统计(机器)学习 Homework-9

1. 设 $X=(X_1,\cdots,X_m)^\top$ 是 m 维随机变量，协方差矩阵为 $\mathrm{cov}(X)\overset{\mathrm{def}}{=}\boldsymbol{\Sigma}$. 设矩阵 $\boldsymbol{\Sigma}$ 的特征值为 $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_m$；这 m 个特征值对应的单位特征向量为 $\alpha_1,\cdots,\alpha_m$ (特征向量矩阵为 $(\alpha_1,\alpha_2,\cdots,\alpha_m)\in\mathbb{R}^{m\times m}$，其中 $\alpha_k=(\alpha_{k1},\alpha_{k2},\cdots,\alpha_{km})^\top$ 为列向量)，若用 Y_k 表示 X 的第 k 个主成分，且 $Y=(Y_1,\cdots,Y_m)^\top$，证明：

(1) $\mathrm{cov}(Y)=\mathrm{diag}(\lambda_1,\cdots,\lambda_m)$;

证：记特征向量矩阵为 $A=(\alpha_1,\cdots,\alpha_m)$，且满足：

$$\Sigma = A\cdot\Lambda\cdot A^{-1} = A\cdot\Lambda\cdot A^T \Leftrightarrow A^T\cdot\Sigma\cdot A = \Lambda$$

$A\cdot\Lambda\cdot A^T$ 为 Σ 的谱分解，$\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_m)$

其中 $\lambda_1,\cdots,\lambda_m$ 为 Σ 的特征值

注意到 $Y=A^T\cdot X$

$$\begin{aligned}\Rightarrow \mathrm{cov}(Y) &= A^T\cdot\mathrm{cov}(X)A\\ &= A^T\cdot\Sigma\cdot A\\ &= \Lambda\\ &= \mathrm{diag}(\lambda_1,\cdots,\lambda_m)\end{aligned}$$

(2) $\sum_i\lambda_i=\sum_i\boldsymbol{\Sigma}_{ii}\,(\boldsymbol{\Sigma}_{ii}=\mathrm{var}(X_i))$.

证：注意到 $\sum_i\Sigma_{ii} = \mathrm{trace}(\Sigma)$

trace 具有可交换性：$\mathrm{trace}(MN)=\mathrm{trace}(NM)$

则有：

$$\begin{aligned}\mathrm{trace}(\Sigma) &= \mathrm{trace}(A\Lambda A^T)\\ &= \mathrm{trace}(A^TA\Lambda)\\ &= \mathrm{trace}(A^{-1}\cdot A\cdot\Lambda)\end{aligned}$$

$$= \mathrm{trace}(\Lambda)$$

$$= \sum_{i=1}^{m} \lambda_i$$

(3) 因子负荷量为

$$\rho(Y_k, X_i) = \operatorname{cor}(Y_k, X_i) = \frac{\sqrt{\lambda_k}\alpha_{ki}}{\sqrt{\boldsymbol{\Sigma}_{ii}}}$$

证: $Y_k = \alpha_k^T \cdot X = \sum_{j=1}^{n} \alpha_{kj} X_j$

$$Cov(Y_k, X_i) = Cov\left(\sum_{j=1}^{n} \alpha_{kj} X_j, X_i\right)$$

$$= \sum_{j=1}^{n} \alpha_{kj} \cdot Cov(X_j, X_i)$$

记 Σ 的第 i 列为 Σ_i，则有 $\Sigma_i = \left(Cov(X_1, X_i), \cdots, Cov(X_m, X_i)\right)^T$

$$\Rightarrow Cov(Y_k, X_i) = \alpha_k^T \cdot \Sigma_i$$

$$= (\alpha_k^T \cdot \Sigma)_i$$

$$= \lambda_k \cdot \alpha_{ki}$$

另一方面:

$$Var(Y_k) = Var(\alpha_k^T X)$$

$$= \alpha_k^T \cdot Cov(X) \cdot \alpha_k$$

$$= \alpha_k^T \cdot \Sigma \cdot \alpha_k$$

$$= \lambda_k$$

则有:

$$\rho(Y_k, X_i) = \operatorname{Cor}(Y_k, X_i)$$

$$= \frac{\operatorname{Cov}(Y_k, X_i)}{\sqrt{\operatorname{Var}(Y_k)}\sqrt{\operatorname{Var}(X_i)}}$$

$$= \frac{\lambda_k \cdot \alpha_{ki}}{\sqrt{\lambda_k}\sqrt{\Sigma_{ii}}}$$

$$= \frac{\sqrt{\lambda_k} \cdot \alpha_{ki}}{\sqrt{\Sigma_{ii}}}$$

(4) $\sum_{i=1}^{m} \boldsymbol{\Sigma}_{ii}\rho\left(Y_k, X_i\right)^2 = \lambda_k$

证：由上题可知 $\rho(Y_k, X_i)^2 = \dfrac{\lambda_k \alpha_{ki}^2}{\Sigma_{ii}}$

则有：

$$\sum_{i=1}^{m} \Sigma_{ii}\, \rho(Y_k, X_i)^2$$

$$= \sum_{i=1}^{m} \Sigma_{ii} \cdot \frac{\lambda_k \alpha_{ki}^2}{\Sigma_{ii}}$$

$$= \lambda_k \cdot \sum_{i=1}^{m} \alpha_{ki}^2$$

由于α_k是单位特征向量，$\sum_{i=1}^{m} \alpha_{ki}^2 = \|\alpha_k\|^2 = 1$

则有 $\sum_{i=1}^{m} \Sigma_{ii}\, \rho(Y_k, X_i)^2 = \lambda_k$

(5) $\sum_k \rho^2(Y_k, X_i) = 1$

证：注意到对于Σ有如下谱分解：

$$\Sigma = A \cdot \Lambda \cdot A^{-1} = A \cdot \Lambda \cdot A^T$$

其中 $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_m)$，$A = (\alpha_1, \cdots, \alpha_m)$

注意到等式右侧：

$$(A \cdot \Lambda \cdot A^T)_{ij} = \sum_{k=1}^{m} \alpha_{ki} \cdot \lambda_k \cdot \alpha_{kj}$$

等式两端的 (i,i) 处的元素满足：

$$\Sigma_{ii} = \sum_{k=1}^{m} \alpha_{ki} \cdot \lambda_k \cdot \alpha_{ki} = \sum_{k=1}^{m} \lambda_k \alpha_{ki}^2$$

$$\Rightarrow \sum_{k=1}^{m} \frac{\lambda_k \cdot \alpha_{ki}^2}{\Sigma_{ii}} = 1$$

$$\Rightarrow \sum_{k=1}^{m} \rho(Y_k, X_i) = 1 \qquad \square$$

2. 给定标准化样本矩阵 $\mathbf{X} \in \mathbb{R}^{N \times p}$，$\mathbf{X}$ 的 SVD 分解为 $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$，其中 $\mathbf{U} \in \mathbb{R}^{N \times p}$ 是左特征向量矩阵，$\mathbf{\Sigma} = \mathrm{diag}\{\sigma_1, \sigma_2, \cdots, \sigma_p\} \in \mathbb{R}^{p \times p}$，其中 $\sigma_1 \geq \cdots \geq \sigma_p$，$\mathbf{V} \in \mathbb{R}^{p \times p}$ 是右特征向量矩阵。对 $\mathbf{X}$ 进行主成分分析，假设保留前 k 个主成分，$\mathrm{rank}(\mathbf{X}) > k$。试证明：

$$\boxed{X} = \boxed{U}\,\boxed{\Sigma}\,\boxed{V}$$

(1) 前 k 个样本主成分可以表示为

$$\mathbf{Z}^{(k)} = \mathbf{U}^{(k)}\mathbf{\Sigma}^{(k)} \in \mathbb{R}^{N \times k},$$

其中，$\mathbf{U}^{(k)}$ 是 $\mathbf{U}$ 的前 k 列，$\mathbf{\Sigma}^{(k)}$ 是 $\mathbf{\Sigma}$ 的前 $k \times k$ 矩阵；

证：对于标准化样本矩阵，每一维度上的样本方差均满足 $s_{ii} = 1$

也即其样本协方差矩阵 S 的对角线为 1，且样本相关系数满足：

$$R = \mathrm{diag}(S)^{-1/2} \cdot S \cdot \mathrm{diag}(S)^{-1/2}$$
$$= S$$
$$= \frac{1}{n-1} \cdot X^T \cdot X$$
$$= \frac{1}{n-1} V \Sigma U^T U \Sigma V^T$$
$$= \frac{1}{n-1} \cdot V \cdot \Sigma^2 \cdot V^T$$

设 R 的特征值为 $\lambda_1, \cdots, \lambda_p$，满足 $\lambda_k = \frac{\sigma_k^2}{n-1}$

对应特征向量为 V_k，为 V 的第 k 列

根据样本主成分的定义，前 k 个样本主成分可表示为：

$$Z^{(k)} = X \cdot V^{(k)}$$

其中 $V^{(k)} = (V_1, \cdots, V_k)$

也即 $Z^{(k)} = U \Sigma V^T \cdot V^{(k)}$

$$= U \cdot \Sigma \begin{bmatrix} I_k \\ O_{p-k} \end{bmatrix}$$

$$= U \cdot \begin{bmatrix} \Sigma^{(k)} \\ O_{p-k} \end{bmatrix}$$

$$= [U^{(k)}, U^{(p-k)}] \cdot \begin{bmatrix} \Sigma^{(k)} \\ O_{p-k} \end{bmatrix}$$

$$= U^{(k)} \cdot \Sigma^{(k)}$$

(2) 试证明下面问题的解为 $\mathbf{B} = \mathbf{V}^{(k)\top}$（其中 $\mathbf{V}^{(k)}$ 是 $\mathbf{V}$ 的前 k 列）：

$$\mathbf{B} = \mathrm{argmin}_{\mathbf{B}} \left\| \mathbf{X} - \mathbf{Z}^{(k)}\mathbf{B} \right\|_F^2.$$

证：$\min\limits_{B} \|X - Z^{(k)}B\|_F^2 \Leftrightarrow \sum\limits_{i=1}^{p} \min\limits_{b_i} \|x_i - Z^{(k)} b_i\|_2^2$

其中 X_i 是 X 的第 i 列，b_i 是 B 的第 i 列

则由最小二乘法：$b_i = (Z^{(k)T} Z^{(k)})^{-1} Z^{(k)T} X_i$

$$\Rightarrow \quad B = (Z^{(k)T} Z^{(k)})^{-1} Z^{(k)T} X$$

注意到 $X = U \Sigma V^T$，$Z^{(k)} = U^{(k)} \Sigma^{(k)}$

则有：

$$B = (\Sigma^{(k)2})^{-1} \cdot \Sigma^{(k)} U^{(k)T} U \Sigma V^T$$

$$= (\Sigma^{(k)})^{-1} \cdot [I_k, O_{p-k}] \cdot \Sigma V^T$$

$$= (I_k, O_{p-k}) \cdot V^T$$

$$= V^{(k)T}$$

(3) 证明以下最优化问题的解为 $\mathbf{X}^{(k)} = \mathbf{U}^{(k)} \mathbf{\Sigma}^{(k)} \mathbf{V}^{(k)\top}$:

$$\min_{\mathbf{L}} \|\mathbf{X} - \mathbf{L}\|_F, \quad \text{s.t.} \quad \text{rank}(\mathbf{L}) \le k.$$

(提示：对于任意矩阵 $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n} (n \le m)$，有 $\sigma_{i+j-1}(\mathbf{A} + \mathbf{B}) \le \sigma_i(\mathbf{A}) + \sigma_j(\mathbf{B})$，其中，$1 \le i, j \le n, i + j - 1 \le n$)

证：由于 $\text{rank}(L) \le k$，则当 $j > k$ 时，有 $\sigma_j(L) = 0$

$$\sigma_{i+j-1}(X) \le \sigma_i(X - L) + \sigma_j(L)$$

$$\Rightarrow \quad \sigma_{i+j}(X) \le \sigma_i(X - L) + \sigma_{j+1}(L)$$

$j \geqslant k$时： $\sigma_{i+j}(X) \leq \sigma_i(X-L)$

$$\|X-L\|_F^2 = \sum_{i=1}^{p} \sigma_i^2(X-L)$$

$$\geqslant \sum_{i=1}^{p-k} \sigma_i^2(X-L)$$

$$\geqslant \sum_{i=1}^{p-k} \sigma_{i+k}^2(X)$$

$$= \sum_{i=k+1}^{p} \sigma_i^2(X)$$

也即 $\|X-L\|_F \geqslant \left(\sum_{i=k+1}^{p} \sigma_i^2(X)\right)^{1/2}$

注意到 $\|\cdot\|_F$ 有正交不变性

设 $\hat{U} = (U, U_\perp)$，其中 $U_\perp$ 是 U 的正交补空间的一组正交基

并且有 $\hat{U}\hat{U}^T = I$， $\hat{U}^T \cdot U = \begin{bmatrix} I_p \\ 0 \end{bmatrix}$， $\hat{U}^T U^{(k)} = \begin{bmatrix} I_k \\ 0 \end{bmatrix}$

则当 $L = X^{(k)} = U^{(k)} \cdot \Sigma^{(k)} \cdot V^{(k)T}$ 时：

$$\|X-L\|_F = \|U\Sigma V^T - U^{(k)}\Sigma^{(k)}V^{(k)T}\|_F$$

$$= \|U\Sigma V^T - U\begin{pmatrix} \Sigma^{(k)} & \\ & 0_{p-k} \end{pmatrix} V^T\|_F$$

$$= \| \hat{U}^T U \Sigma V^T V - \hat{U}^T U \begin{pmatrix} \Sigma^{(k)} & \\ & O_{p-k} \end{pmatrix} V^T V \|_F$$

$$= \| \begin{bmatrix} I_p \\ O_{N-p} \end{bmatrix} \cdot \left(\Sigma - \begin{pmatrix} \Sigma^{(k)} & \\ & O_{p-k} \end{pmatrix} \right) \|_F$$

$$= \left(\sum_{i=k+1}^{p} \sigma_i^2(X) \right)^{1/2}$$

也即 $L = U^{(k)} \Sigma^{(k)} V^{(k)T}$ 时，$\| X - L \|_F$ 恰好取到下界

则有 $\min_L \| X - L \|_F = \left(\sum_{i=k+1}^{p} \sigma_i^2(X) \right)^{1/2}$

且 $L = U^{(k)} \Sigma^{(k)} V^{(k)T}$ 是最优化问题的一个解 $\square$