

L17: Distance Metric Learning

MDS, LDA, DML

Data Mining
Oct 29, 2026
Jeff M. Phillips



Input X data, size n

Distance $D: X \times X \rightarrow \mathbb{R}$
↗ unsure about

Like to have data in $\mathbb{R}^k$ but too big
Dim-Reduction $X \in \mathbb{R}^{d \times n} \Rightarrow \tilde{T}(x_i) \mapsto g_i \in \mathbb{R}^k$

$$D(x_i, x_j) = \|g_i - g_j\|$$

Input distance, correct distance $\neq D(x_i, x_j) \neq \|x_i - x_j\|$

<u>Ex1</u> $X \in \mathbb{R}^d$ $x_i = (\text{height}, \text{weight}, \text{age}, \dots)$	<u>Ex2</u> X cities $D(x_i, x_j) = \text{cost to fly on airplane}$
--	---

Input set X

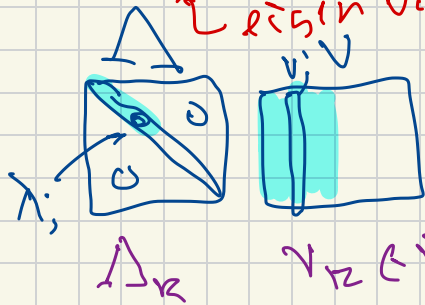
Similarity $s: X \times X \rightarrow \mathbb{R} \stackrel{?}{=} [0, 1]$

$$S \in \mathbb{R}^{n \times n} \quad S_{ij} = s(x_i, x_j)$$

eigenvalues $\lambda_1, \dots, \lambda_n$

$$[\Lambda, V] = \text{eig}(S)$$

eigen vectors $v_1, v_2, \dots$



$$S = V \Lambda V^T$$

$$\text{if } S = A A^T \\ A = V \Lambda^{1/2}$$

$$Q = V \Lambda^{1/2} \in \mathbb{R}^{n \times k} \\ = A$$

if x_i, x_j similar, $s(x_i, x_j)$ large

if x_i, x_j not similar, $s(x_i, x_j)$ small

Graph $s(x_i, x_j) = 1$ if edge
 $= 0$ otherwise

$$K(x_i, x_j) = e^{-\|x_i - x_j\|^2 / w^2}$$

$$\langle x_i, x_j \rangle = \sum_{l=1}^d x_{il} x_{jl}$$

something given as input

Multi-Dimensional Scaling (MDS)

Input X size n

trust $D: X, X \rightarrow \mathbb{R}$

Matrix $D \in \mathbb{R}^{n \times n}$

$$D_{ij} = D(x_i, x_j)$$

Goal map: $X \rightarrow Q \subset \mathbb{R}^{n \times k}$

$$D(x_i, x_j) \approx \|g_i - g_j\|$$

pretend $a_i = 0 \in \mathbb{R}^d$

$$\hookrightarrow \|a_i\|^2 = \|a_i - a_i\|^2 = D(a_i, a_i)^2$$

$$\|a_j\|^2 = D(a_j, a_i)^2$$

not input $\rightarrow$ if data $A \in \mathbb{R}^{n \times d}$
 n d
 A
 $AA^T = S$
 $S_{ij} = \langle a_i, a_j \rangle$

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

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

classical MDS

centering step

$$X \in \mathbb{R}^n \Rightarrow A \in \mathbb{R}^d \text{ centered}$$

centering matrix $C_n = I - \frac{1}{n} \mathbf{1}\mathbf{1}^T$ $\text{mean}(A) = 0$

$$A = C_n X$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

cMDS

1. Double Centering

$$S = -\frac{1}{2} C_n D^{(2)} C_n$$

2. eigen decomp $[\Lambda, V] = \text{eig}(S)$

3. Refun $Q = U_{12} \Lambda_{12}^{1/2}$

$$D_{ij}^{(2)} = (D_{ij})^2$$

Fisher Linear Discriminant Analysis (LDA)

Input $X \subset \mathbb{R}^d$ do not trust Euclidean

Clusters $S_1, S_2, \dots, S_k \subset X$ $S_i \cap S_j = \emptyset$

same clusters $\rightarrow$ close $\nearrow$ diff clusters $\rightarrow$ far

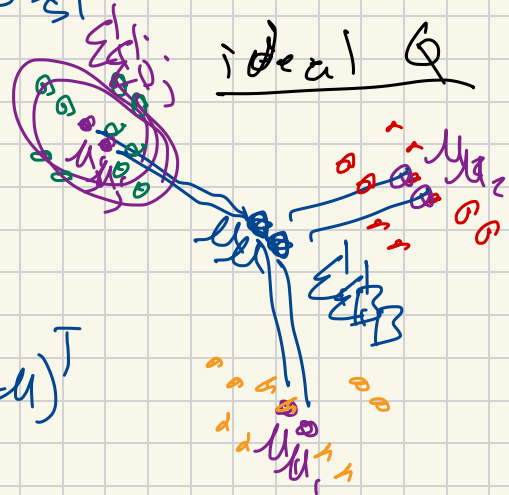
• centre $\mu_j = \frac{1}{|S_j|} \sum_{x \in S_j} x$

• covariance $\Sigma_j = \frac{1}{|S_j|} \sum_{x \in S_j} (x - \mu_j)(x - \mu_j)^T$

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

• class cov $\Sigma_B = \frac{1}{|X|} \sum_{j=1}^k |S_j| (\mu_j - \mu)(\mu_j - \mu)^T$

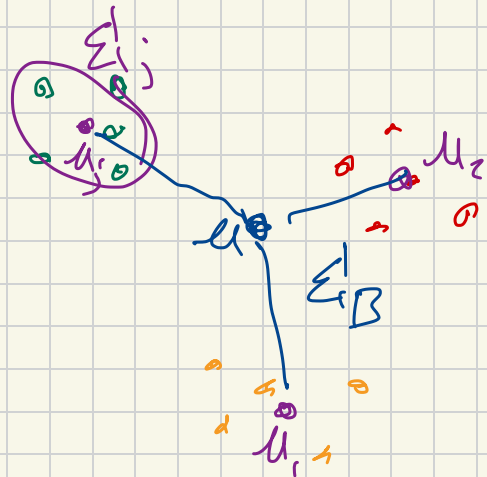
• within class cov $\Sigma_W = \frac{1}{|X|} \sum_{j=1}^k |S_j| \Sigma_j$



① Between class cov $\Sigma_B \Rightarrow \text{large}$

for $v \in \mathbb{R}^d$
 $\|v\|=1$

$$v^T \Sigma_B v \Rightarrow \text{large}$$



② Within class cov $\Sigma_W \Rightarrow \text{small}$

$$v^T \Sigma_W v \Rightarrow \text{small}$$

Find $v \in \mathbb{R}^d$ $\|v\|=1 \Rightarrow \frac{v^T \Sigma_B v}{v^T \Sigma_W v} \Rightarrow \text{large}$

$\hookrightarrow$ eigenvector $\Sigma_W^{-1} \Sigma_B$

LDA: find to k eigenvectors $v_k = [v_1, \dots, v_k]$ of $\Sigma_W^{-1} \Sigma_B$

$$Q = \pi_{v_k} [X] = (Xv_1, Xv_2, \dots, Xv_k)$$

non-linear Dimensionality Reduction

- t-SNE

- UMAP

- Kernel PCA

- try to preserve

↳ trust Euclidean in X

- if X has very tight clusters

→ are preserved.

- not generalizable to new data. No π

Prof Phillips
not a fan!

$$s = (x_i, x_j) =$$

$$K(x_i, x_j) = e$$

$$- \|x_i - x_j\|^2 / \omega^2$$

↑
some
parameter

Linear Distance Matrix Learning

Input $X \in \mathbb{R}^d$, do not know Euclidean

- close pairs $C = \{x_i, x_j\}$

- far pairs $F = \{x_i, y_j\}$

$$D_M(g_i, g_j) = \sqrt{(g_i - g_j)^T M (g_i - g_j)} \quad \text{Mahalanobis}$$

goal learn M

maybe M rank $k \Rightarrow \text{dim-reduced}$

Low Rank Mahal dist M

$M_{12} \leftarrow$ best rank k approx

$$\text{eig}(M) = [\Lambda, V] \quad M_{12} = V_{12} \Lambda_{12} V_{12}^T$$

$$A = \Lambda^T V^T = M = f^T A$$