Metric Learning

Data $X \in \mathbb{R}^{n \times d} = \{x_1, x_2, .. x_n\}$ $x_i \in \mathbb{R}^d$

$$x_i = \begin{bmatrix} \text{height} \\ \text{weight} \\ \text{age} \\ \text{income} \end{bmatrix}$$

Input $= \{a_1, a_2 \ldots a_n\}$

"linear map"

$u(a_i) \to \mathbb{R}^k$

noteably $k = 2 \implies$ Draw picture

$$D(a_i, a_j) = \| u(a_i) - u(a_j) \|_2$$

Multimensional Dimensional Scaling (MDS)

Input either (1) n×n distance matrix

classical MDS $\longrightarrow$ $D \; : \; D_{ij} = d(a_i, a_j)$

(2) function $f(a_i, a_j) = d(a_i, a_j)$

Goal: Embedding $\{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^k$

s.t. $\forall x_i, x_j \quad d(a_i, a_j) \simeq \|x_i - x_j\|_2$

assume $d(a_i, a_j) \leftarrow$ derived from Euclidean

$$\|a_i - a_j\|^2 = \|a_i\|^2 + \|a_j\|^2 - 2\langle a_i, a_j\rangle$$

$$\langle a_i, a_j\rangle = \frac{1}{2}\left(-\|a_i - a_j\|^2 + \|a_i\|^2 + \|a_j\|^2\right)$$

if I know $A = \left[\quad \right] \in \mathbb{R}^{n \times d}$

$M = AA^T \in \mathbb{R}^{n \times n}$

answer!

$\text{eigs}(M) = UVU^T \qquad \boxed{U_k} \in \mathbb{R}^{n \times k}$

$$M_{ij} = \langle a_i, a_j \rangle = \frac{1}{2}\left(\underbrace{\|a_i - a_j\|^2}_{D_{ij}^2} + \underbrace{\|a_i\|^2}_{D_{i1}^2} + \underbrace{\|a_j\|^2}_{D_{j1}^2} \right)$$

Euclidean embedding

shift invariant

↳ $x_1 = 0 \leftarrow$ the origin $\Longrightarrow \|a_1\| = 0$

$$\|a_i - a_1\|^2 = \underbrace{\|a_i\|^2}_{D_{i,1}^2}$$

Linear Discriminant Analysis LDA

$k<d$

Input: $X \in \mathbb{R}^{n\times d}$ assume $d(x_i, x_j) = \|x_i - x_j\|_2$

clusters $S_1, S_2 \dots S_\ell \subset X$ $S_i \cap S_j = \phi$

$\cup S_i = X$

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

mean $\mu_i = \frac{1}{|S_i|} \sum_{x \in S_i} x$

covariance $\Sigma_i = \frac{1}{|S_i|} \sum_{x\in S_i} (x-\mu_i)(x-\mu_i)^T \in \mathbb{R}^{d\times d}$

t-SNE

between class covariance $\Sigma_B = \frac{1}{|X|} \sum_{i=1}^{\ell} |S_i| (\mu_i - \mu)(\mu_i - \mu)^T$

within class covariance $\Sigma_w = \frac{1}{|X|} \sum_{i=1}^{\ell} |S_i| \Sigma_i$

Find vectors u which maximize

$$\frac{u^T \hat{\Sigma}_B u}{u^T \hat{\Sigma}_W u} \qquad \text{Set } \{u_1, \dots u_k\} \quad k < \ell - 1$$

Answer top k eigenvectors of

$$\hat{\Sigma}_W^{-1} \hat{\Sigma}_B$$

Distance Metric Learning

Input $X \in \mathbb{R}^{n \times d}$ ⇒ Mahalanobis Distance

$$d_M(p, q) = \sqrt{(p-q)^T M (p-q)}$$

$p, q \in \mathbb{R}^d$

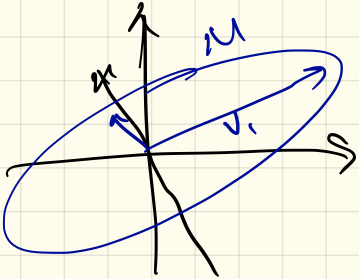
$M \in \mathbb{R}^{d \times d}$

$M = I \Rightarrow$ Euclidean

$$M = \begin{bmatrix} M_{11} & & 0 \\ & M_{22} & \\ 0 & & \ddots M_{dd} \end{bmatrix}$$

weight 2nd coord

general



Input:

$X \in \mathbb{R}^{n\times d}$

close pairs $C \subset X \times X$

far pairs $F \subset X \times X$

Goal $M \in \mathbb{R}^{d\times d}$

s.t. in d_M

C close

F far

JMLR 12 Ying + Li

$$M^* = \underset{M\in\mathbb{R}^{d\times d}}{\arg\max} \; \min_{\{x_i,x_j\}\in F} d_M(x_i,x_j)^2$$

$$\text{s.t.} \quad \sum_{\{x_i,x_j\}\in C} d_M(x_i,x_j)^2 \le K$$

- $$H = \sum_{\{x_i, x_j\} \in C} (x_i - x_j)(x_i - x_j)^T \in \mathbb{R}^{d \times d}$$

  Full Rank : o.w. $H = H + \overset{\overset{\text{small}}{\downarrow}}{\delta} I$

- Restrict Trace$(M) = d$ $\qquad Tr(M) = \sum_{i=1}^{d} eig(M, i)$

  $M \in \mathbb{P}$ $\qquad Tr(I) = d$

- $$\Delta = \{\alpha \in \mathbb{R}^{|F|} \mid \sum_i \alpha_i = 1 \text{ and } \alpha_i \geq 0\}$$

- $T = T_{i,j} \in F$ $\qquad X_T = X_{T_{i,j}} = (x_i - x_j)(x_i - x_j)^T \in \mathbb{R}^{d \times d}$

  $$\tilde{X}_T = H^{-1/2} X_T H^{-1/2}$$

$$M^* = \arg\max_{M \in \mathbb{P}} \min_{\alpha \in \Delta} \sum_{T \in F} \alpha_T \overbrace{\langle \tilde{X}_T, M \rangle}^{d_M(X_T)}$$

$\sigma = d \cdot 10^{-5}$ gradient start $M = I$

$$g_\sigma(M) = \frac{\sum_{\tau \in F} \exp\left(-\langle \tilde{X}_\tau, M\rangle / \sigma\right) \tilde{X}_\tau}{\sum_{\tau \in F} \exp\left(-\langle \tilde{X}_\tau, M\rangle / \sigma\right)} \in \mathbb{R}^{d \times d}$$

F-W DML

```
0. Init $M_0 = I$
1. for $t = 1, 2, \ldots T$ do
    a. Set $G = g_\sigma(M_{t-1})$
    b. Let $v_t = V_\sigma M_{t-1} \leftarrow$ max eigenvector$(G)$
    c. Update $M_t = \frac{t-1}{t} M_{t-1} + \frac{1}{t} v_t v_t^T$
Return $M = M_T$
```