Metric Learning

Mar 5, 2018

Data : $X: \{x_1, x_2, \ldots x_n\}$

map

$f: X \to \mathbb{R}^k$ so $\|f(x_i) - f(x_j)\|$

Input

1 MDS — all distances

2 LDA — clusters

3 Dist Metric Learning — Pairs: C close, F far

Normalization

$$X = \{x_1, x_2, \dots x_n\}$$

$$x_i \in \mathbb{R}^d \qquad x_i = \{x_{i1}, x_{i2} \,..\, x_{id}\}$$

→ Make all coordinates $x_{ij} \in [0,1]$

$$t_j = \max_{i\in[n]} x_{i,j} \qquad\qquad b_j = \min_{i\in[n]} x_{ij}$$

$$\downarrow \qquad\qquad\qquad\qquad \downarrow$$

$$1 \qquad\qquad\qquad\qquad 0$$

→ Make all coordinates $\{x_{ij}\}_i \sim N(0,1)$

$$\operatorname*{mean}_{i\in[n]} \{x_{ij}\} = 0 \qquad \operatorname*{Var}_{i\in[n]} \{x_{ij}\} = 1$$

Multidimensional Scaling (MDS)

~~Start $X \subset \mathbb{R}^d$ d very large~~

Really ① $f: X \times X \to \mathbb{R}$ (distance)

② $D \in \mathbb{R}^{n \times n}$ $D_{ij} = f(x_i, x_j)$ $\longrightarrow \mathbb{R}^k$

- Pretend $D_{ij} = \|a_i - a_j\|$

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$$\|a_i - a_j\|^2 = \|a_i\|^2 + \|a_j\|^2 - 2\boxed{\langle a_i, a_j \rangle} \longrightarrow \boxed{AA^T}_{ij}$$

solution

top k eigenvector of AA^T

solve for

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

$\|a_1\| = 0$

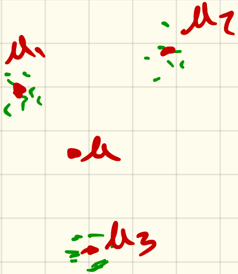
$\|a_i\| = D_{i,1}$

Linear Discriminant Analysis

$X \subset \mathbb{R}^d$ (multiclass)

<u>Input</u> X : k clusters $S_1, S_2, \dots S_k$ $S_i \subset X$

$S_i \cap S_j = \emptyset$



$X \mapsto \mathbb{R}^{k'}$

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

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

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

between class covariance

$$\Sigma_B = \frac{1}{|X|} \sum_{j=1}^{k} |S_i| (\mu_i - \mu)(\mu_i - \mu)^T$$

within class covariance

$$\Sigma_W = \frac{1}{|X|} \sum_{j=1}^{k} |S_i| \Sigma_i = \frac{1}{|X|} \sum_{j=1}^{k} \sum_{x \in S_j} (x - \mu_j)(x - \mu_j)^T$$

$$\left(\max_{u \in \mathbb{R}^d} \frac{u^T \Sigma_B u}{u^T \Sigma_W u} \right)$$

want small

Solution

top k' eigenvectors of $\Sigma_W^{-1} \Sigma_B \in \mathbb{R}^{d \times d}$

$\mapsto$ basis $U = \{u_1, u_2 \dots u_{k'}\}$

map $x \Rightarrow \pi_U(x)$

use $k' \leq k$

Distance Metric Learning

(Ying + Li JMLR 12)

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

two sets of pairs $\underset{\text{close}}{C} \subset X \times X$ $\underset{\text{far}}{F} \subset X \times X$

Mahalanobis Distance

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

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

Goal

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

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

same as $M = I$

if $M = \begin{bmatrix} m_1 & & & 0 \\ & m_2 & & \\ & & \ddots & \\ 0 & & & m_d \end{bmatrix}$

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

close

$$\mathrm{Trace}(M) = \mathrm{Tr}(M) = \sum_{i=1}^{d} \mathrm{eigenvalues}_i(M)$$

$$\mathbb{P} = \{ \text{all psd matrices } M \text{ s.t. } \mathrm{tr}(M) = d \}$$

$$I \in \mathbb{P}$$

for

$$\Delta = \{ \alpha \in \mathbb{R}^{|F|} \mid \sum a_i = 1 \ \& \text{ all } \alpha_i \geq 0 \}$$

$T_{ij} \in F$ define $\cancel{X}_{T_{ij}} = \cancel{X}_{ij} = (x_i - x_j)(x_i - x_j)^T \in \mathbb{R}^{d \times d}$

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

goal now equivalent:

$$\underset{M \in \mathbb{P}}{\arg\max} \ \min_{\alpha \in \Delta} \sum_{T \in F} \alpha_T \underbrace{\langle \tilde{X}_T, M \rangle}_{= d_M(\tilde{x}_T)}$$

Set $\sigma = d \cdot 10^{-5}$

$$g(M) = \frac{\sum_{T \in F}^{1} \exp\left(-\langle \bar{X}_T, M \rangle / \sigma\right) \bar{X}_T}{\sum_{T \in F}^{1} \exp\left(-\langle \bar{X}_T, M \rangle / \sigma\right)}$$

$v_{\sigma, M}$ = top eigenvector $g(M)$

0. Init $M_0 = I$

1. for $t = 1, 2, \ldots T$

   1a $G = g(M)$

   1b $v_t = v_{\sigma, M}$

   1c $M_t = \frac{t-1}{t} M_{t-1} + \frac{1}{t} v_t v_t^T$

2. return M_T

MG

$$\boxed{f_q - \varepsilon n \le \hat{f}_q \le f_q}$$

$n = 1$ million

$n =$ length stream

$\varepsilon = \frac{1}{k+1}$

$k = 9 \qquad \varepsilon = 0.1$

$\varepsilon n =$

$$\mathrm{eig}(H) = U S U^T$$

$$S = \begin{pmatrix} s_1 & & 0 \\ & s_2 & \\ 0 & & \ddots \\ & & & s_d \end{pmatrix}$$

$$S^{-1/2} = \begin{pmatrix} s_1^{-1/2} & & 0 \\ & s_2^{-1/2} & \\ 0 & & \ddots \\ & & & s_d^{-1/2} \end{pmatrix}$$

$$H^{-1/2} = U S^{-1/2}$$