Statistical Learning HW P.

丁泽微 18307110088

证明题部分:

1.1 解: $\because \rho(Y_k, X_i) = \frac{\sqrt{\lambda_k}\alpha_{ki}}{\sqrt{\sigma_{ii}}}$ 且 λ_k 为 Σ 特征值, $\sigma_{ii} = Var(X_i) = \Sigma_{ii}$ (Σ的第i,i个元素)

$$\therefore \sum_k \rho^2(Y_k, X_i) = \sum_{k=1}^m \frac{1}{\sigma_{ii}} \cdot \lambda_k \cdot \alpha_{ki}^2$$

$$= \frac{1}{\sigma_{ii}} \sum_{k=1}^m \lambda_k \cdot \alpha_k^T e_i e_i^T \alpha_k$$ $\because \alpha_k$为λ_k对应特征向量 $\therefore \Sigma\alpha_k = \lambda_k\alpha_k$

$$= \frac{1}{\sigma_{ii}} \sum_{k=1}^m \alpha_k^T e_i e_i^T (\lambda_k \alpha_k)$$

$$= \frac{1}{\sigma_{ii}} \sum_{k=1}^m \alpha_k^T e_i e_i^T \Sigma \alpha_k$$ 令$A = (\alpha_1, \cdots, \alpha_m)$, 且 则$(A^TA)_{ii} = \alpha_i^T\alpha_i$

$$\cancel{= \frac{1}{\sigma_{ii}} \sum_{k=1}^m tr\{\alpha_k\alpha_k \cdot e_i^T \Sigma e_i}$$

$$= \frac{1}{\sigma_{ii}} tr\{A^T e_i e_i^T \Sigma A\}$$

$\because e_i$为基本单位向量, $\therefore e_ie_i^T \in R^{m\times m}$ 且仅有第(i,i)个元素为1, 其余均为0

$\therefore e_ie_i^T\Sigma$ 仅保留Σ的第(i,i)个元素, 即σ_{ii}, 其余均为0

$\therefore A^Te_ie_i^T\Sigma A$ 中仅保留第(i,i)个元素, 其余均为0

$\therefore tr\{A^Te_ie_i^T\Sigma A\} = \alpha_i^T\sigma_{ii}\alpha_i = \sigma_{ii}\alpha_i^T\alpha_i = \sigma_{ii}$ $\therefore$得 $\sum_{k=1}^m \rho^2(Y_k, X_i) = \frac{1}{\sigma_{ii}} \cdot \sigma_{ii} = 1$

证毕.

#

1.2 解. ① 将X看成六个二维样本; 则先进行标准化后, 得: (对每一维, 即每一行标准化)

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_i}{\sqrt{s_{ii}}} \quad (*)$$

S为样本协方差阵, $S = \frac{1}{n-1}\sum_j (x_j - \bar{x})(x_j - \bar{x})^2$; 欲得S, 先计算中心化后的样本:

由数据, $\bar{x}_1 = \frac{1}{6}(2+3+3+4+5+7) = \frac{24}{6} = 4$ $\bar{x}_2 = \frac{1}{6}(2+4+5+5+6+8) = \frac{30}{6} = 5$

$\therefore$ 中心化后的样本为: $X_{中心化} = \begin{pmatrix} -2, & -1, & -1, & 0, & 1, & 3 \\ -3, & -1, & 0, & 0, & 1, & 3 \end{pmatrix}$

$\therefore$ 样本协方差阵 $S = \hat{\Sigma} = \frac{1}{n-1} X_{中心化} X_{中心化}^T = \frac{1}{5} \cdot \begin{pmatrix} 16 & 17 \\ 17 & 20 \end{pmatrix}$

$\therefore \sqrt{S_{11}} = \frac{4}{\sqrt{5}}$ $\sqrt{S_{22}} = 2$

② $\therefore$ 代入(*)式, 得标准化后数据为: $X^* = \begin{pmatrix} -1.12, & -0.56, & -0.56, & 0, & 0.56, & 1.68 \\ -1.5, & -0.5, & 0, & 0, & 0.5, & 1.5 \end{pmatrix}$

∴ 标准化后, $\hat{\Sigma}^* = S^* = R = \frac{1}{n-1} X^* \cdot X^{*T} = \begin{pmatrix} 1 & 0.95 \\ 0.95 & 1 \end{pmatrix}$ (R为相关阵)

② ∴ 对 $\hat{\Sigma}^*$ 特征分解: $\hat{\Sigma}^* = Q\Lambda Q^T \Rightarrow Q = \begin{pmatrix} 0.71 & -0.71 \\ 0.71 & 0.71 \end{pmatrix}, \Lambda = \begin{pmatrix} 1.95 & \\ & 0.05 \end{pmatrix}$ (近似值)

∴ 主成分分析得到的投影阵(即解)为 $Q = \begin{pmatrix} 0.71 & -0.71 \\ 0.71 & 0.71 \end{pmatrix}$ $a_1 = \begin{pmatrix} 0.71 \\ 0.71 \end{pmatrix}$ $a_2 = \begin{pmatrix} -0.71 \\ 0.71 \end{pmatrix}$

按 kaiser 准则, 第二主成分方差贡献率过低、方差也很低 (0.05, 即λ_2), ∴舍去

∴ 选取第一主成分: $Y = a_1^T X = (2.84, 4.97, 5.68, 6.39, 7.81, 10.65)$

#

~~1.3 证明:~~

~~样本方差矩阵~~ $\cancel{S = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})^T}$ ~~x_j为第j个样本~~

$\cancel{= \frac{1}{n-1} (X - \bar{X})(X - \bar{X})^T}$ ~~$X \in R^{p\times n}$, p为各样本维数~~

1.3. 证明:

样本方差矩阵 $S = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})^T$ x_j为第j个样本

设总体期望, 即 $E(x_j) = \mu$

$$= \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \mu + \mu - \bar{x})(x_j - \mu + \mu - \bar{x})^T$$

$$= \frac{1}{n-1} \left[\sum_{j=1}^{n} (x_j - \mu)(x_j - \mu)^T + 2 \sum_{j=1}^{n} (x_j - \mu)(\mu - \bar{x})^T + \sum_{j=1}^{n} (\mu - \bar{x})(\mu - \bar{x})^T \right]$$

$$= \frac{1}{n-1} \left[\sum_{j=1}^{n} (x_j - \mu)(x_j - \mu)^T + \sum_{j=1}^{n} (\bar{x}\bar{x}^T - 2x_j\bar{x}^T + 2x_j\mu^T - \mu\mu^T) \right]$$

$$\therefore E(S) = \frac{1}{n-1} E\left[\sum_{j=1}^{n} (x_j - \mu)(x_j - \mu)^T + \sum_{j=1}^{n} (\bar{x}\bar{x}^T - 2x_j\bar{x}^T + 2x_j\mu^T - \mu\mu^T) \right]$$

$$= \frac{1}{n-1} \left\{ E\left(\sum_{j=1}^{n} (x_j - \mu)(x_j - \mu)^T \right) + E\left[n\bar{x}\bar{x}^T - 2\left(\sum_{j=1}^{n} x_j\bar{x}^T\right) + 2\left(\sum_{j=1}^{n} x_j\mu^T\right) - n\mu\mu^T \right] \right\}$$

$$= \frac{1}{n-1} \left\{ \sum_{j=1}^{n} E\left[(x_j - \mu)(x_j - \mu)^T \right] + E\left(n\bar{x}\bar{x}^T - 2n\bar{x}\bar{x}^T + 2n\bar{x}\mu^T - n\mu\mu^T \right) \right\}$$

由定义, $\Sigma = Var(X) = E(x_j - \mu)(x_j - \mu)^T$ 且 $E(\bar{x}) = \mu$ (x_j from 总体 X)

$$\therefore E(S) = \frac{1}{n-1} \left\{ n\Sigma - E\left(n\bar{x}\bar{x}^T - 2n\bar{x}\mu^T + n\mu\mu^T \right) \right\}$$

又: $E(n\bar{x}\bar{x}^T - 2n\bar{x}\mu^T + n\mu\mu^T) = n E[(\bar{x} - \mu)(\bar{x} - \mu)^T] = n \cdot \frac{1}{n}\Sigma = \Sigma$

(因为 $E[(\bar{x} - \mu)(\bar{x} - \mu)^T] = E[\frac{1}{n^2}(\sum_{i=1}^{n} x_i - \mu)(\sum_{i=1}^{n} x_i - \mu)^T] = \frac{1}{n^2} \cdot n\Sigma = \frac{1}{n}\Sigma$)

∴ 得 $E(S) = \frac{1}{n-1} \left\{ n\Sigma - n \cdot \frac{1}{n}\Sigma \right\} = \frac{n-1}{n-1}\Sigma = \Sigma$

即 ~~是~~ S是Σ(总体方差阵)的无偏估计

#

1.4 证明:

设 $A=\{\alpha_1,\cdots,\alpha_k\}\in R^{p\times k}$ $(k\le p)$ 是一组线性无关向量组, ~~且单位~~

且 $\alpha_i^T\alpha_j=\begin{cases}1, & if\ i=j\\ 0, & else\end{cases}$

设样本矩阵为 $X=(x_1,\cdots,x_n)^T\in R^{n\times p}$,由上课所讲,若X已标准化,则 $\hat{\Sigma}=S=\frac{1}{n-1}X^TX$

由最大可分性,主成分分析的投影阵(系数阵)由下得到。

$$(Q)\quad \max_A\ tr\{A^TX^TXA\}\qquad (A\in R^{p\times k})$$
$$s.t.\ A^TA=I_k$$

∴ 即证题目所给问题与(Q)优化问题等价.

设 ∃ 标准正交基 $W=(w_1,\cdots,w_p)\in R^{p\times p}$, 并设 $X=ZW^T$ (Z是X在W上投影), $Z\in R^{n\times p}$

∵ 要求 $rank(L)\le k$, W任取, 可令 L 落在 W 前k个向量张成的空间中, 即 $L\in span\{w_1,\cdots,w_k\}$

记 $W_k=(w_1,\cdots,w_k)\in R^{p\times k}$, 且记 $Z_k=(z_1,z_2,\cdots,z_k)^T$ 则 $L=Z_kW_k^T$ (*)

(该式成立是因为"最佳低秩矩阵"定理, 大二上数值算法讲过, 较为复杂, 此处略去)

⇒ L可取为保留X $\boxed{\text{奇异值前}}$k个后的阵, 即若 $X=U\Sigma V^T$, 则 $L=U\Sigma_kV^T$

Σ_k 为由 Σ 前k个元素构成的对角阵

∴ $\min_L\|X-L\|_F=\min_L\|X-L\|_F^2$

$=\min_L tr\{(X-L)^T(X-L)\}$

$\Leftrightarrow \min_L tr\{-2X^TL+L^TL\}=\min_L tr\{-2WZ^TZ_kW_k^T+W_kZ_k^TZ_kW_k^T\}$ (#)

∵ $tr(WZ^TZ_kW_k^T)=tr(Z^TZ_kW_k^TW)=tr\left(Z^TZ_k\cdot\begin{pmatrix}I_k\\ & 0\end{pmatrix}\right)=tr(Z_k^TZ_k)$

且 $tr(W_kZ_k^TZ_kW_k^T)=tr(Z_k^TZ_k)$ (W、W_k 均为正交阵)

∴ (#) $\Leftrightarrow \min_L tr(-2Z_k^TZ_k+Z_kZ_k)=\min_L -tr(Z_k^TZ_k)$

将 $Z_k=XW_k$ 代入, 即得:

$$(\#)\Leftrightarrow \min_{W_k} -tr(W_k^TX^TXW_k)$$
$$s.t.\ W_k^TW_k=I$$

这与(Q)等价

∴ 证毕

#