子空间算法

① 主成分分析（PCA）

设 $x_1\ x_2 \cdots x_p$ 为训练样本，每个 x_i 为 N 维。

寻找一个 $M\times N$ 维矩阵 A，使 X 变为 AX，即将 X 的维度由 N 降到 M。

PCA 的要求：

① $A = \begin{bmatrix} — a_1 — \\ — a_2 — \\ \vdots \\ — a_M — \end{bmatrix}$ 其中 a_i 为 $1\times N$ 维。

② $a_i a_j^T = \begin{cases} 1 & i=j \text{ 时} \\ 0 & i\neq j \text{ 时} \end{cases}$ （正交性）

③ 方差最大。

即：ⓐ 寻找 a_1，使

$$\sum_{i=1}^{p} \| a_1 (x_i - \bar{x}) \|^2 \text{ 最大。}$$

$$\text{maximize} \quad \sum_{i=1}^{p} a_1 (x_i - \bar{x})(x_i - \bar{x})^T a_1^T = a_1 \Sigma a_1^T$$

$$\text{subject to:} \quad a_1 a_1^T = 1$$

其中 $\Sigma = \sum_{i=1}^{p} (x_i - \bar{x})(x_i - \bar{x})^T$ 为协方差矩阵

拉格朗日乘子法：

$$E(a_1) = a_1 \Sigma a_1^T - \lambda (a_1 a_1^T - 1)$$

$$\frac{\partial E}{\partial a_1} = \Sigma a_1^T - \lambda a_1^T = 0 \Rightarrow \Sigma a_1^T = \lambda a_1^T$$

而 $a_1 \Sigma a_1^T = a_1 (\lambda a_1^T) = \lambda$

因此，λ为Σ的最大特征值，a_1^T为最大特征值对应的特征向量。

ⓑ 求a_2，a_2满足条件为：

$$\text{Maximize} \quad \sum_{i=1}^{p} \|a_2(x_i-\overline{x})\|^2$$

$$\text{s.b.} \quad a_2a_2^T=1$$

$$a_2a_1^T=0$$

拉格郎日乘子法：

$$E(a_2)=a_2\Sigma a_2^T-\lambda(a_2a_2^T-1)-\beta a_1a_2^T$$

$$\frac{\partial E(a_2)}{\partial a_2}=\Sigma a_2^T-\lambda a_2^T-\beta a_1^T=0$$

① 先证明 $\beta=0$，这是因为：

$$(\Sigma a_2^T-\lambda a_2^T-\beta a_1^T)^T=0$$

$$a_2\Sigma^T-\lambda a_2-\beta a_1=0$$

又由于 $\Sigma^T=\Sigma$，即

$$a_2\Sigma-\lambda a_2-\beta a_1=0$$

两边乘以a_1^T，得：

$$a_2\underbrace{\Sigma a_1^T}_{=\lambda_1 a_1^T}-\lambda\underbrace{a_2a_1^T}_{=0}-\beta\underbrace{a_1a_1^T}_{=1}=0$$

所以 $\lambda_1a_2a_1^T-\lambda a_2a_1^T-\beta=0$

因此 $\beta=0$

$$\frac{\partial E(a_2)}{\partial a_2}=\Sigma a_2^T-\lambda a_2^T=0$$

因此 λ为Σ的第二大特征值，a_2^T为Σ第二大特征值对应的特征向量。

ⓒ a_3为Σ第三大特征值对应的特征向量，依次类推。

Singular Value Decomposition

$$\underset{n\times p}{M} = \underset{n\times n}{U}\ \underset{n\times p}{S}\ \underset{p\times p}{V^T}$$

$$\begin{cases} U^TU = I_{n\times n} \\ V^TV = I_{p\times p} \\ S: \text{对角阵} \end{cases}$$

证明: U^T 为 MM^T 的特征向量, V^T 为 (M^TM) 的特征向量。

证明:
$$MM^T = (USV^T)(USV^T)^T$$
$$= USV^TVSU^T$$
$$= US^2U^T \quad (S^2 \text{为} S \text{对角线上非零元素平方})$$

则有: $(MM^T)U^T = S^2U^T$

同理: $M^TM = VS^2V^T$

则有: $(M^TM)V^T = S^2V^T$

$$M = \overbrace{\begin{bmatrix} | & | & | & & | \\ x_1-\mu & x_2-\mu & x_3-\mu & \cdots & x_p-\mu \\ | & | & | & & | \end{bmatrix}}^{p\text{个}} \Bigg\} n\text{维}$$

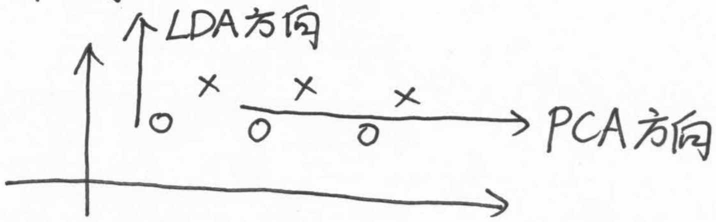
$$\underset{n\times n}{\Sigma} = \underset{n\times p}{M}\ \underset{p\times n}{M^T}$$

svd分解: $M = USV^T$

U^T 为特征向量, S^2 对角线元素为特征值。

Linear Discriminant Analysis

① 举例



② 两类问题：

找一个矩阵 W，使

$$\widetilde{X} = WX,$$

$$\widetilde{\mu_1} = \frac{1}{N_1}\sum_{X\in C_1}\widetilde{X} \qquad \widetilde{\mu_2} = \frac{1}{N_2}\sum_{X\in C_2}\widetilde{X}$$

$$\widetilde{S_1} = \sum_{X\in C_1}(\widetilde{X}-\widetilde{\mu_1})^T(\widetilde{X}-\widetilde{\mu_1})$$

$$\widetilde{S_2} = \sum_{X\in C_2}(\widetilde{X}-\widetilde{\mu_2})^T(\widetilde{X}-\widetilde{\mu_2})$$

要使 $\dfrac{\|\widetilde{\mu_1}-\widetilde{\mu_2}\|^2}{\widetilde{S_1}^2+\widetilde{S_2}^2}$ 最大。

$$\|\widetilde{\mu_1}-\widetilde{\mu_2}\|^2 = W\left[(\mu_1-\mu_2)(\mu_1-\mu_2)^T\right]W^T$$

其中 $\mu_1 = \frac{1}{N_1}\sum_{X\in C_1} X \qquad \mu_2 = \frac{1}{N_2}\sum_{X\in C} X$

$$\widetilde{S_1}^2 = W\left[\sum_{X\in C_1}(X-\mu_1)(X-\mu_1)^T\right]W^T$$

$$\widetilde{S_2}^2 = W\left[\sum_{X\in C_2}(X-\mu_2)(X-\mu_2)^T\right]W^T$$

$$\cancel{\widetilde{S_1}^2+\widetilde{S_2}^2} =$$

设 $S_B = (\mu_1-\mu_2)(\mu_1-\mu_2)^T$

$$S_W = \sum_{X\in C_1}(X-\mu_1)(X-\mu_1)^T + \sum_{X\in C_2}(X-\mu_2)(X-\mu_2)^T$$

则有:

最大化 : $E(W) = \dfrac{WS_BW^T}{WS_WW^T}$

$$\frac{\partial E(W)}{\partial W} = 0 \Rightarrow$$

$$WS_WW^T \cdot S_BW^T - WS_BW^T \cdot S_WW^T = 0$$

$$S_BW^T - \left(\frac{WS_BW^T}{WS_WW^T}\right) S_W W^T = 0$$

$$S_BW^T - c\, S_WW^T = 0$$

$$S_BW^T = c\, S_WW^T$$

$$\left(S_W^{-1} S_B\right) W^T = c\, W^T$$

$$W^* \propto S_W^{-1}(\mu_1 - \mu_2)$$

③ 多类问题

如果有C类，则我们要找C-1个方向。

$$S_i = \sum_{x \in C_i} (x - \mu_i)(x - \mu_i)^T$$

$$S_W = \sum_{i=1}^{C} S_i$$

$$S_B = \sum_{i=1}^{C} N_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (\text{秩 } C-1)$$

最大化 $\dfrac{|W^TS_BW|}{|W^TS_WW|}$ $(Y = W^TX)$

$$\left(S_W^{-1} S_B\right) w_i^* = \lambda_i w_i^*$$

最多只有C-1个非零特征值。