2|2|24

Principal Component Analysis

→ Technique for dimensionality reduction.

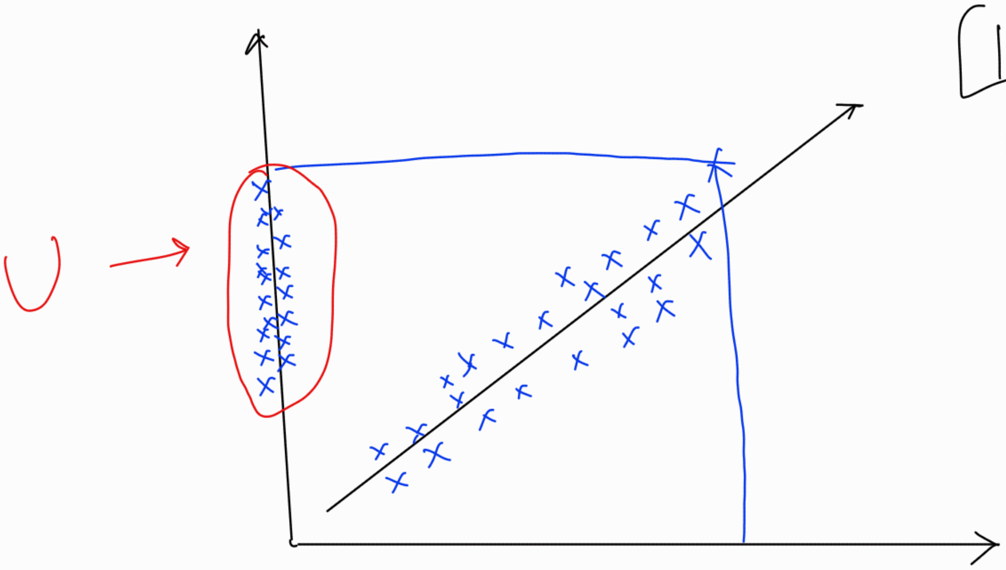
→ Limitation : Linear algorithm.

Linear Classifiers

$$y = \text{sign}(W^T x)$$

Learn values of w's such that the sign will tell us where the point lies wrt to the classifying plane.

State of the art is discriminative classifiers or generative classifiers.



$$\underset{U^T}{[1\times 2]}\ \underset{X}{[2\times N]}$$

U is direction vector

X is original data

$$\boxed{\text{PCA maximizes } var(U^T X)}$$

$$\frac{\sum (x-\mu_x)^2}{N-1} = var(x)$$

For entire population, use N

For sample, use N-1

$$\begin{array}{ccc} & x & y \\ x & \multicolumn{2}{c}{\multirow{2}{*}{$\begin{bmatrix} V_x & cov(x,y) \\ cov(x,y) & V_y \end{bmatrix}$}} \\ y & & \end{array}$$

$$\begin{bmatrix} V_x & cov(x,y) \\ cov(x,y) & V_y \end{bmatrix} \quad (d\times d) \text{ dim.}$$

$$cov(x,y) = \frac{\sum (x-\mu_x)(y-\mu_y)}{N-1}$$

Expectation is weighted mean

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \\ y_1 & y_2 & \cdots & y_N \\ z_1 & z_2 & \cdots & z_N \end{bmatrix}$$

$$X - E[X] = \begin{bmatrix} x_1 - \mu_x & x_2 - \mu_x & \cdots \\ y_1 - \mu_y & y_2 - \mu_y & \cdots \\ z_1 - \mu_z & z_2 - \mu_z & \cdots \end{bmatrix}$$

$$(X - E[X])(X - E[X])^T = S$$

$$S = \begin{bmatrix} var(x,x) & cov(x,y) & cov(x,z) \\ cov(x,y) & var(y,y) & cov(y,z) \\ cov(x,z) & cov(y,z) & var(z,z) \end{bmatrix}$$

Principal component is the first direction where I project to.

We maximize $var(U^T x)$ s.t. $U^T U = 1$

If you don't restrict $U^T U = 1$, i.e. we project to unit vector, the vector in one direction will go on increasing to maximize variance and never stop.

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

$$E[U^T x] = U^T E[x]$$

$$= E\left[U^T \underbrace{(x - E[x]) (x - E[x])^T}_{Var(x)} U \right]$$

$$\boxed{= \max\left[U^T S U \right]} \quad st \quad U^T U = 1$$

By Lagrange's Multiplier,

$$\max_{U} \left[U^T S U - \lambda (U^T U - 1) \right]$$

$$2SU - \lambda 2U = 0$$

$$\therefore \boxed{SU = \lambda U}$$

U is an eigen vector of covariance matrix S.

$$\max U^T S U = \max U^T \lambda U = \max \lambda U^T U$$

$$= \boxed{\max \lambda}$$

eigen vector corresponding to largest eigen value

1] Compute covariance matrix (S)

2] Do eigen decomposition

3] For d dimensions, you get d eigen values.

4] Pick largest eigen value for first principal component.

How do you decide how many dimensions of features to pick?

$$\frac{\sum_{i=1}^{d} \lambda_i}{\sum_{j=1}^{100} \lambda_j} = 0.98$$

This denotes that keeping the first d dimensions will preserve 98% of the accuracy.

→ If the value is just = 0.15, it denotes that there is no point doing a linear dimensionality reduction.