Boyut indirgeme için ilk olarak Principal Component Analysis'den bahsedeceğiz.

Boyut indirgeme nerelerde kullanılır?

-->Görüntü filtreleme. Noise'ı atmak için.

-->Görselleştirme

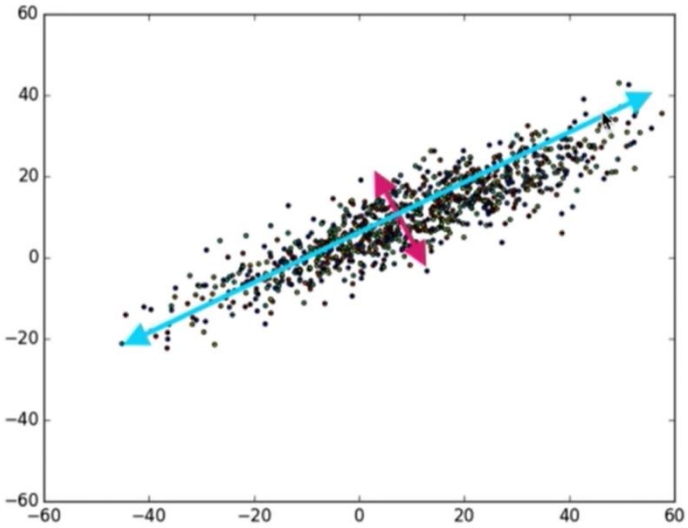
-->Öznitelik çıkarımı. Veriler üzerinden yeni öznitelikler çıkartılabilir.

-->Öznitelik Eleme / Dönüştürme . Her zaman boyut indirgemek zorunda değiliz. Boyut korunarak yeni features da elde edilebilir.

-->Borsa analizi

-->Sağlık verileri / Genetik veriler

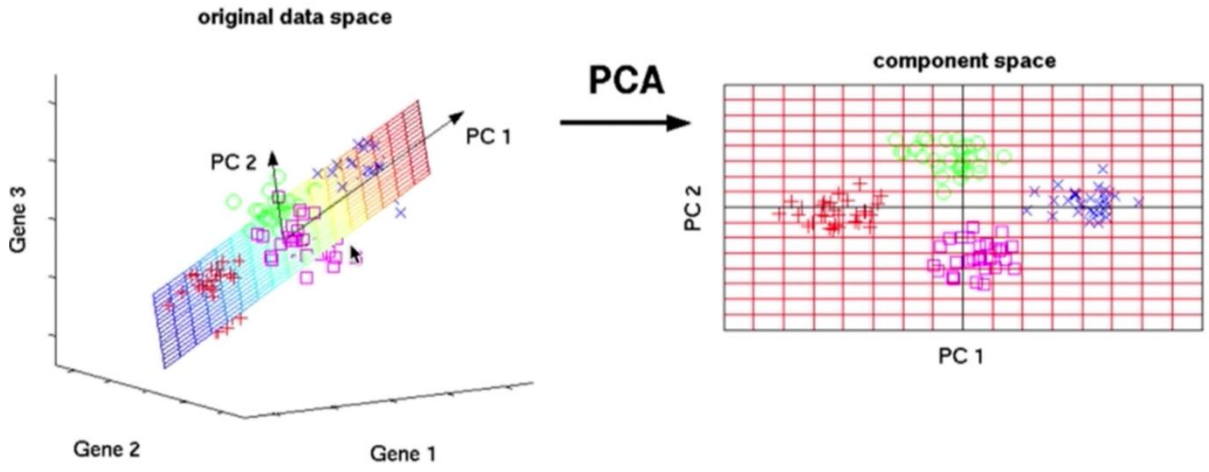
-->Principal component verilerin maximum ayrıklığını elde ektiğimiz axise deniyor.Yani dağılımın en yüksek olduğu axis.



→Mesela yukarıdaki veri setinde mavi eksen variance'ın en yüksek olduğu eksen. Dolayısıyla bu principal component olabilir. Bunun yanında kırmızı eksen de diğer eksen olabilir. Ve aynı veri bu yeni eksenler ile de gösterilebilir.

→Peki bunun avantajı ne? Öncelikle artık principal eksen ile verilerin dağılımı daha net bir şekilde ayırt edilebiliyor. Bu çoğu algoritma için çok şey ifade ediyor mesela KNN çalıştıracağız eski eksenlerde ölçülen mesafe ile bu eksenlerde ölçülen mesafe farklı olacaktır. Yani distance metrics üzerinden çalışan algoritmalarda bu durum fark eder.

→Bunun yanında PCA ile boyut indirgenince bir bilgi kaybından söz edilebilir. Mesela 3D den 2D ye düşüyoruz, yeni surface üzerine aynı noktaya düşer yani üstüste gelen verilen artık ayırt edilemez.



Şimdi bu noktada daha detaya inmeden önce Eigenvalue ve Eigenvector kavramlarından bahsedelim.

- Rasgele bir matris alalım
- Bu matrisi tek boyutlu bir matris ile çarparsak
- Çarpım şayet çarpanın herhangi bir skalar katını veriyorsa, bu skalar eigen value bu vektör eigen vector dür.

$$Av = \lambda v$$

Eigen Value Eigen Vector

1	2	0		1		3			1
0	1	2	x	1	=	3	=	3	1
1	0	2		0		0			0

1	2	0		1		3			1
0	1	2	x	1	=	3	=	3	1
1	0	2		1		3			1

→Zaten detaylı çalışmasını Andrew anlatmıştı. Koda geçebiliriz.

→Bunun yanında boyut indirgeme için bir başka algoritmadan daha bahsedeceğiz.

LİNEAR DISCRIMINANT ANALYSIS

→Bu da PCA gibi bir dimensionality reduction algorithmdir.

→PCA ile bazı farklarından bahsedelim:

→PCA'den farklı olarak sınıflar arasındaki ayrımı önemser ve maksimize etmeye çalışır.

→Yani PCA için unsupervised bir algoritma diyebiliriz. Ancak LDA ise supervised dır. Etiketlere göre bir dimensionality reduction yapar ki etiket grupları birbirinden iyice ayrılsın.

➔ Yani LDA için labeled training set gerekli.