PCA Algorithm

Training

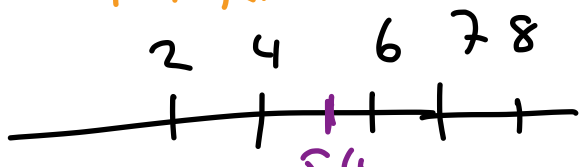
Step 1: Calculate $\overbrace{\Sigma_x}^{D\times D}$

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

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

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

$\hat{\mu}$ ↳ sample mean $\sum_{i=1}^{N} \frac{x_i}{N}$



$-3.4 \quad -1.4 \quad 0.6 \quad 1.6 \quad 2.6$

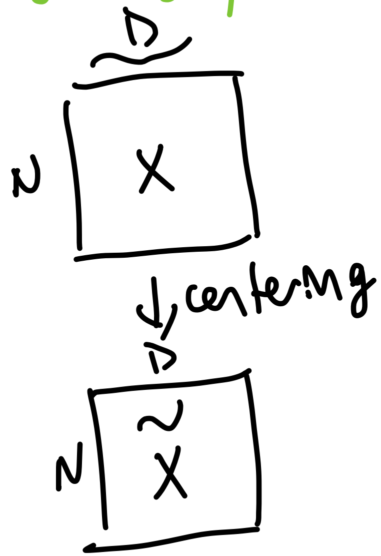
$$\Sigma_x = \frac{\sum_{i=1}^{N} (x_i-\hat{\mu})(x_i-\hat{\mu})^T}{N}$$

center data points

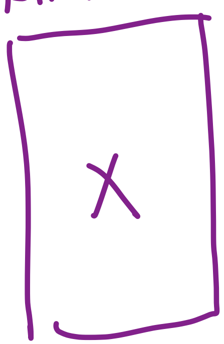
↓

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

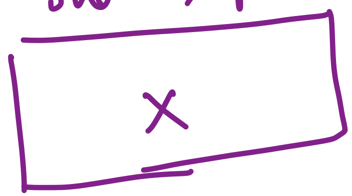
$$\Sigma_x = \frac{\sum_{i=1}^{N} \tilde{x}_i \tilde{x}_i^T}{N} = \frac{\tilde{X}^T \tilde{X}}{N}$$



tall & thin



short & fat



$$X \Sigma_x . W = \alpha X W$$

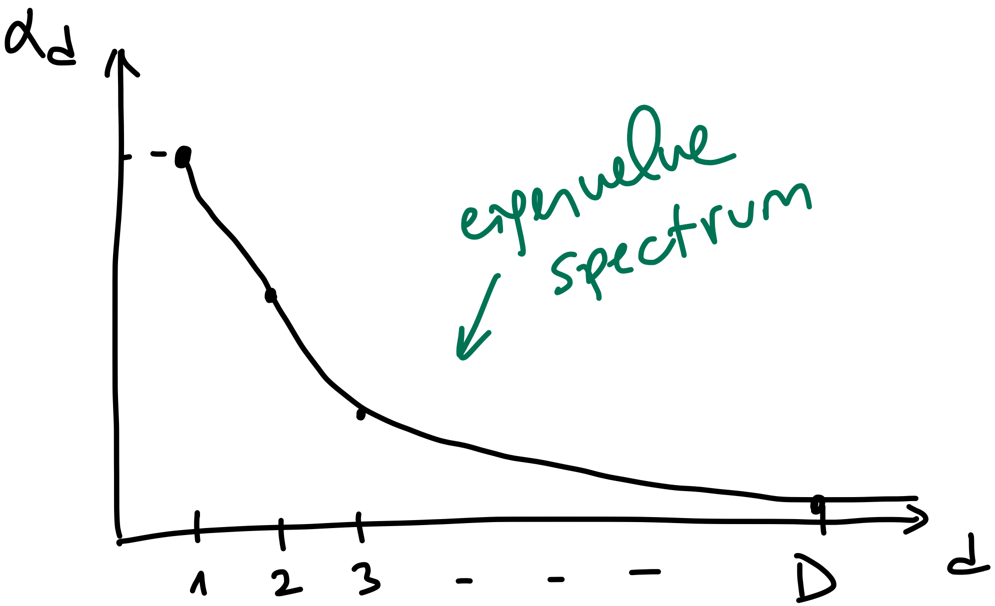
$$\underbrace{\frac{X X^T}{N}}_{\tilde{\Sigma}} \underbrace{X.W}_{\tilde{W}} = \alpha . \underbrace{X.W}_{\tilde{W}}$$

$$\tilde{\Sigma} \tilde{W} = \alpha . \tilde{W}$$

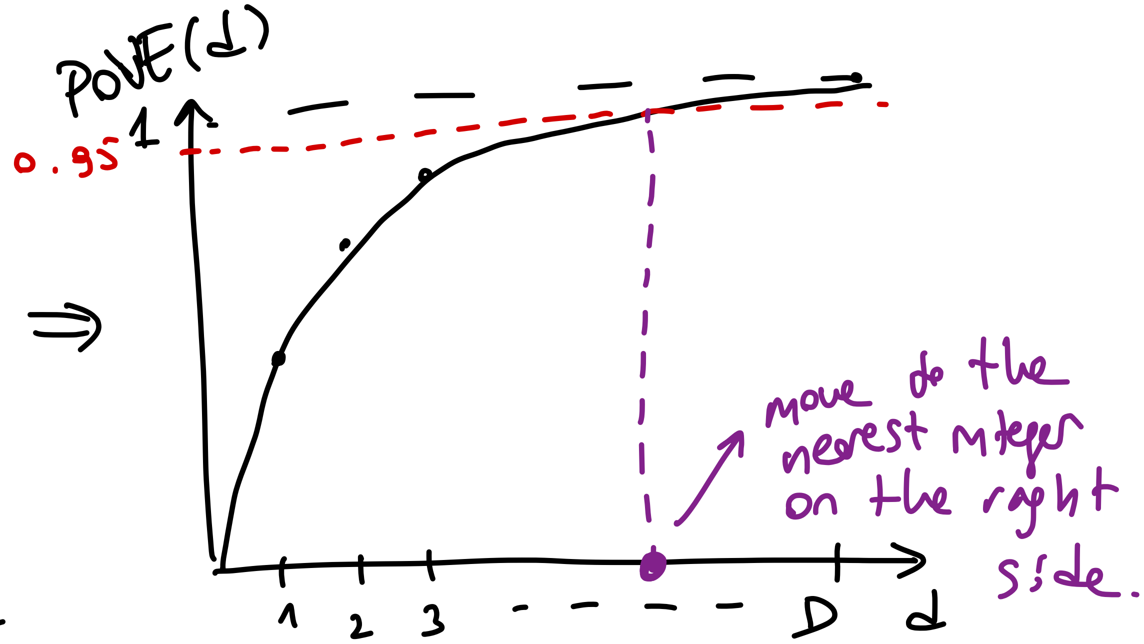
How to pick D'? Proportion of Variance Explained (PoVE)

$$\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_D \geq 0$$

$$\mathrm{PoVE}(D') = \frac{\alpha_1 + \alpha_2 + \dots + \alpha_{D'}}{\alpha_1 + \alpha_2 + \dots + \alpha_D}$$



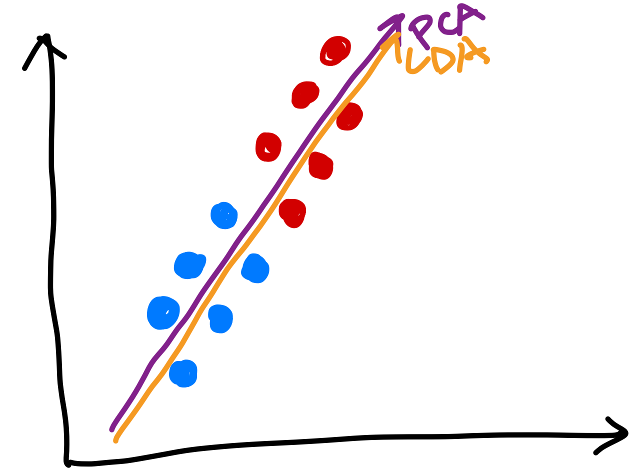
$\Rightarrow$

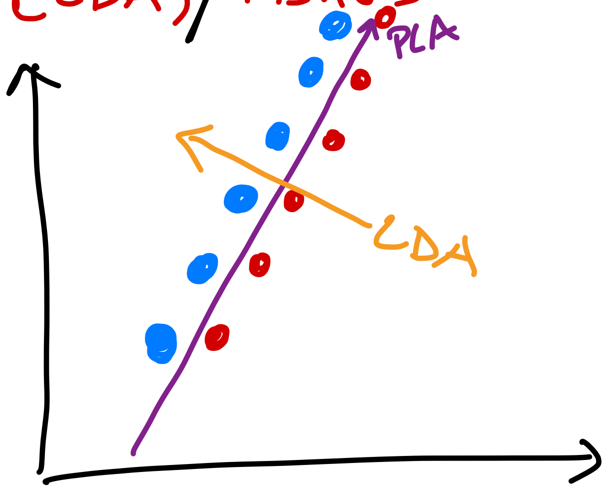


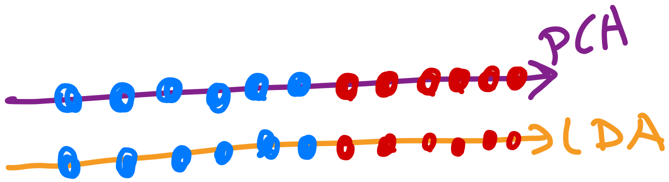
Rule of thumb:

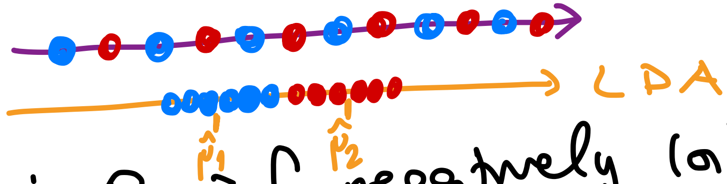
At least 95% of variance should be preserved

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









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

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

projected space

original space

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

$w^* = ?$

Sample means in the projected space $\left\{ \; |\tilde{\mu}_1 - \tilde{\mu}_2| \Rightarrow \text{as large as possible} \right.$

Sample variances in the projected space $\left\{ \; s_1^2 + s_2^2 \Rightarrow \text{as small as possible} \right.$

$$\tilde{\mu}_1 = \frac{\sum_{i=1}^{N} z_i . y_i}{\underbrace{\sum_{i=1}^{N} y_i}_{N_+}} = \frac{\sum_{i=1}^{N} w^T . x_i y_i}{\sum_{i=1}^{N} y_i} = w^T \left(\frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} y_i} \right) = w^T . \mu_1$$

$$\tilde{\mu}_2 = \frac{\sum_{i=1}^{N} z_i (1-y_i)}{\underbrace{\sum_{i=1}^{N} (1-y_i)}_{N_-}} = \frac{\sum_{i=1}^{N} w^T . (x_i)(1-y_i)}{\sum_{i=1}^{N} (1-y_i)} = w^T . \frac{\sum_{i=1}^{N} x_i (1-y_i)}{\sum_{i=1}^{N} (1-y_i)} = w^T . \mu_2$$

$$s_1^2 = \frac{\sum_{i=1}^{N} (z_i - \tilde{\mu}_1)^2 . y_i}{\sum_{i=1}^{N} y_i} = \frac{\sum_{i=1}^{N} (w^T . x_i - w^T . \mu_1)^2 . y_i}{\sum_{i=1}^{N} y_i} = \frac{\sum_{i=1}^{N} \underset{1 \times D}{w^T} . \underset{D \times 1}{(x_i - \mu_1)} \underset{1 \times D}{(x_i - \mu_1)^T} \underset{D \times 1}{w} . y_i}{\sum_{i=1}^{N} y_i}$$

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

$$= w^T . S_1 . w$$

$$s_2^2 = \frac{\sum_{i=1}^{N} (z_i - \tilde{\mu}_2)(1-y_i)}{\sum_{i=1}^{N} (1-y_i)} = w^T . S_2 . w$$

Sample covariance matrices in the original space

$E[aX+b] = aE[X]+b$

3	4	5	6	$\Rightarrow 4.5$
↓	↓	↓	↓	
3x3	3x4	3x5	3x6	
9	12	15	18	$\Rightarrow 13.5$

$$J(w) = \frac{(\tilde{\mu}_1 - \tilde{\mu}_2)^2}{s_1^2 + s_2^2} = \frac{(w^T.\mu_1 - w^T.\mu_2)^2}{w^T.S_1 w + w^T.S_2.w} = \frac{w^T.(\mu_1-\mu_2)(\mu_1-\mu_2)^T w}{w^T(S_1+S_2)w}$$

$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$ ← between class scatter matrix

$S_W = S_1 + S_2$ ← within-class scatter matrix

$$J(w) = \frac{w^T.S_B.w}{w^T.S_W.w} \Rightarrow w^* = ?$$

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

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

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

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

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

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

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

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

W^* = the largest eigenvectors of $(\underbrace{S_W^{-1}}_{\text{full rank } (D)} . \underbrace{S_B}_{\text{rank} = K-1})$

$K-1$ if $K \ll D$

$$\text{RANK}(A . B) \leq \min(\text{Rank}(A), \text{RANK}(B))$$