PRINCIPAL COMPONENT ANALYSIS

1) Calculate Σ_x

2) Find first K eigenvectors of Σ_x

$$m = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$z^* = W^T(x^* - \underline{\underline{m}})$$

→ sample mean

← projection

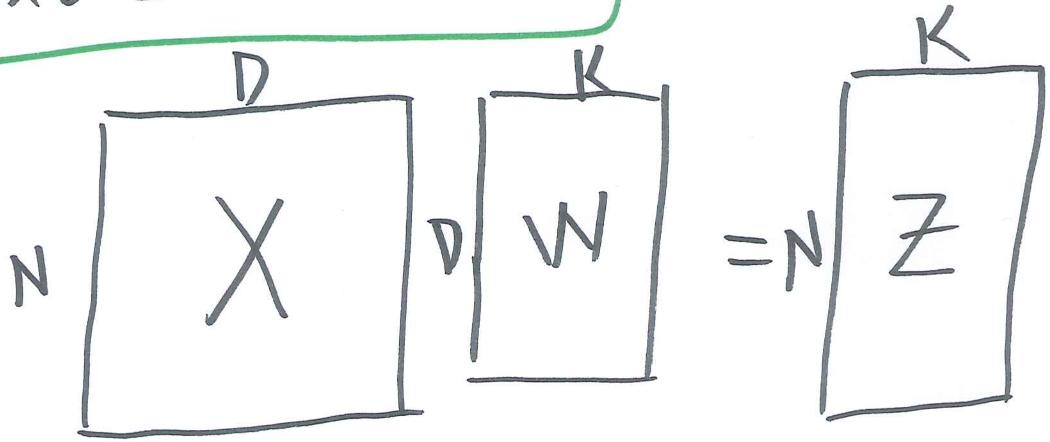
centering

$$W z_i = \underbrace{W W^T}_{I}(x_i - m)$$

$$\Downarrow$$

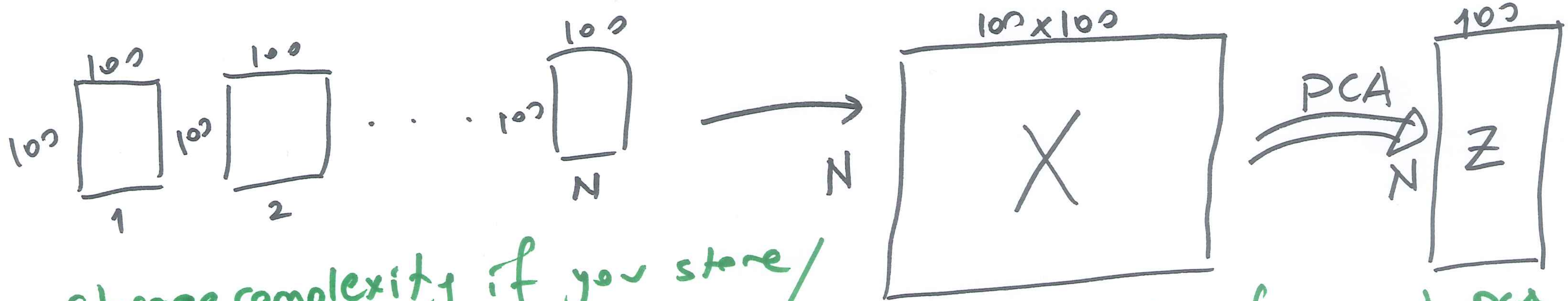
$$\boxed{\hat{x}_i = W.z_i + m}$$

$$WW^T = \begin{bmatrix} \overbrace{w_1^T w_1}^{1} & \overbrace{w_1^T w_2}^{0} & \cdots & \overbrace{w_1^T w_K}^{0} \\ \underbrace{w_2^T . w_1}_{0} & \underbrace{w_2^T w_2}_{1} & \cdots & \underbrace{w_2^T w_K}_{0} \\ & & & \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & & & & & \\ & 1 & & & 0 & \\ & & 1 & & & \\ & & & 1 & & \\ & 0 & & & 1 & \\ & & & & & 1 \end{bmatrix}}_{I}$$



Reconstruction Error

$$\sum_{i=1}^{N} \| x_i - \hat{x}_i \|^2$$



storage complexity if you store the original images.

$N \times (100 \times 100) = \underline{N} \times \underline{10000}$

IF $N = \underline{10000}$ $\quad \underline{\underline{10^8}}$

storage complexity if you do PCA

$$\left.\begin{array}{l} Z \rightarrow N \times 100 \\ W \rightarrow 10000 \times 100 \\ m \rightarrow 10000 \times 1 \end{array}\right\} N \times 100 + \underline{10000} \times \underline{101}$$

IF $N = 10000 \Rightarrow 101\, 10^4 + 10^6 \cong (2.01)\underline{\underline{10^6}}$

FACTOR ANALYSIS

$x \in \mathbb{R}^D$

PCA $\quad x \rightarrow z$

FA $\quad z \rightarrow x$

$\hookrightarrow$ hidden or latent factors

$$x_d - \mu_d = V_{d1} \cdot z_1 + V_{d2} \cdot z_2 + \ldots + V_{dK} \cdot z_K + \epsilon_d \quad \forall d.$$

$$\underset{D\times 1}{x} - \underset{D\times 1}{\mu} = \underset{D\times K}{V} \; \underset{K\times 1}{z} + \epsilon$$

$z \hookrightarrow$ factors

$V \hookrightarrow$ factor loadings

MULTIDIMENSIONAL SCALING

Ankera - London = ...

Ankera - Paris = ...

⋮ ⋮

Input $D = \{ d_{ij} \}_{i=1, j=1}^{N, N}$ all possible pairwise distances.

$\Downarrow$

no exact coordinates for x_i's

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

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

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

$D \approx E$

$D = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$

$E = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$

Sammon mapping (Sammon stress)

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

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

with respect to $z_i \in \mathbb{R}^k$.

out-of-sample embedding. (X)

If we know x_i's

$$z_i = W \cdot x_i$$

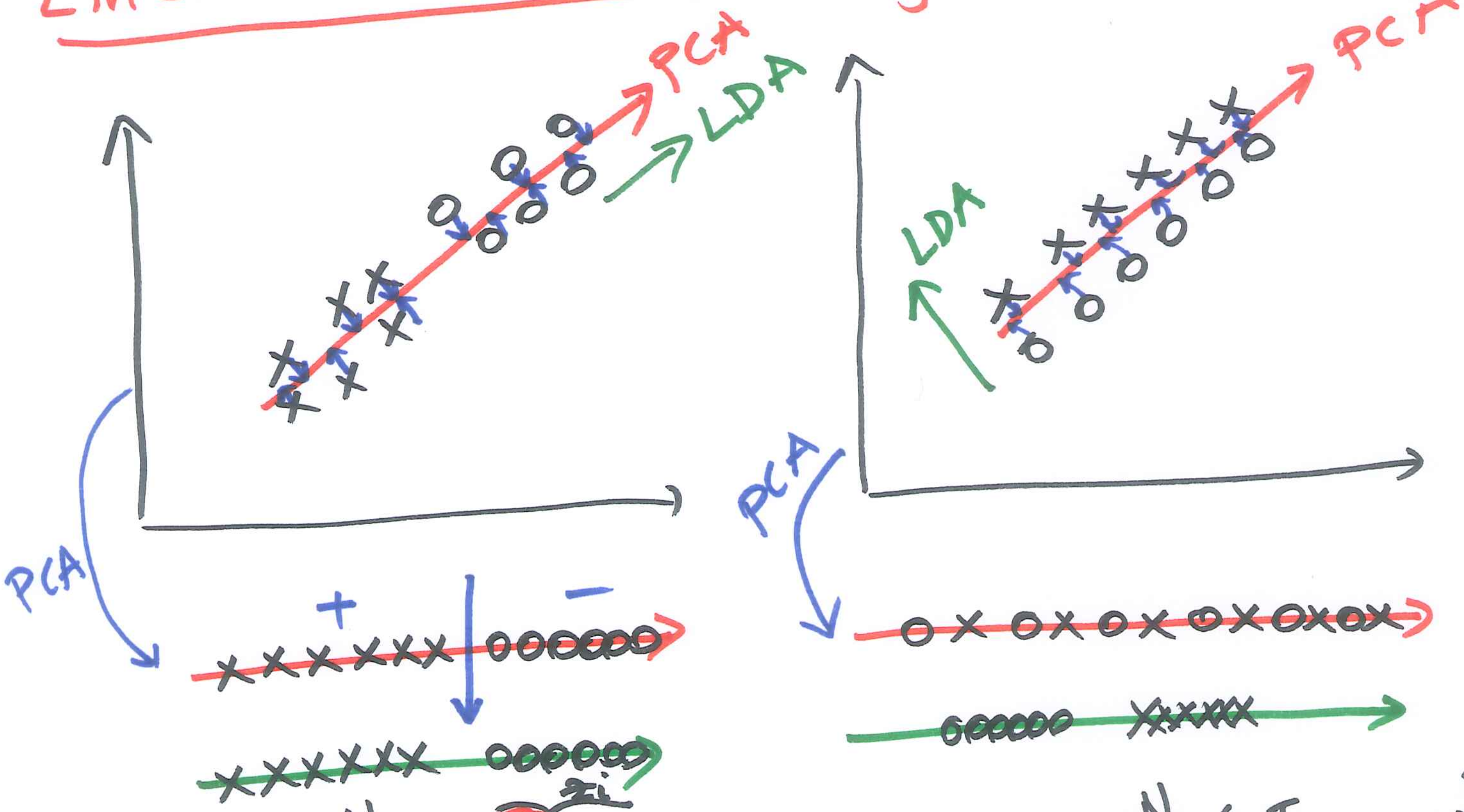
$$\text{minimize} \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{(\|Wx_i - Wx_j\|_2 - d_{ij})^2}{d_{ij}^2}$$

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

out-of-sample embedding (✓)

$$z^* = Wx^*$$

Linear Discriminant Analysis



$z_i = \underline{\underline{w}}^T . x_i$

$|\underline{m_1} - \underline{\underline{m_2}}| \Rightarrow$ as large as possible

$s_1^2 + s_2^2 \Rightarrow$ as small as possible.

$y_i = 1$ if class is C_1

$y_i = 0$ if class is C_2

$$m_1 = \frac{w^T \sum_{i=1}^{N} (w^T . x_i) . y_i}{\sum_{i=1}^{N} y_i} = w^T . \mu_1$$

$$s_1^2 = \sum_{i=1}^{N} (\underline{w}^T . x_i - \underline{\underline{m_1}})^2 . y_i$$

$$m_2 = \frac{\sum_{i=1}^{N} (w^T . x_i)(1 - y_i)}{\sum_{i=1}^{N} (1 - y_i)} = w^T . \mu_2$$

$$s_2^2 = \sum_{i=1}^{N} (w^T . x_i - m_2)^2 (1 - y_i)$$

$$J(w) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2} \Rightarrow (m_1 - m_2)^2 = (w^T \cdot \mu_1 - w^T \cdot \mu_2)^2$$

$$= \underbrace{w^T}_{1 \times D} \underbrace{(\mu_1 - \mu_2)}_{D \times 1} \underbrace{(\mu_1 - \mu_2)^T}_{1 \times D} \cdot \underbrace{w}_{D \times 1}$$

$$= w^T \cdot S_B \cdot w$$

↳ between class scatter matrix.

$$S_B = (\mu_1 - \mu_2) \cdot (\mu_1 - \mu_2)^T$$

$$s_1^2 = \sum_{i=1}^{N} (w^T \cdot x_i - \overset{w^T \cdot \mu_1}{m_1})^2 \cdot y_i$$

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

$$s_2^2 = \sum_{i=1}^{N} (w^T \cdot x_i - \overset{(w^T \cdot \mu_2)}{\underline{m_2}})^2 \cdot (1 - y_i)$$

$$= \sum_{i=1}^{N} w^T (x_i - \mu_2)(x_i - \mu_2)^T \cdot \cancel{y_i} (1 - y_i) \cdot w = w^T \cdot \underbrace{\left[\sum_{i=1}^{N} (x_i - \mu_2)(x_i - \mu_2)^T \cdot (1 - y_i) \right]}_{S_2.} w$$

$$J(w) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2} = \frac{w^T . S_B . w}{w^T . S_1 w + w^T . S_2 . w} = \frac{w^T . S_B . w}{w^T . S_W . w}$$

$$S_W = S_1 + S_2$$

within-class scatter matrix

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

$\boxed{\text{Reading exercise}}$ Read the derivation from the textbook

$$\boxed{\underbrace{w^*}_{D \times 1} = \underbrace{S_W^{-1}}_{D \times D} . (\underbrace{\mu_1 - \mu_2}_{D \times 1})}$$

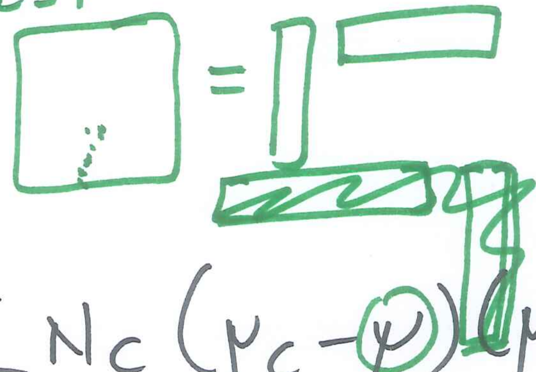
$$y_{ic} = \begin{cases} 1 & \text{if } y_i = c \\ 0 & \text{otherwise} \end{cases}$$

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

$$\boxed{z_i = w^T . x_i}$$

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

rank=1



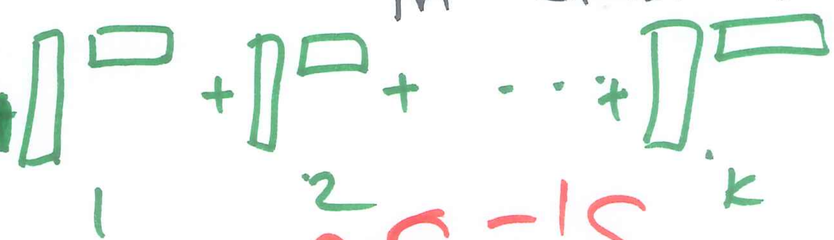
$$S_W = \sum_{c=1}^{K} S_c$$

class mean, overall mean

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

N_c → # of data points in class c.

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



$A . w = \lambda . w$

$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ $\; 0$

$W^* \Rightarrow$ the largest eigenvectors of $S_W^{-1} . S_B$

S_W: DxD, S_B: DxD

$$\boxed{\text{rank}(S_B) = K - 1}$$

$$\underline{\underline{A}} + \underline{\underline{B}} = C$$

$\underline{\underline{(K-1)}}$ nonzero $\underline{\underline{\text{eigenvalues}}}$

$$\mu = \frac{N_1 . \mu_1 + N_2 . \mu_2 + \ldots + N_K \mu_K}{N}$$

$$\underline{\underline{\mu}} = \frac{N_1}{N} . \mu_1 + \frac{N_2}{N} . \mu_2 + \ldots + \frac{N_K}{N} \mu_K .$$