L17: Metric Learning

10.27.2025

Data Mining

Jeff M. Phillips

Data Mining

Input X data, size n

Distance $D : X \times X \to \mathbb{R}$

Like $X \subset \mathbb{R}^d$ d not too big

$D(p, q) = \|p - q\|$

X cities

Dist = cost airfare to fly

$X \subset \mathbb{R}^d$

$x_i = (\text{height}, \text{weight}, \text{age})$

<u>Input</u> set X, similarity $s: X \times X \to \mathbb{R}$

if x, x' similar $s(x, x')$ large

x, x' not similar $s(x, x')$ small

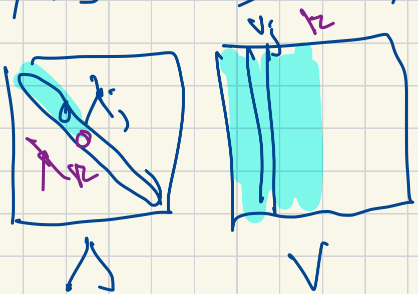
$S \in \mathbb{R}^{n \times n}$

eigenvalues

$S_{ij} = s(x_i, x_j)$

eigenvectors

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



- $K(x, x') = e^{-\|x - x'\|^2 / \omega^2}$
- Graph $s(x, x') = 1$ if edge
  $= 0$ o.w.
- $\langle x, x' \rangle = \sum_j x_j x'_j$

$\Phi \in \mathbb{R}^{n \times k}$ $\qquad \Phi = V_k \Lambda_k^{1/2}$

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

MultiDimensional Scaling (MDS)

Input X ~~=~~ $A \in \mathbb{R}^{n \times d}$ size n

$D : X \times X \to \mathbb{R}$

Matrix $D \in \mathbb{R}^{n \times n}$ $\quad D_{ij} = D(x_i, x_j)$

Goal map : $X \to Q$ $\quad u(x_i) \to q_i \in \mathbb{R}^k$ k small

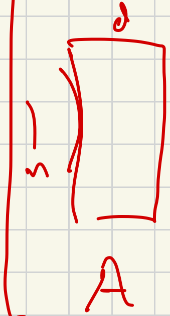
$$D(x_i, x_j) \approx \|q_i - q_j\| \qquad Q \subset \mathbb{R}^k$$

$$\|a_i - a_j\|^2 = \|a_i\|^2 + \|a_j\|^2 - 2\langle a_i, a_j \rangle \qquad AA^T = S$$

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

put a_1 at 0 $\Rightarrow \|a_i\|^2 = \|a_i - a_1\|^2 \equiv D(a_i, a_1)^2$

if data $A \in \mathbb{R}^{n \times d}$



$$AA^T = S$$

$$S_{ij} = \langle a_i, a_j \rangle$$

classical MDS

centering step $\quad X \in \mathbb{R}^d \Longrightarrow A \in \mathbb{R}^d$

A is centered.

centering matrix $\quad C_n = \underline{I} - \frac{1}{n} \underbrace{\underline{1}\,\underline{1}^T}_{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}}$

$$A = C_n X$$

1. Double Centering
$$S = -\frac{1}{2} C_n D^{(2)} C_n \qquad D^{(2)}_{ij} = (D_{ij})^2$$

2. eigendecomp. $\quad [\Lambda, V] = \mathrm{eig}(S)$

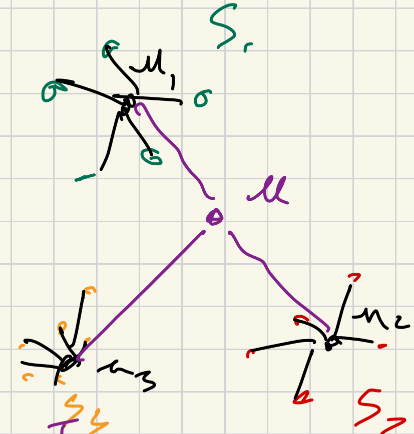
3. $Q = V_{1:2} \Lambda^{1/2}_{1:2}$

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 cluster $\Rightarrow$ close // diff cluster $\Rightarrow$ far



center $\mu_j = \frac{1}{|S_j|} \sum_{x_i \in S_j} x_i$

covariance $\Sigma_j = \frac{1}{|S_j|} \sum_{x_i \in S_j} (x_i - \mu_j)(x - \mu_j)^T$ $\underbrace{\in \mathbb{R}^{d \times d}}$

mean $\mu = \frac{1}{n} \sum_{i=1} x_i$

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

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

Quantify ① Between class Σ_B large

if for $u \in \mathbb{R}^d$ $u^T \Sigma_B u$ is large
$\|u\|$

② Within class cov to be small

if $u^T \Sigma_w u$ small

---

find $u \in \mathbb{R}^d$ $\|u\| = 1$ so $\dfrac{u^T \Sigma_B u}{u^T \Sigma_w u}$ large

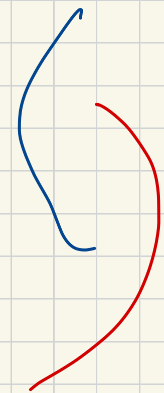
$\hookrightarrow$ top eigenvector of $\left[\Sigma_w^{-1} \Sigma_B\right]$

LDA find top k eigvectors $U_k = \{u_1, u_2, \dots u_k\}$

$$Q = \pi_{U_k}(X)$$

nonlinear Dimensionality Reduction

- t-SNE
- UMAP



Prof Phillips

not a big

fan.

Linear Distance Metric Learning

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

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

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

Mahalanobis

goal learn M.

Low-rank Mahalanobis dist M.

$M_k \leftarrow$ best rank k approx

$\text{eigs}(M) = [\Lambda, V] \qquad M_k = V_k \Lambda_k V_k^T$

$A = \Lambda^{1/2} V^T \qquad M = A^T A$

$A_k = \Lambda_k^{1/2} V_k^T \qquad M_k = A_k^T A_k$

$q_i = \bar{\pi}(x_i)$

$q_i = A_k x_i$

$$\|q_i - q_j\| = \sqrt{(Ax_i - Ax_j)^T (Ax_i - Ax_j)}$$

$$= \sqrt{(x_i - x_j)^T A^T A (x_i - x_j)}$$

$$= \sqrt{(x_i - x_j)^T M (x_i - x_j)} = D_M(x_i, x_j)$$

small $\alpha > 0$

$$M = \left(\alpha I + \frac{1}{n_C} \sum_{x_i, x_i' \in C} (x_i - x_i')(x_i - x_i')^T - \frac{1}{n_F} \sum_{x_j, x_j' \in F} (x_j - x_j')(x_j - x_j')^T \right)^{-1}$$