Linear Discriminant Analysis

$$J(w) = \frac{(\hat{\mu}_1 - \hat{\mu}_2)^2}{s_1^2 + s_2^2}$$

w: projection vector

$$(\overset{1\times 1}{\hat{\mu}_1} - \overset{1\times 1}{\hat{\mu}_2})^2 = (\overset{1\times D}{w^T} \cdot \overset{D\times 1}{\mu_1} - \overset{1\times D}{w^T} \cdot \overset{D\times 1}{\mu_2})^2$$

$$= \left[w^T(\mu_1 - \mu_2)\right]\left[w^T \cdot (\mu_1 - \mu_2)\right]^T$$

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

$$(A \cdot B)^T = B^T \cdot A^T$$

$$= w^T \cdot \underbrace{\left[\overbrace{(\mu_1 - \mu_2)}^{D\times 1}\ \overbrace{(\mu_1 - \mu_2)^T}^{1\times D}\right]}_{S_B} \cdot w$$

S_B between class scatter matrix.

$$s_1^2 = w^T \cdot \underbrace{\left[\sum_{i=1}^{N} \underbrace{(x_i - \mu_1)(x_i - \mu_1)^T}_{\text{rank-one matrix}} \cdot y_i\right]}_{S_1} \cdot w$$

$$s_2^2 = w^T \cdot \underbrace{\left[\sum_{i=1}^{N} (x_i - \mu_2) \cdot (x_i - \mu_2)^T \cdot (1 - y_i)\right]}_{S_2} \cdot w$$

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

$$J(w) = \frac{w^T \cdot S_B \cdot w}{w^T \cdot \underbrace{[S_1 + S_2]}_{S_W} \cdot w} = \frac{w^T \cdot S_B \cdot w}{w^T \cdot S_W \cdot w}$$

S_W = within-class scatter matrix

$$\Rightarrow w^* = ?$$

$$\underbrace{w^*}_{D\times 1} = \underbrace{S_W^{-1}}_{D\times D} \cdot \underbrace{(\mu_1 - \mu_2)}_{D\times 1} \Longrightarrow z_i = (w^*)^T \cdot x_i$$

$$z_i = (W^*)^T \cdot x_i$$

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

$$z_i = \mathbf{W}^T \cdot x_i$$

→ can be projected to at most [K−1] dimensions where K is the # of classes.

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

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

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

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

class mean

overall mean

W^* = the largest eigenvectors of $S_W^{-1} \cdot S_B$

$$\frac{\text{rank of } S_W}{D} \qquad \frac{\text{rank of } S_B}{K-1}$$

Multidimensional Scaling (MDS)

Ankara – London $= d_{AL}$

Ankara – Paris $= d_{AP}$

London – Paris $= d_{LP}$

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

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

$\cancel{z_i = W^T x_i}$ → we have no access to x_i's.

$$d_{ij} = \|x_i - x_j\|_2 \qquad e_{ij} = \|z_i - z_j\|_2$$

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

$$D \cong E$$

Sammon Mapping (Sammon Stress)

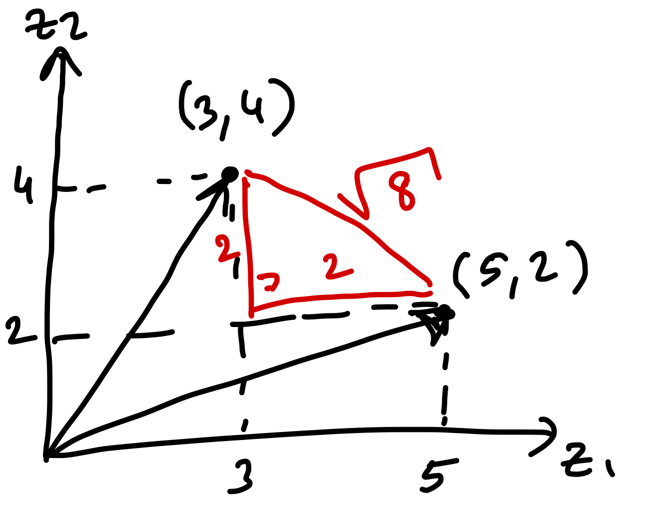
$y - \hat{y}$

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

$$\boxed{\text{minimize} \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{(d_{ij} - \|z_i - z_j\|_2)^2}{d_{ij}^2} \quad \text{with respect to: } z_i \in \mathbb{R}^{D'}}$$

Cannot perform out-of-sample embedding

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



$$(3\ 4)\begin{bmatrix}3\\4\end{bmatrix} - 2\,[3\ 4]\begin{bmatrix}5\\2\end{bmatrix} + [5\ 2]\begin{bmatrix}5\\2\end{bmatrix}$$

$$25 - 46 + 29 = 8$$

If we have access to x_i's

$$z_i = W^T \cdot x_i$$

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

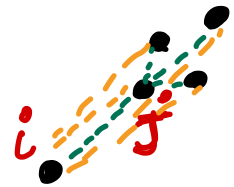
with respect to: $W \in \mathbb{R}^{D \times D'}$

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

t-Distributed Stochastic Neighbor Embedding (t-SNE)

SNE Algorithm :

$p_{j|i}$ = probability of x_j is a neighbor of x_i.



$$p_{j|i} = \frac{\exp\left[-\|x_i - x_j\|_2^2 / 2\sigma_i^2\right]}{\sum_{k \neq i} \exp\left[-\|x_i - x_k\|_2^2 / 2\sigma_i^2\right]}$$

$q_{j|i}$ = probability of z_j is a neighbor of z_i.

$$\boxed{KL(P \| Q) = \sum \sum p_{j|i} \cdot \log\left[\frac{p_{j|i}}{q_{j|i}}\right]}$$

t-SNE Algorithm:

$$\underline{\underline{p_{ij}}} = \frac{p_{j|i} + p_{i|j}}{2N}$$

$$\begin{bmatrix} 2 & 3 \\ 5 & 2 \end{bmatrix} \Rightarrow \begin{bmatrix} 2 & 4 \\ 4 & 2 \end{bmatrix}$$

$p_{j|i}$ $\quad$ p_{ij}

$$\underline{\underline{q_{ij}}} = \frac{\left(1 + \|z_i - z_j\|_2^2\right)^{-1}}{\sum_{k=1}^{N} \sum_{l \neq k} \left(1 + \|z_k - z_l\|_2^2\right)^{-1}}$$

$$KL(P \| Q) = \sum_{i=1}^{N} \sum_{j=1}^{N} p_{ij} \log\left[\frac{p_{ij}}{q_{ij}}\right]$$

$$KL_{\text{SYMMETRIC}} = \frac{KL(P\|Q) + KL(Q\|P)}{2}$$

t-SNE / UMAP