

# ELM472 Makine Öğrenmesinin Temelleri

## PCA Kullanarak Boyut Azaltma Ödev-5

Selimhan Aygün  
s.aygun2019@gtu.edu.tr  
Elektronik Mühendisliği Bölümü, GTÜ, Kocaeli, Türkiye

### ÖZET

Bu çalışmada PCA metodu kullanılarak verilen verinin boyutları azaltılmış, çıkan sonucun öbekleme algoritmasıyla karşılaştırması yapılmıştır.

### GİRİŞ

PCA tanıma, sınıflandırma, görüntü sıkıştırma alanlarında kullanılan yararlı bir istatistiksel tekniktir. Temel amacı yüksek boyutlu verilerde en yüksek varyans ile veri setini tutmak ancak bunu yaparken boyut indirgemeyi sağlamak olan bir tekniktir. Fazla boyutlu verilerdeki genel özellikleri bularak boyut sayısının azaltılmasını, verinin sıkıştırılmasını sağlar.

PCA'nın arkasında yatan temel mantık çok boyutlu bir veriyi, verideki temel özellikleri yakalayarak daha az sayıda değişkenle göstermektir. Uzayda bir noktalar kümesi için, tüm noktalara ortalama uzaklığı en az olan "en uygun doğru" seçilir. Daha sonra bu doğruya dik olanlar arasından yine en uygun doğru seçilerek, bu adımlar, yeni bir boyutun varyansı belirli bir eşiğin altına inene kadar tekrarlanır. Bu sürecin sonunda elde edilen doğrular, bir doğrusal uzayın tabanlarını oluşturur. Bu taban vektörlerine temel bileşen denir. Verinin temel bileşenleri birbirinden bağımsız olur.

### TEORİ VE YÖNTEM

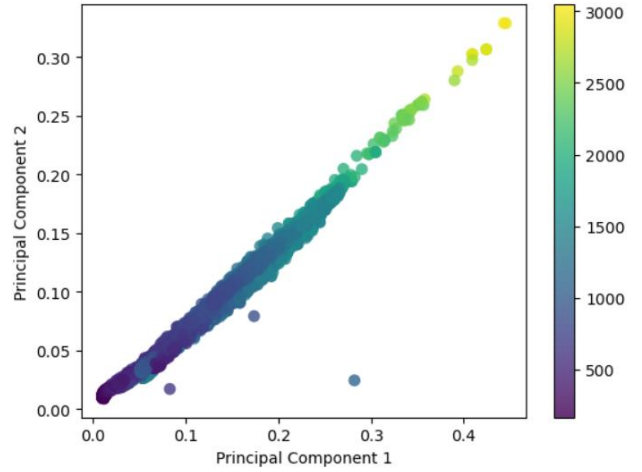
İlk olarak alınan verilerin boyutu *reshape()* fonksiyonu ile değiştirildi ve PCA analizi için kullanılabilir hale getirildi. Ardından verilerin normalizasyonu yapıldı. Değişkenlerin birbirlerine göre ortalamadan nasıl değiştiğini anlamak, başka bir deyişle aralarında bir ilişki olup olmadığını görmek için kovaryans matrisi hesabı yapıldı. Değişkenlerin gereksiz bilgi içerecek şekilde yüksek korelasyona sahip olmasına karşın bu işlem gerçekleştirildi. Bu korelasyonları belirlemek için kovaryans matrisini hesaplandı.

Ardından, özdeğer ve özvektörler hesaplandı. En büyük özdeğerle ilişkili özvektör, verinin en fazla varyansa sahip olduğu yönü gösterir. Dolayısıyla, özdeğerleri kullanarak, verilerdeki en fazla değişkenliği hangi özvektörlerin

yakaladığı bilinir. Özvektörler, özdeğerlerine göre azalan sırada sıralanır ve ilk k adet özvektör seçilir, bu yeni boyutu belirler. İç çarpım yapılarak orijinal, n boyutlu verinin k boyuta dönüşümü (projeksiyonu) yapılır.

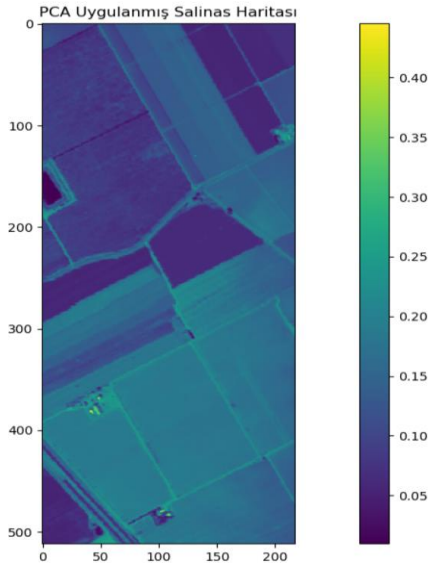
$$Ax - \lambda x = 0$$
$$(A - \lambda I)x = 0$$

(1.1)



Şekil 1 PCA Uygulanmış Veri

Şekil 1' de görülen yeniden şekillendirilmiş veriye PCA işlemi uygulanarak boyutu azaltıldı. Üç boyutlu olan veri PCA yönteminin kullanılabilmesi adına 2'ye düşürüldü. Özellik sayısı da 2'ye düşürüldü. Bu veri tekrardan 3 boyutlu 2 özellikli bir veriye dönüştürüldüğünde Şekil 2'deki sonuç çıkmaktadır.

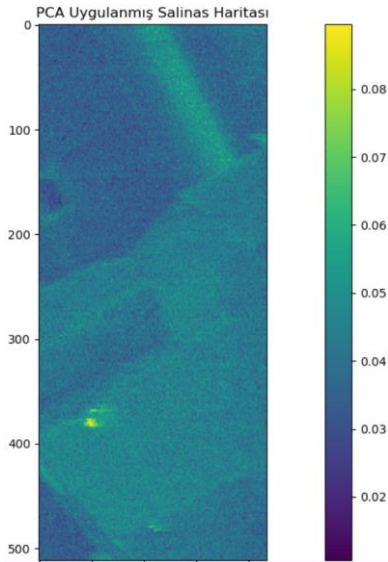


Şekil 2 Salinas verisi, PCA uygulandıktan sonra yeniden şekillendirilmiş

## KAYNAKÇA

- [1] [Temel bileşen analizi - Vikipedi \(wikipedia.org\)](https://tr.wikipedia.org/wiki/Temel_bileşen_analizi)
- [2] [https://colab.research.google.com/drive/1E2RViy7xmor0mhqskZV14\\_NUj2jMpJz3](https://colab.research.google.com/drive/1E2RViy7xmor0mhqskZV14_NUj2jMpJz3)
- [3] [https://www.youtube.com/watch?v=Rjr62b\\_h7S4](https://www.youtube.com/watch?v=Rjr62b_h7S4)

Aynı şekilde baştaki veri setine öbekleme algoritması uygulanıp yeniden şekillendirildiğinde Şekil 3'teki sonuç çıkmaktadır.



Şekil 3 Öbekleme uygulandıktan sonra yeniden şekillendirilmiş salinas verisi

## ANALİZ VE YORUM

PCA yöntemi için boyutun 2'ye düşürüldü çünkü algoritmanın en düzgün çalıştığı yer burasıydı. Özellik sayısının artırılıp azaltılmasında bir değişiklik görülemedi. Çıkan sonuçlara bakıldığında bu veri için PCA yönteminin, öbeklemeye göre daha iyi çalıştığı görülmektedir.