Great Likelihood Estimation of the Multivariate Normal Distribution

2333105 Ning Zhiheng

Multivariate normal distribution

Give the probability density function of the normal distribution in multivariate form($\boldsymbol{x} \sim N_p(\boldsymbol{u}, \boldsymbol{\Sigma})$):

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

s.t. $\boldsymbol{x} \in \mathbb{R}^p, \boldsymbol{u} \in \mathbb{R}^p, \boldsymbol{\Sigma} \in \mathcal{S}_+^p$.
Then for the sample set $\mathcal{D} = \{(\boldsymbol{x_1}, y_1), \ldots, (\boldsymbol{x_n}, y_n)\}$, s.t. $\boldsymbol{x_i} \in \mathbb{R}^n$, $y_i \in \mathbb{R}$, $|\mathcal{D}| = n$. With reference to the definition, the great likelihood function is given:

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

Take the logarithm of it and simplify it:

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

Matrix derivation formula

Give the derivation formulas that may be used:

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

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

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

Then write $\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}} = \boldsymbol{A}$, which $\boldsymbol{X} \in \mathbb{R}^{n \times n} f : \mathbb{R}^{n \times n} \to \mathbb{R}$. For a symmetric matrix $\boldsymbol{X} \in S^n$, We have the following form:

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

$\circ$ is Hadamard product, proof is not given here.
It is known that $\boldsymbol{X} \in \mathcal{S}^n$, then (1)、(2) can transfer to (5)、(6):

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

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

From equation (3):

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

From equation (5)、(6):

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

let (7) equals $\mathbf{0}$, multiply the left and right sides by Σ^{-1}, then $\boldsymbol{\hat{u}} = \sum_{\boldsymbol{i=0}}^{\boldsymbol{n}} \boldsymbol{x_i} = \boldsymbol{\overline{x}}$
Take the result into equation (8), and let (8) equals $\mathbf{0}$, then $\boldsymbol{\hat{\Sigma}} = \frac{\boldsymbol{1}}{\boldsymbol{n}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}(\boldsymbol{x_i}-\boldsymbol{\overline{x}})(\boldsymbol{x_i}-\boldsymbol{\overline{x}})^{\boldsymbol{T}}$
It can be seen that the great likelihood estimate of the multivariate normal distribution $\boldsymbol{\hat{u}}$ is the sample mean, $\boldsymbol{\hat{\Sigma}}$ is the covariance matrix of the sample.