerindeki etkisini inceledik. PCA uygulandığında ve uygulanmadığında hata oranlarının nasıl değiştiğini gözlemekteyiz. Elde edilen sonuçlar, PCA'nın spektral veri analizi için önemini vurgulamaktadır. PCA'nın uygulanmasıyla hata oranlarının düştüğünü ve verinin daha iyi kümeleme sonuçları verdiğini gözlemlemekteyiz.

Bu çalışma, spektral veri analizinde PCA'nın önemini ve uygulama sonuçları gözlemlenmiştir. PCA, veriyi daha yönetilebilir hale getirirken önemli bilgileri korumanın etkili bir yolunu sunar. Elde edilen sonuçlar, PCA'nın spektral veri analizinde kullanılmasının faydalarını göstermektedir. Hata oranlarının düşüşünü gözlemlemiş olduk.

## KAYNAKÇA

- [1] <https://medium.com/machine-learning-t%C3%BCrkiye/pca-b4d745de33f3>