PCA SVD

Unsupervised Learning : Data without Labels

① Find structures in data

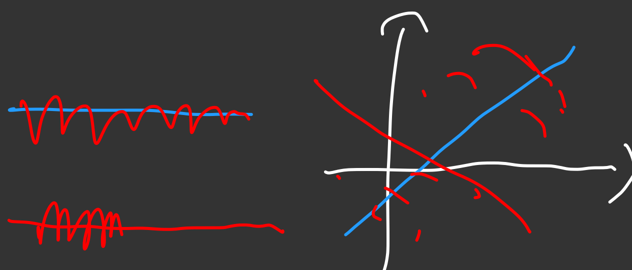
② Data compression and dimension reduction.

③ Explore or explore data (generate data)

④ Construct features for supervised learning

Eg. Kmeans ; PCA ; Gan ; VAE

PCA : Find subspace with highest variance.



Eg. $X = \begin{bmatrix} | & & | \\ x^{(1)} & \cdots & x^{(|D|)} \\ | & & | \end{bmatrix}$

① Subspace should contain 0

centered data

$$\bar{X} = \begin{bmatrix} | & & | \\ x^{(1)} - \mu & \cdots\cdots & x^{(|D|)} - \mu \\ | & & | \end{bmatrix}$$

$$\mu = \frac{1}{|D|} \sum_{x \in D} x$$

② find max var.

$$\max_{w : \|w\|_2^2 = 1} \mathrm{Var}\left[w^T \bar{X} \right]$$

s.t. $\mathrm{Var}(X) = E\left[(X - \mu)^2 \right]$

$$\Rightarrow \max_{w : \|w\|_2^2 = 1} E\left[w^T \bar{X} \bar{X}^T w \right] = \max_{w : \|w\|_2^2 = 1} w^T \Sigma w$$

$$\max \quad w^T \Sigma w$$

$$w: \|w\|_2^2 = 1$$

$$L(\,) = w^T \Sigma w + \alpha \left(\|w\|_2^2 - 1 \right)$$

$$\frac{\partial L}{\partial w} : 2\Sigma w - 2\alpha w = 0$$

$$\underbrace{\Sigma w = \alpha w}$$

w are eigenvectors.

$$w^T \Sigma w = \alpha\, w^T w = \alpha$$

So find eigenvector w corresponds

to max α

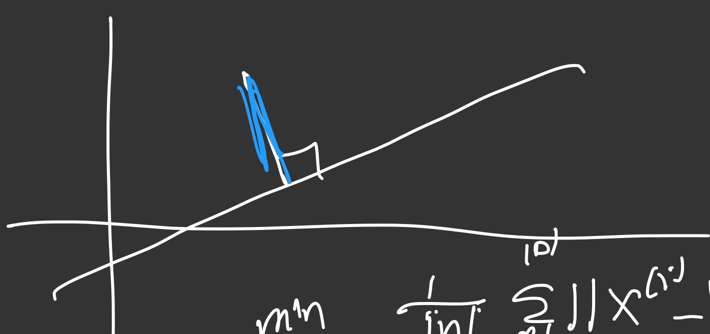
① $U = \begin{bmatrix} | & & | \\ w_1 & \cdots & w_d \\ | & & | \end{bmatrix}$ (subspace dir)

② $\hat{x} = U^T(x - \nu) \in \mathbb{R}$ (proj on subspace)

③ $\tilde{x} = U\hat{x} + \nu$ (reconstruction)

$\boxed{\text{Alternative view of PCA}}$:

Find axis that min sum of sqr distance from points to their orthogonal projections.



$$\min_{w : \|w\|_2^2 = 1} \frac{1}{|D|} \sum_{i=1}^{|D|} \|x^{(i)} - ww^T x^{(i)}\|_2^2$$

$\boxed{\text{Frobenius form}}$

$$\|A\|_F^2 = \sum_{i,j} a_{i,j}^2 = Tr(A^T A)$$

$$\text{so } \frac{1}{|D|} \sum_{i=1}^{|D|} \|x^{(i)} - ww^T x^{(i)}\|_2^2$$

$$= \frac{1}{|D|} \|\underbrace{\bar{x} - ww^T \bar{x}}_{P\bar{x} = (I - ww^T)(\bar{x})}\|_F^2$$

$$= \frac{1}{|D|} Tr((P\bar{x})^T (P\bar{x}))$$

$$= \frac{1}{|D|} Tr(\bar{x}\bar{x}^T P^T P)$$

$$= Tr(\Sigma P)$$

$$= Tr(\Sigma) - Tr(\Sigma ww^T)$$

$$\underset{w}{\operatorname{argmin}}\ Tr(\Sigma) - Tr(\Sigma ww^T)$$

$$= \underset{w}{\operatorname{argmax}}\ Tr(\Sigma ww^T)$$