PCA Algorithm

Step 1: Calculate $\overbrace{\Sigma_x}^{D\times D}$

Step 2: $\underbrace{\text{Find first } D' \text{ eigenvectors}}_{\text{eigenvectors that correspond to } D' \text{ largest eigenvalues}}$ of Σ_x.

$$W = \overbrace{\begin{bmatrix} | & | & & | \\ w_1 & w_2 & \cdots & w_{D'} \\ | & | & & | \end{bmatrix}}^{D\times D'}$$

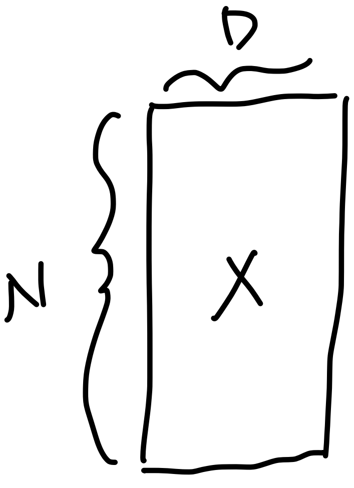
Projection Step: $z_i = W^T . (x_i - \hat{\mu})$

$$\Sigma_x = \frac{\sum_{i=1}^{N} (x_i - \hat{\mu}).(x_i - \hat{\mu})^T}{N}$$

if we center data points

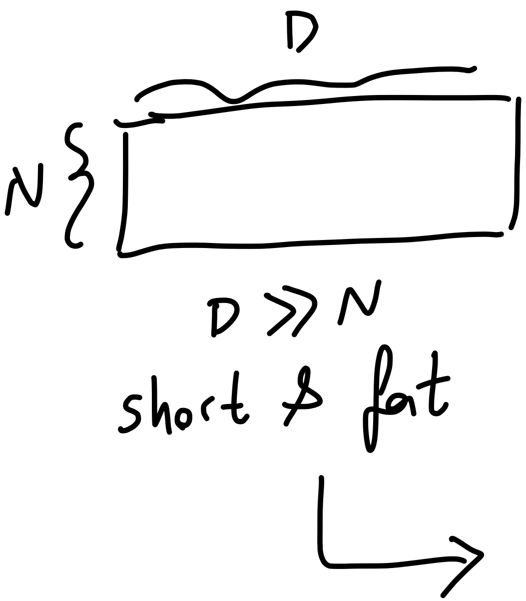
$$\tilde{x}_i = x_i - \hat{\mu}$$

$$\Sigma_x = \frac{\tilde{X}\tilde{X}^T}{N}$$



$N \gg D$

tall & thin



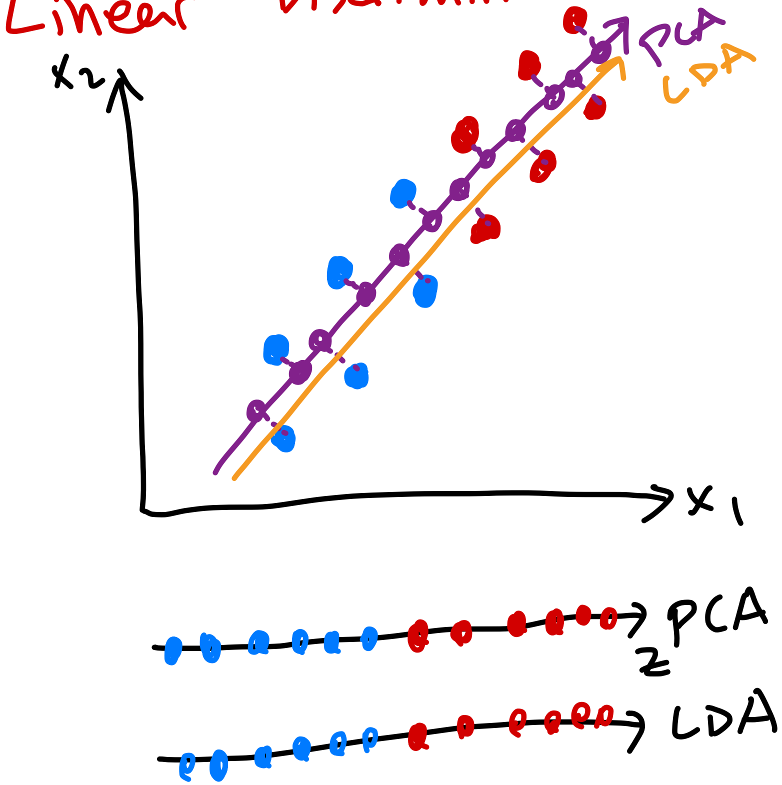
$D \gg N$

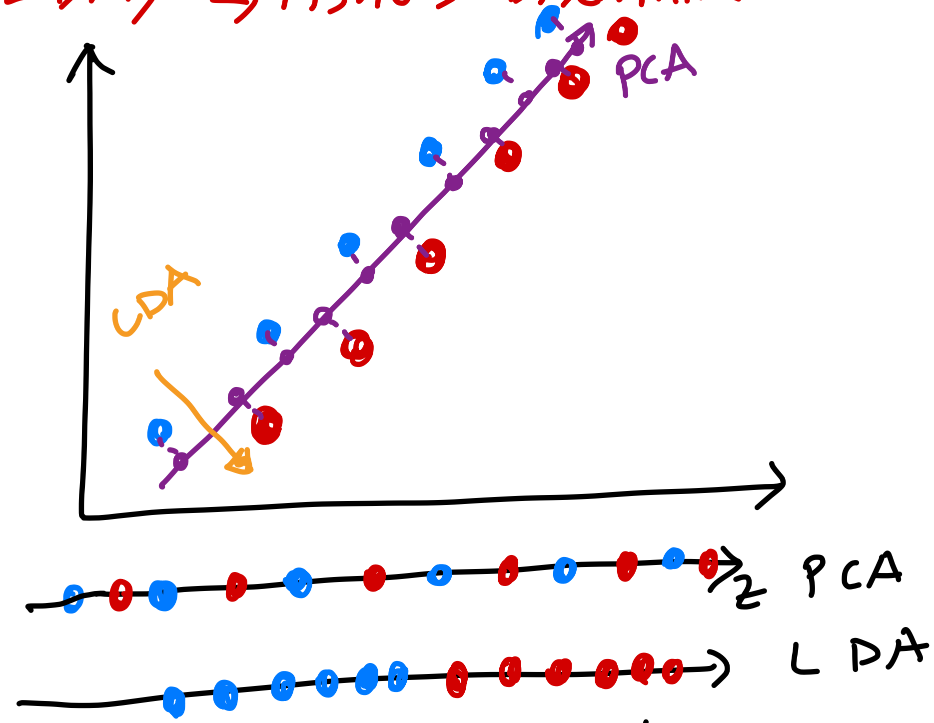
short & fat

$$\overbrace{\Sigma_x}^{D\times D} . w = \alpha . w$$

$$X^T \frac{X X^T}{N} . w = \alpha X^T w \Rightarrow \frac{\overbrace{X^T . X}^{N\times N}}{N} . u = \alpha . u$$

Linear Discriminant Analysis (LDA) ⇒ Fisher's Discriminant Analysis (FDA)





$$\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N}$$

$$\begin{cases} y_i = 0 & \text{if negatively labeled} \\ y_i = 1 & \text{if positively labeled} \end{cases}$$

$$\underbrace{z_i}_{1\times 1} = \underbrace{w^T}_{1\times D} \cdot \underbrace{x_i}_{D\times 1}$$

w^*

sample means in the projected space ⇐ $|\hat{\mu}_1 - \hat{\mu}_2|$ ⇒ as large as possible

$|s_1^2 + s_2^2|$ ⇒ as small as possible

Sample variances in the projected space.

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{N} z_i . y_i}{\sum_{i=1}^{N} y_i} = \frac{\sum_{i=1}^{N} w^T x_i . y_i}{\sum_{i=1}^{N} y_i} = w^T . \underbrace{\left[\frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} y_i} \right]}_{} = \underline{w^T . \mu_1}$$

$$\hat{\mu}_2 = \frac{\sum_{i=1}^{N} z_i (1-y_i)}{\sum_{i=1}^{N} (1-y_i)} = \frac{\sum_{i=1}^{N} w^T . x_i (1-y_i)}{\sum_{i=1}^{N} (1-y_i)} = w^T . \underbrace{\left[\frac{\sum_{i=1}^{N} x_i (1-y_i)}{\sum_{i=1}^{N} (1-y_i)} \right]}_{\mu_2} = w^T . \mu_2$$

μ_1

$$(AB)^T = B^T . A^T$$

$$s_1^2 = \sum_{i=1}^{N} \left(z_i - \hat{\mu}_1 \right)^2 . y_i = \sum_{i=1}^{N} \boxed{\left(w^T . x_i - w^T . \mu_1 \right)^2} . y_i$$

$$= \sum_{i=1}^{N} \underline{w^T (x_i - \mu_1)(x_i - \mu_1)^T . w} . y_i$$

$$s_2^2 = w^T . S_2 . w$$

$$= w^T . \underbrace{\left[\sum_{i=1}^{N} (x_i - \mu_1)(x_i - \mu_1)^T . y_i \right]}_{S_1} . w$$

S_1 sample covariance of the first class in the original domain

$$\overbrace{\left[w^T . (x_i - \mu_1) \right]^T}$$

$$(w^T . x_i - w^T . \mu_1)(w^T x_i - w^T . \mu_1)^T = \underline{w^T . (x_i - \mu_1) . (x_i - \mu_1)^T . w}$$

$$J(w) = \frac{(\hat{\mu}_1 - \hat{\mu}_2)^2}{S_1^2 + S_2^2}$$

$$(w^T.\mu_1 - w^T.\mu_2)(w^T.\mu_1 - w^T.\mu_2)^T$$

$$w^T.\underbrace{(\mu_1-\mu_2)(\mu_1-\mu_2)^T}_{S_B = \text{between class scatter matrix}}.w$$

$$w^T.S_1.w + w^T.S_2.w$$

$$w^T.\underbrace{(S_1+S_2)}_{S_W = \text{within-class scatter matrix}}.w$$

maximize $\dfrac{w^T.S_B.w}{w^T.S_W.W}$ $\quad w^* = ?$

$$\begin{bmatrix}3\\2\\1\end{bmatrix}\begin{bmatrix}3 & 2 & 1\end{bmatrix} = \begin{bmatrix}9 & 6 & 3\\6 & 4 & 2\\3 & 2 & 1\end{bmatrix}$$

$$S_1 = \sum_{i=1}^{N}(x_i-\mu_1)(x_i-\mu_1)^T.y_i$$

$$w^* = S_W^{-1}.(\mu_1-\mu_2)$$

$$\begin{bmatrix}25 & 11\\11 & 5\end{bmatrix} \Leftarrow \begin{bmatrix}3\\1\end{bmatrix}\begin{bmatrix}3 & 1\end{bmatrix} = \begin{bmatrix}9 & 3\\3 & 1\end{bmatrix} \quad \begin{bmatrix}4\\2\end{bmatrix}\begin{bmatrix}4 & 2\end{bmatrix} = \begin{bmatrix}16 & 8\\8 & 4\end{bmatrix}$$

$$\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N} \qquad x_i \in \mathbb{R}^D \qquad y_{ic} = \begin{cases} 1 & \text{if } y_i = c \\ 0 & \text{otherwise} \end{cases}$$

$$\underbrace{z_i}_{D' \times 1} = \underbrace{W^T}_{D' \times D} \cdot \underbrace{x_i}_{D \times 1}$$

$$S_c = \sum_{i=1}^{N} (x_i - \mu_c)(x_i - \mu_c)^T \cdot y_{ic}$$

$$J(W) = \frac{\det(W^T \cdot S_B \cdot W)}{\det(W^T \cdot S_W \cdot W)}$$

$$S_W = S_1 + S_2 + \cdots + S_K$$

$$S_B = \sum_{i=1}^{N} \sum_{c=1}^{K} (\mu_c - \mu)(\mu_c - \mu)^T \cdot y_{ic}$$

W^* = the largest eigenvectors of $\underbrace{S_W^{-1}}_{\substack{\text{full rank} \\ \Rightarrow D}} \cdot \underbrace{S_B}_{K-1 \Leftarrow \text{rank}}$

$$\text{rank}(A \cdot B) \leq \min(\text{rank}(A), \text{rank}(B))$$

Multidimensional Scaling (MDS)

No access to x_i's.

$$\boxed{x_{Ankara}, x_{London}, x_{Paris}} \, ?$$

Ankara – London $= d_{AL}$

Ankara – Paris $= d_{AP}$

London – Paris $= d_{LP}$

$$\begin{array}{c|ccc|} & \text{Ankara} & \text{London} & \text{Paris} \\ \hline \text{Ankara} & 0 & d_{AL} & d_{AP} \\ \text{London} & d_{AL} & 0 & d_{LP} \\ \text{Paris} & d_{AP} & d_{LP} & 0 \\ \hline \end{array}$$

new city.

$z_4 = ?$

Input $D = \{d_{ij}\}_{i=1, j=1}^{N, N}$

Output $z_1, z_2, \ldots, z_N \in \mathbb{R}^{D'}$

$$d_{ij} = \|x_i - x_j\|_2$$

$$e_{ij} = \|z_i - z_j\|_2$$

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1N} \\ d_{21} & d_{22} & \cdots & d_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & \cdots & d_{NN} \end{bmatrix} \qquad E = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1N} \\ e_{21} & e_{22} & \cdots & e_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ e_{N1} & e_{N2} & \cdots & e_{NN} \end{bmatrix}$$

$$D \cong E$$

Sammon mapping (Sammon Stress)

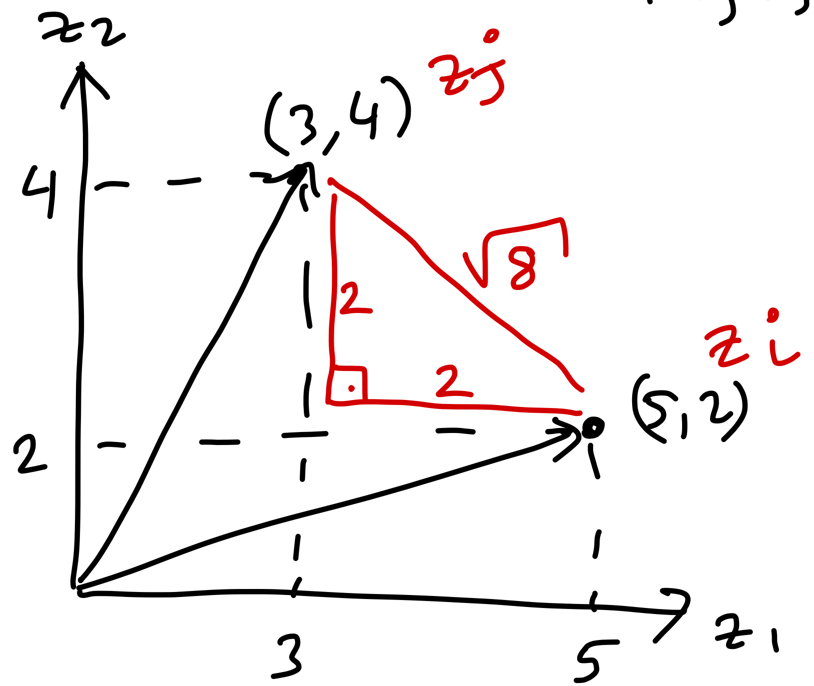
$$\text{Error} = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{(d_{ij} - e_{ij})^2}{d_{ij}^2} = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{(d_{ij} - \|z_i - z_j\|_2)^2}{d_{ij}^2}$$

$$\text{minimize} \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{(d_{ij} - \|z_i - z_j\|_2)^2}{d_{ij}^2}$$

with respect to: $z_i \in \mathbb{R}^{D'}$

$\Downarrow$ out-of-sample embedding is not possible

$$\Rightarrow \|z_i - z_j\|_2 = \sqrt{z_i^T z_i - 2 z_i^T z_j + z_j^T z_j}$$



If we have access to x_i's

$$z_i = W^T . x_i$$

$$\text{minimize} \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{(d_{ij} - \|W^T.x_i - W^T.x_j\|_2)^2}{d_{ij}^2}$$

with respect to: $W \in \mathbb{R}^{D \times D'}$

out-of-sample embedding $\Rightarrow$ $z_{N+1} = W^T . x_{N+1}$

$$\sqrt{\begin{bmatrix}5 & 2\end{bmatrix}\begin{bmatrix}5\\2\end{bmatrix} - 2\begin{bmatrix}5 & 2\end{bmatrix}\begin{bmatrix}3\\4\end{bmatrix} + \begin{bmatrix}3 & 4\end{bmatrix}\begin{bmatrix}3\\4\end{bmatrix}}$$

$$29 - 46 + 25 = 8$$