Principal Component Analysis

Main feature selection

which components (features) are important to keep ?

significance = variance

intelligence = recognize significance

starting point is a file with a table.

features

observations	x_1	x_2	x_3	...	x_n
1					
2					
3					
:					
m					

no change

$$\sigma^2 = \frac{1}{n-1} X \cdot X^T$$

$n = |x|$ (cardinality of x)

$$x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad x^T = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

$$\sigma^2 = \frac{1}{2}[1+4+9]$$

$$\Sigma = E\left[(x - E[x])(x - E[x])^T\right]$$

$$C = E\left[x x^T\right]$$

$$\sigma^2 = \text{variance}(x) = E\left[(x - E(x))^2\right]$$

variance $\longrightarrow$ 1D does x_3 change ?

covariance $\longrightarrow$ 2D does x_2, x_3 change ?

covariance $\quad \Sigma = XX^T$

$$
\begin{array}{c|ccc}
 & x_1 & x_2 & x_3 \\
\hline
x_1 & x_1^2 & x_1x_2 & x_1x_3 \\
x_2 & x_2x_1 & x_2^2 & x_2x_3 \\
x_3 & x_3x_1 & x_3x_2 & x_3^2
\end{array}
$$

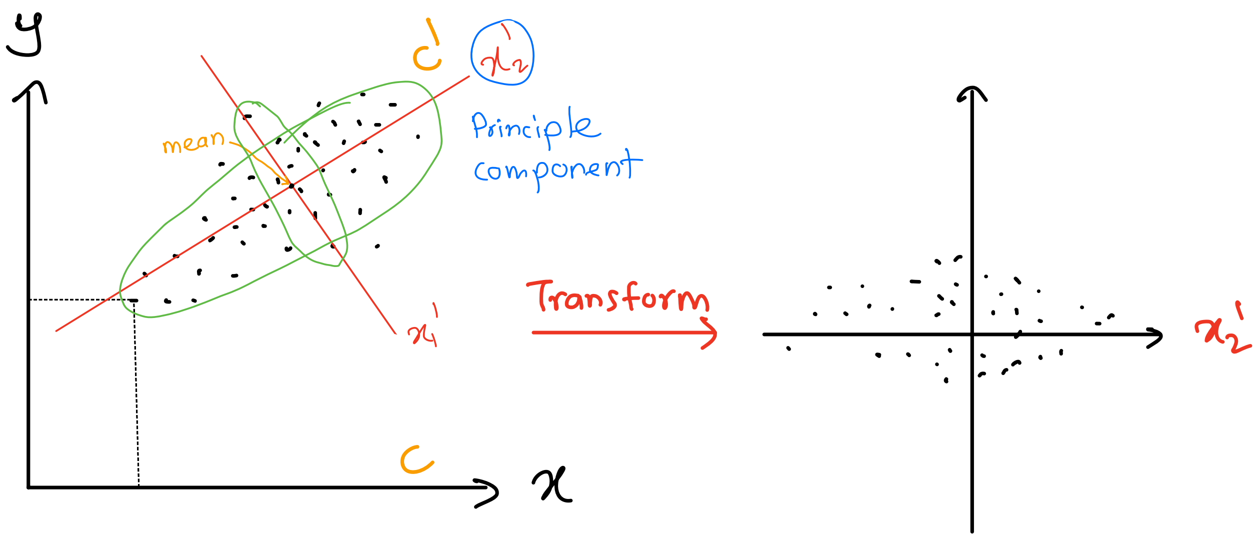
variance

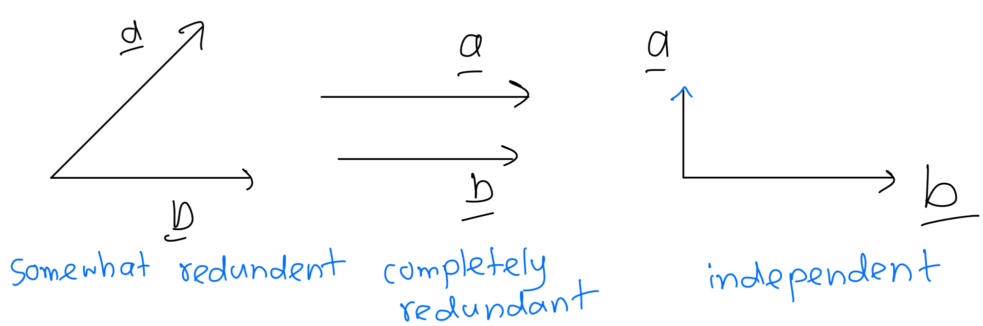
variance $\quad$ covariance

things against same thing $\longrightarrow$ variance

change of 2 different things $\longrightarrow$ covariance

Diagonalizing C using a suitable orthogonal transformation matrix A by obtaining N orthogonal "special vectors" u_i with "special parameters" λ_i





$$\underbrace{\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{u_i} = \begin{bmatrix} 3 \\ 3 \end{bmatrix} = \underset{\lambda}{3} \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{u_i}$$

C- covariance matrix

eigen vector

eigen value

	Apple		Banana		
	1	1	1	1	1
	3	9	0	0	0
	-1	1	-1	1	1
mean	1	11/3	0	2/3	
	A		B		

$Cov(A,B) = E(AB) - E(A)E(B)$
$= E(AB)$
$= 2/3$

$$\begin{array}{c c} & \\ A & \\ B & \end{array} \underbrace{\begin{bmatrix} Cov(A,A) & Cov(A,B) \\ Cov(B,A) & Cov(B,B) \end{bmatrix}}_{\text{Covariance matrix}}$$

$Cov(A,A) = E(A^2) - E(A)^2$
$= 11/3 - 1^2$
$= 8/3$

$Cov(B,B) = 2/3$

$Cov(A,A) = var(A)$
$Cov(A,B) = Cov(B,A)$

$$C_m = \begin{bmatrix} 8/3 & 2/3 \\ 2/3 & 2/3 \end{bmatrix}$$

Linear Transformation

$C \longrightarrow$ symmtric

$C = C^T$

$$u_i = A(x_i - m)$$

$$x_i = m + A^T u_i$$

since $A^{-1} = A^T$ for orthogonal matrices

so $c' = ACA^T$ s.t

$$c' = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ \vdots & \lambda_2 & & & \\ \vdots & & \ddots & & \\ 0 & & & & \lambda_n \end{bmatrix}$$

most important

least important

In an orthogonal transformation, the trace of a matrix remain the same. (inherent characteristic of Data)

$$trace(c) = trace(c') = \sum_{i=1}^{n} \lambda_i$$

Principal Components $\lambda_1, \lambda_2 \cdots \lambda_{N-1}, \lambda_N$

pick $N' \ll N$

$\hookrightarrow$ dimensionality reduction correlation

-1 ——— 1

$var(A) = cov(A, A)$

PCA → a linear transformation
→ unsupervised
→ uses statistics, calculus
→ dimensionality reduction algo
→ a visuallization algorithm
→ intelligent (Recognize significance)

AI and Data

Data types – numbers, symbols, text, images, videos, audio files

Preprocessing – filtering, normallization, outlier detection, dim reduction, augmentation

Data representation –
- Hand crafted features
  stats, SIFT (relate to images)
- Automatic feature extraction
  deep features

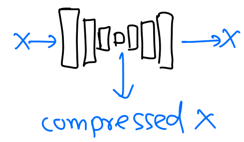
Encoding – compression/embedding
- No learning
- with learning

PCA, Fisher Vector

LDA, VLAD

Auto encoders (painful :c)

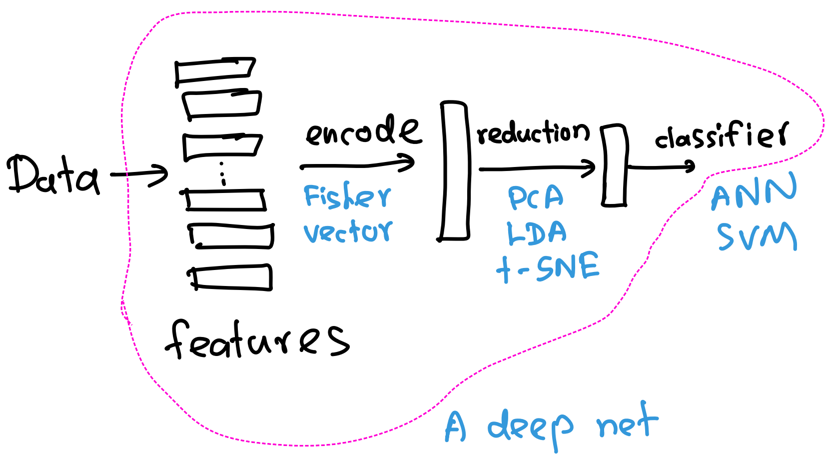
t - SNE



Applications of PCA

* Data reduction
* Data visuallization
* Data classification
* Factor analysis
* Trend analysis
* Noise removal

A meaningful chain :



Project

1. Find a problem
2. Analyse the problem (input/output/knowledge)

3. Select approach (architecture, parameters..)
4. Design the approach
5. Training
6. Retrain if needs
7. Recall phase (unseen data)
8. Compare against other methods