多元正态分布的极大似然估计

宁之恒

多元正态分布

给定多元形式下的正态分布概率密度函数 ($\boldsymbol{x} \sim N_p(\boldsymbol{u}, \boldsymbol{\Sigma})$)：

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{u})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{u})]$$

其中 $\boldsymbol{x} \in \mathbb{R}^p, \boldsymbol{u} \in \mathbb{R}^p, \boldsymbol{\Sigma} \in \mathcal{S}_+^p$。

接着对于样本集 $\mathcal{D} = \{(\boldsymbol{x_1}, y_1), \ldots, (\boldsymbol{x_n}, y_n)\}$，其中 $\boldsymbol{x_i} \in \mathbb{R}^n, y_i \in \mathbb{R}, |\mathcal{D}| = n$。

参照定义，给出极大似然函数：

$$\begin{aligned} L(\boldsymbol{u}, \boldsymbol{\Sigma}) &= \prod_{i=1}^{n} f(\boldsymbol{x_i}) \\ &= \frac{1}{(2\pi)^{\frac{pn}{2}}|\boldsymbol{\Sigma}|^{\frac{n}{2}}} \exp[-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x_i}-\boldsymbol{u})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x_i}-\boldsymbol{u})] \end{aligned}$$

对其取对数，并进行化简：

$$\ln L(\boldsymbol{u}, \boldsymbol{\Sigma}) = -\frac{pn}{2}\ln(2\pi) - \frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x_i}-\boldsymbol{u})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x_i}-\boldsymbol{u})$$

矩阵求导公式

给出可能会用到的求导公式：

$$\frac{\partial \ln|\boldsymbol{X}|}{\partial \boldsymbol{X}} = \boldsymbol{X^{-T}} \tag{1}$$

$$\frac{\partial \ln(\boldsymbol{\lambda^T X^{-1} \lambda})}{\partial \boldsymbol{X}} = -(\boldsymbol{X^{-1}\lambda\lambda^T X^{-1}})^T \tag{2}$$

$$\frac{\partial(\boldsymbol{\lambda}-\boldsymbol{x})^T\boldsymbol{\Sigma^{-1}}(\boldsymbol{\lambda}-\boldsymbol{x})}{\partial \boldsymbol{x}} = [\boldsymbol{(\lambda-x)^T(\Sigma^{-T}+\Sigma^{-1})}]^T \tag{3}$$

接着记 $\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}} = \boldsymbol{A}$，其中 $\boldsymbol{X} \in \mathbb{R}^{n\times n}, f : \mathbb{R}^{n\times n} \to \mathbb{R}$。那么对于对称矩阵 $\boldsymbol{X} \in S^n$，我们有如下形式：

$$\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}} = \boldsymbol{A^T} + \boldsymbol{A} - \boldsymbol{A} \circ \boldsymbol{E} \tag{4}$$

其中 $\circ$ 为 Hadamard product，此处不给出证明。

已知 $\boldsymbol{X} \in \mathcal{S}^n$，则式(1)，(2) 可以变成如下的 (5)，(6)：

$$\frac{\partial \ln |\boldsymbol{X}|}{\partial \boldsymbol{X}} = \boldsymbol{X^{-1}} + \boldsymbol{X^{-T}} - \boldsymbol{X^{-T}} \circ \boldsymbol{E} \tag{5}$$

$$\begin{aligned} \frac{\partial \ln(\boldsymbol{\lambda^T X^{-1} \lambda})}{\partial \boldsymbol{X}} = &-(\boldsymbol{X^{-1} \lambda \lambda^T X^{-1}}) \\ &-(\boldsymbol{X^{-1} \lambda \lambda^T X^{-1}})^T \\ &+(\boldsymbol{X^{-1} \lambda \lambda^T X^{-1}})^T \circ \boldsymbol{E} \end{aligned} \tag{6}$$

由式(3)可得：

$$\begin{aligned} \frac{\partial \ln L(\boldsymbol{u}, \boldsymbol{\Sigma})}{\partial \boldsymbol{u}} &= -\frac{1}{2}\sum_{i=0}^{n} \frac{\partial (\boldsymbol{x_i} - \boldsymbol{u})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x_i} - \boldsymbol{u})}{\partial \boldsymbol{u}} \\ &= \sum_{i=1}^{n} \boldsymbol{\Sigma^{-1}(x_i - u)} \\ &= \boldsymbol{\Sigma^{-1} \sum_{i=1}^{n} (x_i - u)} \end{aligned} \tag{7}$$

由式 (5),(6)可得：

$$
\begin{aligned}
\frac{\partial \ln L(\boldsymbol{u},\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}} &= -\frac{n}{2}\frac{\partial \ln|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\Sigma}} - \frac{1}{2}\sum_{i=0}^{n}\frac{\partial (\boldsymbol{x}_i-\boldsymbol{u})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{u})}{\partial \boldsymbol{\Sigma}} \\
&= \frac{1}{2}[\sum_{\boldsymbol{i=0}}^{\boldsymbol{n}}(\boldsymbol{\Sigma^{-1}}(\boldsymbol{x}_i-\boldsymbol{u})(\boldsymbol{x}_i-\boldsymbol{u})^{\boldsymbol{T}}\boldsymbol{\Sigma^{-1}} - n\boldsymbol{\Sigma^{-1}})^T \\
&+ \sum_{\boldsymbol{i=0}}^{\boldsymbol{n}}(\boldsymbol{\Sigma^{-1}}(\boldsymbol{x}_i-\boldsymbol{u})(\boldsymbol{x}_i-\boldsymbol{u})^{\boldsymbol{T}}\boldsymbol{\Sigma^{-1}} - n\boldsymbol{\Sigma^{-1}}) \\
&- \sum_{\boldsymbol{i=0}}^{\boldsymbol{n}}(\boldsymbol{\Sigma^{-1}}(\boldsymbol{x}_i-\boldsymbol{u})(\boldsymbol{x}_i-\boldsymbol{u})^{\boldsymbol{T}}\boldsymbol{\Sigma^{-1}} - n\boldsymbol{\Sigma^{-1}}) \circ \boldsymbol{E}] \\
&= \sum_{\boldsymbol{i=0}}^{\boldsymbol{n}}(\boldsymbol{\Sigma^{-1}}(\boldsymbol{x}_i-\boldsymbol{u})(\boldsymbol{x}_i-\boldsymbol{u})^{\boldsymbol{T}}\boldsymbol{\Sigma^{-1}} - n\boldsymbol{\Sigma^{-1}}) \\
&- \frac{1}{2}\sum_{\boldsymbol{i=0}}^{\boldsymbol{n}}(\boldsymbol{\Sigma^{-1}}(\boldsymbol{x}_i-\boldsymbol{u})(\boldsymbol{x}_i-\boldsymbol{u})^{\boldsymbol{T}}\boldsymbol{\Sigma^{-1}} - n\boldsymbol{\Sigma^{-1}}) \circ \boldsymbol{E}
\end{aligned}
\tag{8}
$$

令式(7)等于 $\mathbf{0}$, 左右两边乘以 Σ^{-1}, 可得 $\hat{\boldsymbol{u}} = \sum_{\boldsymbol{i=0}}^{\boldsymbol{n}}\boldsymbol{x}_i = \overline{\boldsymbol{x}}$
把结果带入式(8), 并令式(8)等于 $\mathbf{0}$, 可得 $\hat{\boldsymbol{\Sigma}} = \frac{\boldsymbol{1}}{\boldsymbol{n}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}(\boldsymbol{x}_i-\overline{\boldsymbol{x}})(\boldsymbol{x}_i-\overline{\boldsymbol{x}})^{\boldsymbol{T}}$
可以看出多元正态分布的极大似然估计值 $\hat{\boldsymbol{u}}$ 为样本的均值，而 $\hat{\boldsymbol{\Sigma}}$ 为样本的协方差矩阵