Principal Component Analysis (PCA)

PCA: bir veri kümesindeki değişkenlik miktarını en çok olan değişkenlerin bulunmasını sağlayarak boyut indirme yapar. Temel amacı, veri setindeki değişkenler arasındaki korelasyonu azaltmaktır(bir veri setindeki değişkenlerin birbirleriyle olan ilişkilerini veya bağımlılıklarını azaltmayı ifade eder).

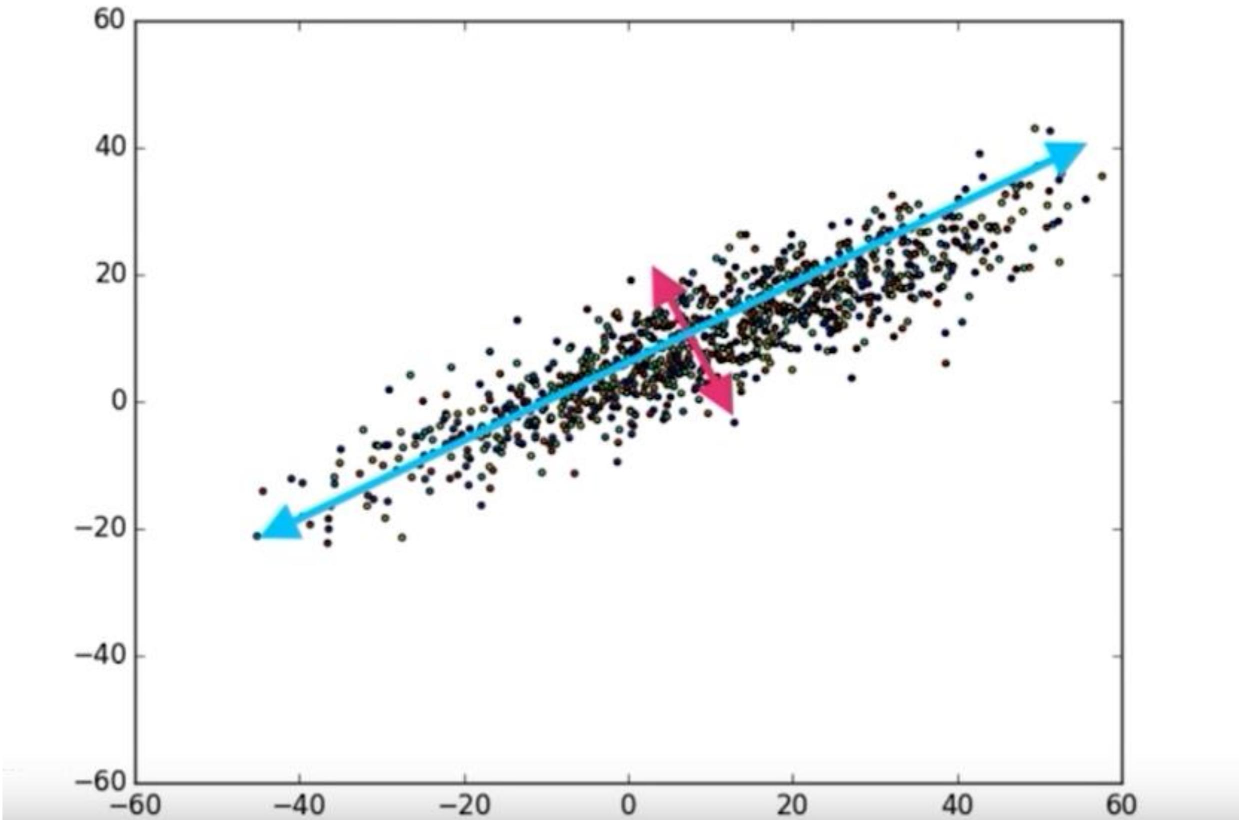
Kullanım Alanları

- Gürültü Filitreleme
- Görselleştirme
- Öznitelik Çıkarımı
- Öznitelik Eleme / Dönüştürme
- Borsa analizi
- Sağlık Verileri / Genetik veriler

NE İÇİN KULLANILIR

- Boyut dönüştürme
- Boyut indirgeme (gereksiz boyutlardan kurtulma veya bazı boyutları birleştirme)
- Değişkenler arasındaki bağlantıları açığa çıkarma

PCA Nedir?



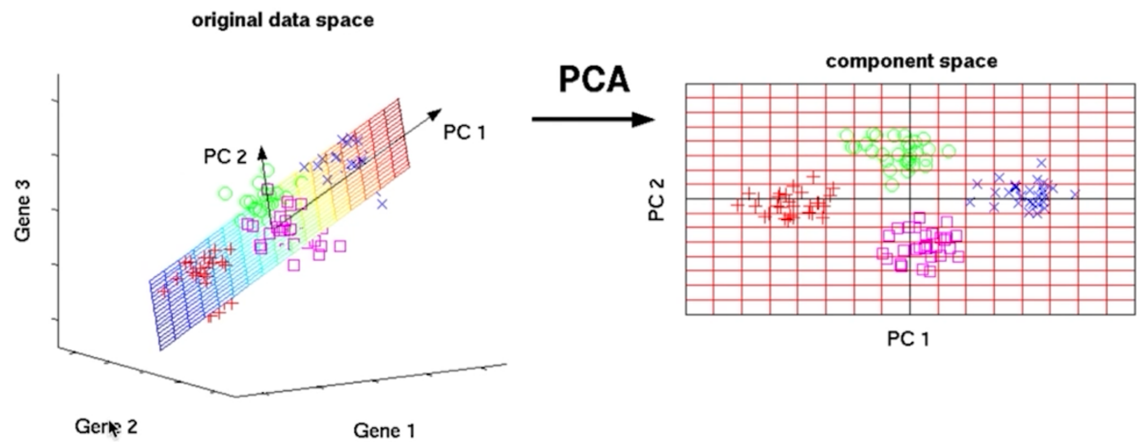
Öyle bir boyut elde edebilir miyim ki verileri birbirinden maximum birbirinden ayırt edilsin.

Gözetimsizdir(unsupervised) sınıflandırılmamıs veriler için

Uzaklık(distance) metricleri üzerinden çalışan algoritmalar için çok şey ifade eder(KNN)

Yeni boyut elde edildiginde veri kaybolabilir.

Boyut İndirgeme



Öyle bir noktalar vardır ki 3 boyuttan 2 boyuta gecildiginde üst üste binebilir ve veri kaybolabilir.

Eigen Value (Oz Değer) ve Eigen Vector (Öz Yöney)

- Rasgele bir matris alalım

scaler bir sayıyla çarpımı
şeklinde yazılabiliyor ise

- Bu matrisi tek boyutlu bir matris ile çarparsak

Öz değer — Vektör

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ -1 & 0 & -2 \end{bmatrix} \mathbf{x} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 0 \end{bmatrix} = \mathbf{3} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Öz değer — Vektör

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 1 & 0 & 2 \end{bmatrix} \mathbf{x} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \mathbf{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Eigen Value (Öz Değer) ve Eigen Vector (Öz Yöney)

- Rasgele bir matris alalım
- Bu matrisi tek boyutlu bir matris ile çarparsak
- Çarpım şayet çarpanın herhangi bir skalar katını veriyorsa, bu skalar öz değer, bu vektör öz yöneydir

$$Av = \underbrace{\lambda}_{\text{Öz Değer}} \underbrace{v}_{\text{Öz Yöney}}$$

PCA Algoritması

- Indirgenmek istenen boyut k olsun
- Veriyi Standartlaştır
- Covariance (Kovaryans) veya Corellation (Korelasyon) matrisinden öz değerleri ve öz vektörleri elde et. Veya SVD kullan.
- Öz değerleri büyükten küçüğe sırala ve k tanesini al.
- Seçilen k özdeğerden W projeksiyon matrisini oluştur
- Orjinal veri kümesi X'i W kullanarak dönüştür ve k-boyutlu Y uzayını elde et.

Linear Discriminant Analysis(LDA)

LDA: sınıflar arasındaki ayrımı maksimize etmek için veri setindeki özellikleri dönüştürür.

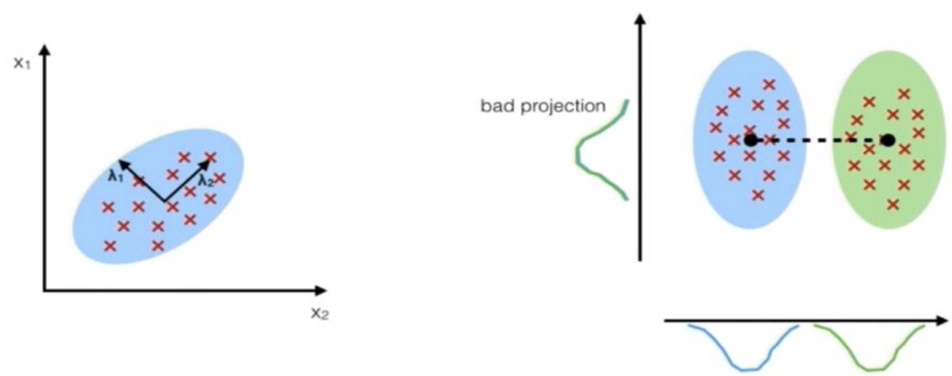
Gözetimlidir(supervised)

Sınıflandırma problemlerinde boyut indirgeme ve sınıflandırma adımlarını bir araya getirir.

LDA
Linear Discriminant Analysis

- PCA benzeri bir boyut dönüştürme / indirgeme algoritmasıdır
- PCA'den farklı olarak sınıflar arasındaki ayrımı önemser ve maksimize etmeye çalışır.
- PCA bu açıdan gözetimsiz (unsupervised) LDA ise gözetimli (supervised) özelliktedir.

PCA ve LCA



- Kaynak: https://sebastianraschka.com/Articles/2014_python_lda.html