Homework 9

Name: 詹远瞩, Number: 17300180094

2020 年 12 月 22 日

1 1.1

$$
\begin{aligned}
\because Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} &= \begin{pmatrix} \alpha_1^T \\ \alpha_2^T \\ \vdots \\ \alpha_m^T \end{pmatrix} X = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1m} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mm} \end{pmatrix} X \triangleq AX \\
\therefore X = A^T Y &= \begin{pmatrix} \alpha_{11} & \alpha_{21} & \cdots & \alpha_{m1} \\ \alpha_{12} & \alpha_{22} & \cdots & \alpha_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1m} & \alpha_{2m} & \cdots & \alpha_{mm} \end{pmatrix} Y \\
\therefore X_i &= \sum_{k=1}^{m} \alpha_{ki} Y_k \\
Var(X_i) &= Var(\sum_{k=1}^{m} \alpha_{ki} Y_k) \\
&= \sum_{k=1}^{m} \alpha_{ki}^2 Var(Y_k) \\
&= \sum_{k=1}^{m} \alpha_{ki}^2 \lambda_k = \sigma_{ii} \\
\therefore \sum_{k=1}^{m} \rho^2(Y_k, X_i) &= \sum_{k=1}^{m} \frac{\lambda_k \alpha_{ki}^2}{\sigma_{ii}} \\
&= \frac{\sum\limits_{k=1}^{m} \lambda_k \alpha_{ki}^2}{\sigma_{ii}} \\
&= 1
\end{aligned}
$$

2 1.2

首先对样本阵 X 进行标准化处理:

$$\bar{x}_1 = \frac{\sum_{i=1}^{n} x_{1i}}{n} = 4 \qquad s_{11} = \frac{\sum_{i=1}^{n} (x_{1i} - \bar{x_1})^2}{n-1} = \frac{16}{5}$$

$$\bar{x}_2 = \frac{\sum_{i=1}^{n} x_{2i}}{n} = 5 \qquad s_{22} = \frac{\sum_{i=1}^{n} (x_{2i} - \bar{x_2})^2}{n-1} = 4$$

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

$$\therefore X^* = (x_{ij}^*) = \begin{pmatrix} -\frac{\sqrt{5}}{2} & -\frac{\sqrt{5}}{4} & -\frac{\sqrt{5}}{4} & 0 & \frac{\sqrt{5}}{4} & \frac{3\sqrt{5}}{4} \\ -\frac{3}{2} & -\frac{1}{2} & 0 & 0 & \frac{1}{2} & \frac{3}{2} \end{pmatrix}$$

为方便起见，下面的 X 代表已进行标准化的样本阵

$$S = \frac{XX^T}{n-1} = \begin{pmatrix} 1 & \frac{17\sqrt{5}}{40} \\ \frac{17\sqrt{5}}{40} & 1 \end{pmatrix}$$

对 S 进行特征值分解，得到:

$$\lambda_1 = \frac{40 + 17\sqrt{5}}{40} = 1.95 \qquad \alpha_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^T$$

$$\lambda_2 = \frac{40 - 17\sqrt{5}}{40} = 0.05 \qquad \alpha_2 = (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})^T$$

$$\therefore y_1 = \frac{1}{\sqrt{2}} x_1 + \frac{1}{\sqrt{2}} x_2 \qquad y_2 = \frac{1}{\sqrt{2}} x_1 - \frac{1}{\sqrt{2}} x_2$$

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = 0.975$$

样本的绝大部分信息可以由 y_1 来表示

3 1.3

只要证明对 S 的任一元素 S_{ij} 都有 $\mathbb{E}S_{ij} = \Sigma_{ij}$ 即可

$$
\begin{aligned}
S_{ij} &= \frac{1}{n-1}\sum_{k=1}^{n}(x_{ik}-\bar{x}_i)(x_{jk}-\bar{x}_j) \\
&= \frac{1}{n-1}\sum_{k=1}^{n}(x_{ik}-\mu_i+\mu_i-\bar{x}_i)(x_{jk}-\mu_j+\mu_j-\bar{x}_j) \\
&= \frac{1}{n-1}\sum_{k=1}^{n}\big((x_{ik}-\mu_i)(x_{jk}-\mu_j)+(x_{ik}-\mu_i)(\mu_j-\bar{x}_j)+(\mu_i-\bar{x}_i)(x_{jk}-\mu_j)+(\mu_i-\bar{x}_i)(\mu_j-\bar{x}_j)\big) \\
&= \frac{1}{n-1}\left(\sum_{k=1}^{n}(x_{ik}-\mu_i)(x_{jk}-\mu_j)+(\mu_j-\bar{x}_j)\sum_{k=1}^{n}(x_{ik}-\mu_i)+(\mu_i-\bar{x}_i)\sum_{k=1}^{n}(x_{jk}-\mu_j)+n(\mu_i-\bar{x}_i)(\mu_j-\bar{x}_j)\right)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\sum_{k=1}^{n}(x_{ik}-\mu_i)(x_{jk}-\mu_j) &= \sum_{k=1}^{n}\mathbb{E}(x_{ik}-\mu_i)(x_{jk}-\mu_j) \\
&= \sum_{k=1}^{n}Cov(X_i,X_j) = n\Sigma_{ij} \\
\mathbb{E}(\mu_j-\bar{x}_j)\sum_{k=1}^{n}(x_{ik}-\mu_i) &= -\frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\mathbb{E}(x_{jl}-\mu_j)(x_{ik}-\mu_i) \\
&= -\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(x_{jk}-\mu_j)(x_{ik}-\mu_i) \\
&= -\Sigma_{ij} \\
\mathbb{E}(\mu_i-\bar{x}_i)\sum_{k=1}^{n}(x_{jk}-\mu_j) &= -\frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\mathbb{E}(x_{il}-\mu_i)(x_{jk}-\mu_j) \\
&= -\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(x_{ik}-\mu_i)(x_{jk}-\mu_j) \\
&= -\Sigma_{ij} \\
\mathbb{E}n(\mu_i-\bar{x}_i)(\mu_j-\bar{x}_j) &= \frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\mathbb{E}(x_{il}-\mu_i)(x_{jk}-\mu_j) \\
&= \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(x_{ik}-\mu_i)(x_{jk}-\mu_j) \\
&= \Sigma_{ij}
\end{aligned}
$$

$$
\begin{aligned}
\therefore \mathbb{E}S_{ij} &= \frac{1}{n-1}\big(n\Sigma_{ij} - \Sigma_{ij} - \Sigma_{ij} + \Sigma_{ij}\big) \\
&= \Sigma_{ij}
\end{aligned}
$$

因此 S 是 Σ 的无偏估计

4 1.4

设 $A \in \mathbb{R}^{m\times n}, m \geq n$，$A$ 的奇异值分解为 $A = U\Sigma V^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T$，令 $A_r = \sum_{i=1}^{r} \sigma_i u_i v_i^T$，我们首先证明：$\underset{rank(L)\leq r}{\arg\min} \|A - L\|_2 = A_r$，$\underset{rank(L)\leq r}{\arg\min} \|A - L\|_F = A_r$:

由于 $r(L) \leq r$，所以 $\exists X, Y \in \mathbb{R}^{n\times r}$ 满足 $L = XY^T$，由于 Y 的秩最多为 r，所以 $\exists w = \sum_{i=1}^{r+1} a_i v_i$ 满足 $Y^T w = 0, w^T w = 1$

$$
\begin{aligned}
\|A - L\|_2^2 &= \max_{\|x\|_2=1} \|(A-L)x\|_2^2 \\
&\geq \|(A-L)w\|_2^2 \\
&= \|Aw - XY^T w\|_2^2 \\
&= \|Aw\|_2^2 \\
\because w^T w = 1 &\quad \therefore \sum_{i=1}^{r+1} a_i^2 = 1 \\
\because Aw &= (\sum_{i=1}^{n} \sigma_i u_i v_i^T)(\sum_{j=1}^{r+1} a_j v_j) \\
&= \sum_{i=1}^{r+1} a_i \sigma_i u_i \\
\therefore \|Aw\|_2^2 &= \sum_{i=1}^{r+1} a_i^2 \sigma_i^2 \\
&\geq \sigma_{r+1}^2 \\
\because \|A - A_r\|^2 &= \sigma_{r+1}^2 \\
&\therefore \underset{rank(L)\leq r}{\arg\min} \|A - L\|_2 = A_r
\end{aligned}
$$

设有矩阵 C, D, C_k, D_k，其中 $C = \sum\limits_{i=1}^{n} \sigma_i u_i v_i^T, C_k = \sum\limits_{i=1}^{k} \sigma_i u_i v_i^T$，$C_k$ 是 C 的截断奇异值分解，类似定义 D 和 D_k，则 $\sigma_i(C) = \sigma_1(C - C_{i-1})$，其中 $\sigma_i(C)$ 代表 C 的第 i 大的奇异值

$$\begin{aligned}
\sigma_i(C) + \sigma_j(D) &= \sigma_1(C - C_{i-1}) + \sigma_1(D - D_{j-1}) \\
&= \|C - C_{i-1}\|_2 + \|D - D_{j-1}\|_2 \\
&\geq \sigma_1((C + D) - (C_{i-1} + D_{j-1}))
\end{aligned}$$

记 $A = C + D$

$$\begin{aligned}
\sigma_1((C + D) - (C_{i-1} + D_{j-1})) &= \sigma_1(A - (C_{i-1} + D_{j-1})) \\
\because rank(C_{i-1} + D_{j-1}) &\leq rank(C_{i-1}) + rank(D_{j-1}) \leq i + j - 2 \\
\therefore \sigma_1(A - (C_{i-1} + D_{j-1})) &\geq \sigma_1(A - A_{i+j-2})
\end{aligned}$$

令 $C = A - L, D = L, i \geq 1, j = r + 1$

$$\begin{aligned}
\sigma_i(A - L) + \sigma_{r+1}(L) &\geq \sigma_1(A - A_{i+r-1}) \\
\because rank(L) \leq r \quad &\therefore \sigma_{r+1}(L) = 0 \\
\therefore \sigma_i(A - L) &\geq \sigma_1(A - A_{i+r-1}) \\
\therefore \sigma_i(A - L) &\geq \sigma_{r+i}(A) \\
\therefore \|A - L\|_F^2 &= \sum_{i=1}^{n} \sigma_i^2(A - L) \\
&\geq \sum_{i=1}^{n-r} \sigma_{r+i}^2(A - L) \\
&= \sum_{i=r+1}^{n} \sigma_i^2(A) \\
&= \|A - A_r\|_F^2 \\
\therefore \mathop{\arg\min}_{rank(L) \leq r} \|A - L\|_F &= A_r
\end{aligned}$$

设 X 的奇异值分解为 $X = U\Sigma V^T$，X 的奇异值由大到小为：$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0$，则使该最优化问题得到最优解的 $L = \sum\limits_{i=1}^{k} \sigma_i u_i v_i^T$，记 $U_k = (u_1, u_2, \cdots, u_k)$ 是 U 的前 k 列构成的截断矩阵，$V_k = (v_1, v_2, \cdots, v_k)$ 是 V 的前 k 列构成的截断矩阵,$\Sigma_k = diag\{\sigma_1, \sigma_2, \cdots, \sigma_k\}$，则 $L = U_k \Sigma_k V_k^T$，而根据教材中的奇异值分解求主成分的算法，最终的主成分为 $Y = V_k^T X^T = \Sigma_k U_k^T$，$L$ 与 Y 只相差一个正交变换，因此两者包含的信息是相同的，所以主成分等价于求此最优化问题。