多元统计分析—R 与 Python 的实现

矩阵代数回顾

吴喜之

June 26, 2019

矩阵定义 矩阵是数值或数值变量的一个矩形数字阵, m 行 n 列 (亦称 $m \times n$ 维或 $m \times n$ 阶) 的矩阵 $\boldsymbol{A}$ 的形式为:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

一些矩阵形式:

- 对称矩阵
- 上三角矩阵, 下三角矩阵, 单位矩阵, 矩阵的转置

矩阵基本运算和概念:

- 加、减、乘
- 行列式计算
- 矩阵的逆及广义逆
- 矩阵 Kronecker 乘积
- 幂等矩阵
- 向量空间
- 矩阵的秩
- 矩阵的迹
- 矩阵的特征值及广义特征值
- 分块矩阵

矩阵的分解: 特征值分解

特征值分解 (eigenvalue decomposition) 也称为谱分解 (spectral decomposition) 为从空间 $\mathbb{R}^p$ 到其本身通过 $p \times p$ 方阵 $\boldsymbol{A}$ 的映射, $\boldsymbol{A} : \mathbb{R}^p \mapsto \mathbb{R}^p$. 如果秩为 r 的矩阵 $\boldsymbol{A}$ 的非零特征值和相应的特征向量矩阵分别为 $\lambda_1, \lambda_2, \ldots, \lambda_r$ 和 $\boldsymbol{\Gamma} = [\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_r]$, 则有

$$\boldsymbol{A}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda},$$

这里, $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$. 如果 $\boldsymbol{A}$ 满秩, 则有特征值分解

$$\boldsymbol{A} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^{-1}.$$

矩阵的分解: 奇异值分解

奇异值分解 (singular value decomposition, SVD) 是上述特征值分解的推广: 每个具有秩 r 的矩阵 $\boldsymbol{A}(n \times p)$ 能够分解成

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H, \tag{0.1}$$

这里矩阵 $\boldsymbol{U}_{n\times n}$ 和 $\boldsymbol{V}_{p\times p}$ 都是酉矩阵 (对于复数矩阵 $\boldsymbol{A}$)[1]或列正交 (对于实矩阵 $\boldsymbol{A}$) 的. 矩阵 $\boldsymbol{\Sigma}$ 为非负实元素组成的 $n \times p$ 对角矩阵. 矩阵 $\boldsymbol{\Sigma}$ 的非零对角线元素称为 $\boldsymbol{A}$ 的非零奇异值, 这些奇异值为 $\boldsymbol{A}\boldsymbol{A}^H$ 和 $\boldsymbol{A}^H\boldsymbol{A}$ 的共同非零特征值的平方根.[2]

[1]矩阵 $\boldsymbol{U}$ 为酉矩阵 (unitary matrix, 也称幺矩阵) 的条件是其共轭转置 (亦称 Hermitian 转置) $\boldsymbol{U}^H$ 是其逆矩阵, 即 $\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^H = \boldsymbol{I}$.

[2]矩阵 $\boldsymbol{A}\boldsymbol{A}^H$ 和 $\boldsymbol{A}^H\boldsymbol{A}$ 的正交特征向量分别称为 $\boldsymbol{A}$ 的左奇异向量 (left-singular vectors) 和右奇异向量 (right-singular vectors).

矩阵的分解: 广义奇异值分解

下面的广义奇异值分解 (generalized singular value decomposition, GSVD) 是前面奇异值分解的带约束的推广. 假定矩阵 $\boldsymbol{A}_{n\times p}$ 的奇异值分解为:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^H,$$

这里的约束为

$$\boldsymbol{U}^H\boldsymbol{W}_u\boldsymbol{U} = \boldsymbol{V}^H\boldsymbol{W}_v\boldsymbol{V} = \boldsymbol{I}.$$

这意味着, 给定约束 $\boldsymbol{W}_u$ 和 $\boldsymbol{W}_v$, 矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 为正交的. $\boldsymbol{W}_u$ 和 $\boldsymbol{W}_v$ 往往为正定的, 甚至是正定对角矩阵.

广义奇异值分解 (定义 2)

还有一种广义奇异值分解: 对于 $m\times n$ 矩阵 $\boldsymbol{A}$ 和 $p\times n$ 矩阵 $\boldsymbol{B}$, 它们的广义奇异值分解由下面一对因子化定义:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}_1[0, \boldsymbol{R}]\boldsymbol{Q}^\top;$$

$$\boldsymbol{B} = \boldsymbol{V}\boldsymbol{\Sigma}_2[0, \boldsymbol{R}]\boldsymbol{Q}^\top.$$

矩阵的分解: QR 分解

一个 $n \times n$ 实方阵 $\boldsymbol{A}$ 可以分解为

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R},$$

这里的 $\boldsymbol{Q}$ 是一个 $n \times n$ 正交矩阵, 即 $\boldsymbol{Q}^\top \boldsymbol{Q} = \boldsymbol{I}$, 而 $\boldsymbol{R}$ 是一个 $n \times n$ 上三角矩阵. 如果 $\boldsymbol{A}$ 可逆, 则该分解是唯一的. 如果矩阵 $\boldsymbol{A}$ 的列线性独立, 则 $\boldsymbol{Q}$ 的列形成 $\boldsymbol{A}$ 的列空间的正交基. 对于复数矩阵, 上述分解中的 $\boldsymbol{Q}$ 为酉矩阵, 即 $\boldsymbol{Q}^H \boldsymbol{Q} = \boldsymbol{I}$, 其中 $\boldsymbol{Q}^H$ 为 $\boldsymbol{Q}$ 的共轭转置.

矩阵的分解: Cholesky 分解

如果 $\boldsymbol{A}$ 为实质对称正定矩阵, Cholesky 分解为:

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^\top,$$

这里的 $\boldsymbol{L}$ 是一个下三角矩阵. 如果 $\boldsymbol{A}$ 为更一般的 Hermitian 正定矩阵, Cholesky 分解为:

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^H.$$

这里 $\boldsymbol{L}$ 是一个下三角矩阵, 而 $\boldsymbol{L}^H$ 是 $\boldsymbol{L}$ 的共轭转置矩阵, 对于实对称正定矩阵 (及更一般的 Hermitian 正定矩阵), 都有唯一的 Cholesky 分解.

二次型: 定义

二次型 $Q(\boldsymbol{x})$ 是一个 $\mathbb{R}^p \mapsto \mathbb{R}$ 的投影, 由对称矩阵 $\boldsymbol{A}_{p\times p} = \{a_{ij}\}$ $(a_{ij} = a_{ji})$ 和向量 $\boldsymbol{x} = (x_1, x_2, \ldots, x_p) \in \mathbb{R}^p$ 组成:

$$Q(x) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}x_i x_j.$$

二次型具有唯一性, 如果对于对称的 $\boldsymbol{A}$ 和 $\boldsymbol{B}$, $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{B}\boldsymbol{x}, \forall \boldsymbol{x} \in \mathbb{R}^p$, 那么 $\boldsymbol{A} = \boldsymbol{B}$.

对于常数 a, 满足 $Q(\boldsymbol{x}) = a$ 的 $\boldsymbol{x}$ (记为 $\{\boldsymbol{x} | Q(\boldsymbol{x}) = a\}$) 形成的曲面称为二次曲面, 而 $\{\boldsymbol{x} | Q(\boldsymbol{x}) \leqslant a\}$ 称为二次区域.

二次型: 矩阵的定性

考虑对称矩阵 $\boldsymbol{A} \in \mathbb{R}^{p\times p}$. 如果 $Q(x) > 0, \ \forall x \neq 0$, 称 $\boldsymbol{A}$ 为正定的, 并记为 $\boldsymbol{A} > 0$. $\boldsymbol{A} > 0$ 的充分必要条件为其最小特征值大于零. 如果 $Q(x) \geqslant 0, \ \forall x \neq 0$, 称 $\boldsymbol{A}$ 为半正定的, 并记为 $\boldsymbol{A} \geqslant 0$. $\boldsymbol{A} \geqslant 0$ 的充分必要条件为其特征值均为非负. 类似地, 可以定义负定和半负定.

二次型: 距离

令 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^p$, 而 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 之间的距离定义为下面的函数 d.

$$d : \mathbb{R}^{2p} \mapsto \mathbb{R}_+ \text{ 满足 } \begin{cases} d(\boldsymbol{x}, \boldsymbol{y}) > 0, & \forall \boldsymbol{x} \neq \boldsymbol{y}; \\ d(\boldsymbol{x}, \boldsymbol{y}) = 0, & \text{当且仅当 } \boldsymbol{x} = \boldsymbol{y}; \\ d(\boldsymbol{x}, \boldsymbol{y}) \leqslant d(\boldsymbol{x}, \boldsymbol{z}) + d(\boldsymbol{z}, \boldsymbol{y}), & \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}\ (\boldsymbol{z} \in \mathbb{R}^p). \end{cases}$$

$\boldsymbol{x}$ 和 $\boldsymbol{y}$ 之间的欧式距离 d 定义为:

$$d^2(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} - \boldsymbol{y})^\top \boldsymbol{A} (\boldsymbol{x} - \boldsymbol{y}),$$

这里的 $\boldsymbol{A}$ 为正定矩阵 $(\boldsymbol{A} > 0)$. $\boldsymbol{A}$ 称为一个度量, 特例为 $\boldsymbol{A} = \boldsymbol{I}_p$.

二次型: 椭球

集合 $E_d = \{x \in \mathbb{R}^p | (\boldsymbol{x} - \boldsymbol{x}_0)^\top (\boldsymbol{x} - \boldsymbol{x}_0) = d^2\}$, 即以 $\boldsymbol{x}_0$ 为中心、d 为半径的球面.
更一般的情况为:

$$E_d = \{x \in \mathbb{R}^p | (\boldsymbol{x} - \boldsymbol{x}_0)^\top \boldsymbol{A} (\boldsymbol{x} - \boldsymbol{x}_0) = d^2\}.$$

它是以 $\boldsymbol{x}_0$ 为中心的椭球面. 如果这一等式中的等号改成小于等于号 ($\leqslant d^2$), 则为以 $\boldsymbol{x}_0$ 为中心的椭球.
令 $\gamma_1, \gamma_2, \ldots, \gamma_p$ 为 $\boldsymbol{A}$ 的相应于特征值 $\lambda_1, \lambda_2, \ldots, \lambda_p$ 的正交特征向量, 有

- (i) E_d 的主轴方向为 $\gamma_i, i = 1, 2, \ldots, p$;
- (ii) 半轴长度为 $\sqrt{d^2/\lambda_i}, i = 1, 2, \ldots, p$;
- (iii) 围绕 E_d 的矩形由下列不等式定义:

$$x_{0i} - \sqrt{d^2 a^{ii}} \leqslant x_i \leqslant x_{0i} + \sqrt{d^2 a^{ii}}, \quad i = 1, 2, \ldots, p,$$

这里 a^{ii} 为 $\boldsymbol{A}^{-1}$ 的 (i, i) 元素. 围绕 E_d 的矩形的边平行于坐标轴.

矩阵的导数: 向量关于数量的偏导数

对于向量 $\boldsymbol{y} = [y_1, y_2, \ldots, y_m]^\top$ 及数量 x, 定义

$$\frac{\partial \boldsymbol{y}}{\partial x} = \left[\frac{\partial y_1}{\partial x}, \frac{\partial y_2}{\partial x}, \ldots, \frac{\partial y_m}{\partial x}\right]^\top = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x}, \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix},$$

这是投影 $\mathbb{R} \mapsto \mathbb{R}^m$.

矩阵的导数: 数量关于向量的偏导数

对于数量 y 及向量 $\boldsymbol{x}=[x_1,x_2,\ldots,x_n]^\top$, 定义

$$\frac{\partial y}{\partial \boldsymbol{x}}=\left[\frac{\partial y}{\partial x_1},\frac{\partial y}{\partial x_2},\ldots,\frac{\partial y}{\partial x_n}\right].$$

矩阵的导数：向量关于向量的偏导数

对于向量 $\boldsymbol{y} = [y_1, y_2, \ldots, y_m]^\top$ 及向量 $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^\top$，定义

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$$

矩阵的导数: 矩阵关于数量的偏导数

对于 $m \times n$ 矩阵 $\boldsymbol{y} = \{y_{ij}\}$ 及数量 x, 定义

$$\frac{\partial \boldsymbol{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x} & \frac{\partial y_{12}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x} \\ \frac{\partial y_{21}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{2n}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{m1}}{\partial x} & \frac{\partial y_{m2}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x} \end{bmatrix}.$$

矩阵的导数：数量关于矩阵的偏导数

对于数量 y 及 $p \times q$ 矩阵 $\boldsymbol{X} = \{x_{ij}\}$, 定义

$$\frac{\partial y}{\partial \boldsymbol{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{21}} & \cdots & \frac{\partial y}{\partial x_{p1}} \\ \frac{\partial y}{\partial x_{12}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{p2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{1q}} & \frac{\partial y}{\partial x_{2q}} & \cdots & \frac{\partial y}{\partial x_{pq}} \end{bmatrix}.$$

矩阵的导数：关于内积和二次型的偏导数

考虑向量 $\boldsymbol{x}=[x_1,x_2,\ldots,x_n]^\top$. 此外, 向量 $\boldsymbol{a},\boldsymbol{b}$ 及矩阵 $\boldsymbol{A}$ 不是 $\boldsymbol{x}$ 的函数. 有

$$\frac{\partial \boldsymbol{a}^\top \boldsymbol{x}}{\partial \boldsymbol{x}}=\frac{\partial \boldsymbol{x}\boldsymbol{a}^\top}{\partial \boldsymbol{x}}=\boldsymbol{a} \text{ 或 } \boldsymbol{a}^\top;$$

$$\frac{\partial \boldsymbol{b}^\top \boldsymbol{A}\boldsymbol{x}}{\partial x}=\boldsymbol{b}^\top\boldsymbol{A} \text{ 或 } \boldsymbol{A}^\top\boldsymbol{b};$$

$$\begin{aligned}\frac{\partial \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}}&=\boldsymbol{x}^\top(\boldsymbol{A}+\boldsymbol{A}^\top) \text{ 或 } (\boldsymbol{A}+\boldsymbol{A}^\top)\boldsymbol{x};\\&=2\boldsymbol{x}^\top\boldsymbol{A} \text{ 或 } 2\boldsymbol{A}\boldsymbol{x}\ (\text{当 } \boldsymbol{A} \text{ 对称时});\end{aligned}$$

$$\frac{\partial \boldsymbol{x}^\top \boldsymbol{x}}{\partial \boldsymbol{x}}=2\boldsymbol{x}^\top \text{ 或 } 2\boldsymbol{x};$$

$$\frac{\partial \boldsymbol{a}^\top \boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{b}}{\partial \boldsymbol{x}}=\boldsymbol{x}^\top(\boldsymbol{a}\boldsymbol{b}^\top+\boldsymbol{b}\boldsymbol{a}^\top) \text{ 或 } (\boldsymbol{a}\boldsymbol{b}^\top+\boldsymbol{b}\boldsymbol{a}^\top)\boldsymbol{x}.$$

矩阵的导数：函数的偏导数

考虑 $f: \mathbb{R}^p \mapsto \mathbb{R}$ 和 $p \times 1$ 向量 $\boldsymbol{x}$, 那么 $\partial f(\boldsymbol{x})/\partial \boldsymbol{x}$ 为一个偏导数的列向量 $\{\partial f(\boldsymbol{x})/\partial x_j\}(j = 1, \ldots, p)$, 而 $\partial f(\boldsymbol{x})/\partial \boldsymbol{x}^\top$ 为同一偏导数的行向量. $\partial f(\boldsymbol{x})/\partial \boldsymbol{x}$ 称为 f 的梯度.

$\partial^2 f(\boldsymbol{x})/\partial \boldsymbol{x}\partial \boldsymbol{x}^\top$ 为二阶导数, 是 $p \times p$ 矩阵, 元素为 $\left\{\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}\right\}$ $(i = 1, 2, \ldots, p;\ j = 1, 2, \ldots, p)$. $\partial^2 f(\boldsymbol{x})/\partial \boldsymbol{x}\partial \boldsymbol{x}^\top$ 称为 f 的 Hessian 矩阵.

二次型 $Q(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ 的 Hessian 矩阵为:

$$\frac{\partial \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}\partial \boldsymbol{x}^\top} = 2\boldsymbol{A}$$