

HW7 参考答案

奇异值分解 (SVD, singular value decomposition)

Exercise 1. 对任何矩阵 $A = (a_{ij})$ (即 A 的 (i, j) 元为 a_{ij}) , 证明 $\sum a_{ij}^2 = \sum \sigma_i^2$, 其中 $\sigma_1, \sigma_2, \dots$ 为 A 的所有奇异值。

Proof. 对 A 做奇异值分解 $A = UDV^\top$, $D = \text{diag}(\sigma_1, \sigma_2, \dots)$ 。注意 $\sum_{i,j} a_{ij}^2 = \text{tr}(A^\top A)$, 故

$$\begin{aligned}\sum_{i,j} a_{ij}^2 &= \text{tr}(A^\top A) \\ &= \text{tr}(VDU^\top UDV^\top) \\ &= \text{tr}(VD^2V^\top) \\ &= \text{tr}(D^2V^\top V) \\ &= \text{tr}(D^2) = \sum_i \sigma_i^2.\end{aligned}$$

Exercise 2. 说明正定矩阵 A 的奇异值分解与谱分解是一致的 (因此 Eckart-Young-Mirsky 奇异值最佳逼近定理也适用于主成分, 参见 lecture 13, P18)。

Proof. 记 $A_{n \times n}$ 的奇异值分解为 $A = P\Sigma Q^\top$, 则 $A^\top A = Q\Sigma^2 Q^\top$, 这表明 Σ^2 的对角元是 $A^\top A$ 的特征值。另一方面, A 正定故对称, 其谱分解为 $A = U\Lambda U^\top$, 于是 $A^\top A = A^2 = U\Lambda^2 U^\top$, 即 Λ^2 的对角元也是 $A^\top A$ 的特征值。从而 $\sigma_i^2 = \lambda_i^2$; 又 $\sigma_i > 0$, 且 A 正定使 $\lambda_i > 0$, 故 $\sigma_i = \lambda_i, \forall i = 1, \dots, n$ 。因此 A 的奇异值分解与谱分解一致: 奇异值即特征值, 奇异向量即特征向量。

Exercise 3. 假设 $n \times 2$ 数据矩阵 $X = (x_1, x_2)$, $\|x_1\| = a, \|x_2\| = b, a > b, x_1^\top x_2 = 0$, 求 X 的奇异值分解。

Proof. 由 $x_1^\top x_2 = 0$ 知

$$X^\top X = \begin{pmatrix} a^2 & 0 \\ 0 & b^2 \end{pmatrix}$$

已是对角阵, 故取 $V = E_2, D = \text{diag}(a, b)$, 则 $U = XVD^{-1} = \begin{pmatrix} \frac{x_1}{a} & \frac{x_2}{b} \end{pmatrix}$ 。于是 X 的一个奇异值分解为

$$X = \begin{pmatrix} \frac{x_1}{a} & \frac{x_2}{b} \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} E_2.$$

Exercise 4. 假设 A 是 $n \times m$ 矩阵, $u \in \mathbb{R}^n$, $v \in \mathbb{R}^m$, 且 $Av = \alpha u$, $A^\top u = \beta v$, 其中实数 $\alpha, \beta \neq 0$. 证明 α, β 同号且 $\sqrt{\alpha\beta}$ 是 A 的一个奇异值。

Proof. 由 $A^\top Av = A^\top(\alpha u) = \alpha(A^\top u) = \alpha\beta v$, 从而 v 是 $A^\top A$ 的一个特征向量, 对应特征值 $\alpha\beta$. 又 $A^\top A$ 半正定, 故 $\alpha\beta \geq 0$; 因 $\alpha, \beta \neq 0$, 所以 $\alpha\beta > 0$, 即 α, β 同号。既然 $\alpha\beta$ 是 $A^\top A$ 的特征值, 按奇异值的定义, $\sqrt{\alpha\beta}$ 即为 A 的一个奇异值。 ■

Exercise 5. (普氏分析) 假设 X, Y 都是 $n \times p$ 矩阵, $p \times p$ 方阵 $X^\top Y$ 的奇异值分解为 $X^\top Y = UDV^\top$, 其中 U, V 都是 p 阶正交矩阵。证明 $Q = UV^\top$ 使得 $\|X - YQ^\top\|_F^2$ 达到最小。

Proof. 由于 Q 正交,

$$\begin{aligned} \|X - YQ^\top\|_F^2 &= \text{tr}(X^\top X) + \text{tr}(QY^\top YQ^\top) - 2 \text{tr}(X^\top YQ^\top) \\ &= \text{tr}(X^\top X) + \text{tr}(Y^\top Y) - 2 \text{tr}(X^\top YQ^\top), \end{aligned}$$

前两项与 Q 无关, 故问题转化为最大化 $\text{tr}(X^\top YQ^\top)$ 。由奇异值分解,

$$\text{tr}(X^\top YQ^\top) = \text{tr}(UDV^\top Q^\top) = \text{tr}(DV^\top Q^\top U) = \sum_i \sigma_i (V^\top Q^\top U)_{ii}.$$

记 u_i, v_i 分别为 U, V 的第 i 列, 则 $(V^\top Q^\top U)_{ii} = v_i^\top Q^\top u_i$, 由 Cauchy-Schwarz 及 Q 正交, $(V^\top Q^\top U)_{ii} = (Qv_i)^\top u_i \leq \|Qv_i\| \|u_i\| = 1$ 。因此 $\text{tr}(X^\top YQ^\top) \leq \sum_i \sigma_i$, 当 $V^\top Q^\top U = E$ 即 $Q = UV^\top$ 时各对角元同时取到 1, 达到最大。 ■

典则相关分析 (CCA, canonical correlation analysis)

Exercise 6. 假设 y 是随机变量, x 是 $p \times 1$ 随机向量, 其协方差矩阵

$$\Sigma = \text{Cov} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} > 0,$$

其中 $\Sigma_{yy} = \sigma_y^2$ 是正实数, Σ_{xy} 是 $p \times 1$ 向量。

(a) 求 x, y 的第一典则相关系数 $\sqrt{\lambda_1}$ (也是唯一的非零典则相关系数) 和第一对典则变量 (ξ_1, η_1) 。

(b) 假设 y, x 均值都为 0, 满足线性模型 $y = \beta^\top x + \epsilon$, $\epsilon \sim (0, \sigma^2)$, $\epsilon \perp x$, 证明该模型等价于 $\eta_1 = \sqrt{\lambda_1} \xi_1 + \delta_1$, $\delta_1 \sim (0, 1 - \lambda_1)$, $\delta_1 \perp \xi_1$ 。

Proof. (a) 按定义需寻找 a, b 使 $\text{Corr}(a^\top x, by)$ 最大, 即最大化

$$\frac{a^\top \Sigma_{xy} b}{\sqrt{a^\top \Sigma_{xx} a} \sqrt{b^\top \Sigma_{yy} b}}.$$

y 为标量时该比值与 b 无关, 只需最大化 $\frac{a^\top \Sigma_{xy}}{\sigma_y \sqrt{a^\top \Sigma_{xx} a}}$ 。令 $u = \Sigma_{xx}^{-1/2} a$, 由 Cauchy-Schwarz,

$$\frac{a^\top \Sigma_{xy}}{\sqrt{a^\top \Sigma_{xx} a}} = \frac{u^\top \Sigma_{xx}^{-1/2} \Sigma_{xy}}{\|u\|} \leq \|\Sigma_{xx}^{-1/2} \Sigma_{xy}\| = \sqrt{\Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy}},$$

取等于 $u \propto \Sigma_{xx}^{-1/2} \Sigma_{xy}$ 即 $a \propto \Sigma_{xx}^{-1} \Sigma_{xy}$ 时。于是第一典则相关系数

$$\sqrt{\lambda_1} = \frac{\sqrt{\Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy}}}{\sigma_y},$$

第一对典则变量（标准化为单位方差）为

$$\xi_1 = \frac{\Sigma_{yx} \Sigma_{xx}^{-1} x}{\sqrt{\Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy}}}, \quad \eta_1 = \frac{y}{\sigma_y}.$$

(b) 模型 $y = \beta^\top x + \epsilon$ 中 $\epsilon \perp x$ ，两边与 x 求协方差得 $\Sigma_{xy} = \Sigma_{xx} \beta$ ，故 $\beta = \Sigma_{xx}^{-1} \Sigma_{xy}$ ，从而

$$\beta^\top x = \Sigma_{yx} \Sigma_{xx}^{-1} x = \frac{\Sigma_{yx} \Sigma_{xx}^{-1} x}{\sqrt{\Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy}}} \cdot \frac{\sqrt{\Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy}}}{\sigma_y} \cdot \sigma_y = \sqrt{\lambda_1} \sigma_y \xi_1.$$

又 $\eta_1 = y/\sigma_y$ ，故 $y = \sigma_y \eta_1$ 。代入并除以 σ_y ：

$$\eta_1 = \frac{y}{\sigma_y} = \sqrt{\lambda_1} \xi_1 + \frac{\epsilon}{\sigma_y} = \sqrt{\lambda_1} \xi_1 + \delta_1, \quad \delta_1 := \frac{\epsilon}{\sigma_y}.$$

由 $\epsilon \perp x$ 且 ξ_1 是 x 的线性函数，得 $\delta_1 \perp \xi_1$ 。再由方差分解 $\sigma_y^2 = \Sigma_{yy} = \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} + \sigma^2 = \lambda_1 \sigma_y^2 + \sigma^2$ ，得 $\sigma^2 = (1 - \lambda_1) \sigma_y^2$ ，故

$$\text{Var}(\delta_1) = \frac{\sigma^2}{\sigma_y^2} = 1 - \lambda_1,$$

即 $\delta_1 \sim (0, 1 - \lambda_1)$ 。两模型等价。 ■

Exercise 7. 假设随机向量 $\begin{pmatrix} y_{q \times 1} \\ x_{p \times 1} \end{pmatrix}$ 的协方差矩阵为相关系数矩阵 $R = (\rho_{ij})$ ， $\rho_{ii} = 1$ 。证明 x 和 y 的第一典则相关系数 $\sqrt{\lambda_1} \geq \rho_{ij}$ ， $i = 1, \dots, q$ ； $j = q + 1, \dots, q + p$ 。

Proof. 按第一典则相关系数的定义，

$$\sqrt{\lambda_1} = \max_{a^\top \Sigma_{xx} a = b^\top \Sigma_{yy} b = 1} \text{Cov}(a^\top x, b^\top y) = \max_{a^\top \Sigma_{xx} a = b^\top \Sigma_{yy} b = 1} a^\top \Sigma_{xy} b.$$

由于 $\Sigma_{xx} = E_p$ 、 $\Sigma_{yy} = E_q$ （相关系数矩阵对角为 1），约束化为 $\|a\| = \|b\| = 1$ 。取 a 的第 i 个分量为 1、其余为 0， b 的第 $(j - q)$ 个分量为 1、其余为 0，则 $a^\top \Sigma_{xy} b = \rho_{ij}$ 。由于 $\sqrt{\lambda_1}$ 是对所有满足约束的 (a, b) 取的最大值，故 $\sqrt{\lambda_1} \geq \rho_{ij}$ 。 ■

Exercise 8. 假设随机向量 $\begin{pmatrix} y_{q \times 1} \\ x_{p \times 1} \end{pmatrix}$ 的协方差矩阵为 $\Sigma = (1 - \rho)I_m + \rho \mathbf{1}_m \mathbf{1}_m^\top$ ， $m = p + q$ ， $\rho > -1/(m - 1)$ 。求 x, y 的第一典则相关系数和典则变量。

Proof. 按定义求解

$$\max_{a^\top \Sigma_{xx} a = b^\top \Sigma_{yy} b = 1} a^\top \Sigma_{xy} b.$$

由分块结构 $\Sigma_{xx} = (1-\rho)I_p + \rho\mathbf{1}_p\mathbf{1}_p^\top$ 、 $\Sigma_{yy} = (1-\rho)I_q + \rho\mathbf{1}_q\mathbf{1}_q^\top$ 、 $\Sigma_{xy} = \rho\mathbf{1}_p\mathbf{1}_q^\top$ ，优化问题写为

$$\max \rho(\mathbf{1}_p^\top a)(\mathbf{1}_q^\top b) \quad \text{s.t.} \quad \begin{cases} (1-\rho)a^\top a + \rho(\mathbf{1}_p^\top a)^2 = 1, \\ (1-\rho)b^\top b + \rho(\mathbf{1}_q^\top b)^2 = 1. \end{cases}$$

拉格朗日函数

$$L = \rho(\mathbf{1}_p^\top a)(\mathbf{1}_q^\top b) + \lambda[(1-\rho)a^\top a + \rho(\mathbf{1}_p^\top a)^2 - 1] + \mu[(1-\rho)b^\top b + \rho(\mathbf{1}_q^\top b)^2 - 1].$$

一阶条件

$$\frac{\partial L}{\partial a} = \rho(\mathbf{1}_q^\top b)\mathbf{1}_p + \lambda[2(1-\rho)a + 2\rho(\mathbf{1}_p^\top a)\mathbf{1}_p] = 0, \quad (1)$$

$$\frac{\partial L}{\partial b} = \rho(\mathbf{1}_p^\top a)\mathbf{1}_q + \mu[2(1-\rho)b + 2\rho(\mathbf{1}_q^\top b)\mathbf{1}_q] = 0. \quad (2)$$

对 (1) 两边左乘 $\mathbf{1}_p^\top$ （即各分量求和）得

$$p\rho(\mathbf{1}_q^\top b) + \lambda[2(1-\rho)(\mathbf{1}_p^\top a) + 2p\rho(\mathbf{1}_p^\top a)] = 0,$$

$$\mathbf{1}_p^\top a = \frac{-p\rho(\mathbf{1}_q^\top b)}{2\lambda(1-\rho + p\rho)}. \quad (3)$$

将 (3) 回代 (1)，可见 a 被表示成 $\mathbf{1}_p$ 的标量倍，即 a 各分量相等；同理 b 各分量相等。记 $a = a_0\mathbf{1}_p$ 、 $b = b_0\mathbf{1}_q$ ，代入约束

$$(1-\rho)pa_0^2 + \rho p^2 a_0^2 = 1, \quad (1-\rho)qb_0^2 + \rho q^2 b_0^2 = 1,$$

解得

$$a_0 = \pm \frac{1}{\sqrt{p(1-\rho + p\rho)}}, \quad b_0 = \pm \frac{1}{\sqrt{q(1-\rho + \rho q)}}.$$

目标函数 $\rho(\mathbf{1}_p^\top a)(\mathbf{1}_q^\top b) = \rho pq a_0 b_0$ ，其符号由 $\rho a_0 b_0$ 决定。为使目标最大（典则相关系数取非负），应令 $\text{sgn}(a_0 b_0) = \text{sgn}(\rho)$ ，此时目标取到

$$\sqrt{\lambda_1} = \frac{|\rho|\sqrt{pq}}{\sqrt{(1-\rho + p\rho)(1-\rho + \rho q)}}.$$

对应的第一对典则变量为

$$\xi_1 = a^\top x = \frac{\text{sgn}(\rho)}{\sqrt{p(1-\rho + p\rho)}} \sum_{i=1}^p x_i, \quad \eta_1 = b^\top y = \frac{1}{\sqrt{q(1-\rho + \rho q)}} \sum_{j=1}^q y_j$$

（符号取在 ξ_1 上，使二者正相关； $\rho > 0$ 时即去掉 sgn ， $\rho < 0$ 时 ξ_1 反号， $\rho = 0$ 时 $\sqrt{\lambda_1} = 0$ ）。 ■