线性回归第七周作业

宋歌 2015080086 数52

4/11/2018

1. 考虑线性模型

$$Y = X\beta + e,\ e \sim N_n(0, \sigma^2 I_n)$$

其中Y是$n\times 1$的响应值向量，X是$n\times p$的设计矩阵，β是$p\times 1$的未知参数向量，e是$n\times 1$的随机误差向量，σ^2是未知的方差，$rank(X) = p < n$. 考虑假设

$$H_0 : L\beta = c$$

其中L是$r\times p$的常数矩阵，c是$r\times 1$的常数列向量，$rank(L) = r$. 试求出上述假设的似然比检验统计量的表达式。

According to the given conditions, we know that $y_1, \ldots, y_n$ are independent and

$$y_i \sim N(x_i^T\beta, \sigma^2),\ i = 1, \ldots, n$$

Then we can write the likelihood function

$$L(\beta, \sigma^2|X, Y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - x_i^T\beta)^2\right) = (\frac{1}{\sqrt{2\pi\sigma^2}})^n \exp\left(-\frac{1}{2\sigma^2}||Y - X^T\beta||^2\right)$$

By the definition of the likelihood ratio test statistic, we have

$$\lambda(X, Y) = \frac{\sup\limits_{\theta\in\Theta_0} L(\theta|X, Y)}{\sup\limits_{\theta\in\Theta} L(\theta|X, Y)},\ \theta = (\beta, \sigma^2)$$

When $\theta \in \Theta$, we already know that

$$\arg\sup_{\beta,\sigma^2} L(\beta, \sigma^2|X, Y) = (\hat{\beta}, \hat{\sigma}^2) = \left((X^TX)^{-1}X^TY,\ \frac{1}{n}||Y - \hat{Y}||^2\right)$$

When $\theta \in \Theta_0$ *i.e.* $L\beta = c$, we can see it as r constraints, that is

$$L\beta = c \Leftrightarrow l_j^T\beta = c_j,\ j = 1, \dots, r$$

Now we apply Lagrange Multiplier. First we write the Lagrange function as follows

$$\begin{aligned} l(\beta, \sigma^2, \lambda) &= logL(\beta, \sigma^2|X, Y) + \sum_{j=1}^{r} \lambda_j(l_j^T\beta - c_j) \\ &= nlog\frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2}||Y - X\beta||^2 + \lambda^T(L\beta - c) \end{aligned}$$

We first fix σ^2 and solve the following equations

$$\begin{cases} \frac{\partial l}{\partial \beta} = -\frac{1}{2\sigma^2}[-2X^TY + 2X^TX\beta] + L^T\lambda = 0 \\ \frac{\partial l}{\partial \lambda} = (L\beta - c)^T = 0 \end{cases}$$

Since $rank(X) = p < n$, we know that X^TX is invertible. Thus, by solving the first equation, we obtain

$$\tilde{\beta} = (X^TX)^{-1}[\sigma^2L^T\lambda + X^TY]$$

Since $rank(L) = r$ *i.e.* L is row full rank, we know that for any nonzero vector z, z^TL and L^Tz is nonzero. Thus, for any nonzero vector z, $z^TL(X^TX)^{-1}L^Tz > 0$, that is, $L(X^TX)^{-1}L^T$ is positive definite. Therefore, we can obtain $\tilde{\lambda}$ by plugging $\tilde{\beta}$ into the second equation, then we can also obtain $\tilde{\beta}$

$$\begin{aligned} \tilde{\lambda} &= \frac{1}{\sigma^2}\left[L(X^TX)^{-1}L^T\right]^{-1}c - \frac{1}{\sigma^2}\left[L(X^TX)^{-1}L^T\right]^{-1}L(X^TX)^{-1}X^TY \\ \tilde{\beta} &= (X^TX)^{-1}L^T\left[L(X^TX)^{-1}L^T\right]^{-1}\left[c - L(X^TX)^{-1}X^TY\right] + (X^TX)^{-1}X^TY \end{aligned}$$

When σ^2 is fixed, the $\tilde{\beta}$ we obtained by Lagrange Multiplier is the local extremum point of the likelihood function $L(\beta, \sigma^2)$ under the constraint $L\beta = c$. And because for any β

$$\frac{\partial^2 logL}{\partial \beta^2} = -\frac{X^TX}{\sigma^2} < 0$$

we know that when σ^2 is fixed, $\tilde{\beta}$ is the global (under the constraint $L\beta = c$) maximum point of L, that is

$$\sup_{\beta, L\beta=c} L(\beta, \sigma^2) = L(\tilde{\beta}, \sigma^2)$$

Now we are left to solve

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}||Y - X\beta||^2 = 0$$

Using the above $\tilde{\beta}$ and $\tilde{\lambda}$, we obtain

$$\begin{aligned}\tilde{\sigma}^2 &= \frac{1}{n}||Y - X\tilde{\beta}||^2 \\ &= \frac{1}{n}\left\| Y - X(X^TX)^{-1}L^T\left[L(X^TX)^{-1}L^T\right]^{-1}\left[c - L(X^TX)^{-1}X^TY\right] - X(X^TX)^{-1}X^TY\right\|^2\end{aligned}$$

We also find that

$$\begin{aligned}\frac{\partial l}{\partial \sigma^2} > 0 \quad if\ \sigma^2 < \tilde{\sigma}^2 \\ \frac{\partial l}{\partial \sigma^2} < 0 \quad if\ \sigma^2 > \tilde{\sigma}^2\end{aligned}$$

That is, $\tilde{\sigma}^2$ is the global maximum point of $L(\tilde{\beta}, \sigma^2)$ under the constraint $L\beta = c$. Thus,

$$\arg\sup_{\Theta_0} L(\beta, \sigma^2) = \arg\sup_{L\beta=c} L(\beta, \sigma^2) = (\tilde{\beta}, \tilde{\sigma}^2)$$

Therefore, the likelihood ratio test statistic can be written as

$$\lambda(X, Y) = \frac{L(\tilde{\beta}, \tilde{\sigma}^2)}{L(\hat{\beta}, \hat{\sigma}^2)} = \left(\frac{||Y - X^T\hat{\beta}||}{||Y - X^T\tilde{\beta}||}\right)^n$$

2. 阅读教材并完成题目：11.4, 11.8, 11.11

11.4 Suppose $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V})$ for known positive definite $n \times n$ matrix $\mathbf{V}$. Then the likelihood function is

$$L(\boldsymbol{\beta}, \sigma^2) = (\frac{1}{\sqrt{2\pi}})^n \frac{1}{|\mathbf{V}|^{1/2}} \frac{1}{\sigma^n} \exp\left(\frac{-1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right)$$

.

a) Suppose that $\hat{\boldsymbol{\beta}}_G$ minimizes $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$. Show that $\hat{\boldsymbol{\beta}}_G$ is the MLE of $\boldsymbol{\beta}$.

The log likelihood function is

$$l(\boldsymbol{\beta}, \sigma^2) = logL(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}log(2\pi) - \frac{1}{2}log|\mathbf{V}| - \frac{n}{2}log\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

For any fixed σ^2, we know that

$$\hat{\boldsymbol{\beta}}_{MLE} = \max_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) = \max_{\boldsymbol{\beta}} l(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \hat{\boldsymbol{\beta}}_G$$

b) Find the MLE $\hat{\sigma}^2$ of σ^2.

From a) we obtain the log likelihood function

$$l(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}log(2\pi) - \frac{1}{2}log|\mathbf{V}| - \frac{n}{2}log\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Then by setting

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0$$

We have

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_G)^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_G)$$

Since

$$\frac{\partial l}{\partial \sigma^2} > 0 \quad if\ \sigma^2 < \hat{\sigma}^2$$
$$\frac{\partial l}{\partial \sigma^2} < 0 \quad if\ \sigma^2 > \hat{\sigma}^2$$

Therefore

$$\hat{\sigma}^2_{MLE} = \arg\max_{\sigma^2} l(\hat{\boldsymbol{\beta}}_G, \sigma^2) = \hat{\sigma}^2$$

11.8 Let $Y \sim N(\mu, \sigma^2)$ so that $E(Y) = \mu$ and $Var(Y) = \sigma^2 = E(Y^2) - [E(Y)]^2$. If $k \geqslant 2$ is an integer, then

$$E(Y^k) = (k-1)\sigma^2 E(Y^{k-2}) + \mu E(Y^{k-1})$$

Let $Z = (Y - \mu)/\sigma \sim N(0, 1)$. Hence $\mu_k = E(Y - \mu)^k = \sigma^k E(Z^k)$. Use this fact and the above recursion relationship $E(Z^k) = (k-1)E(Z^{k-2})$ to find μ_3 and μ_4.

$$\mu_3 = \sigma^3 E(Z^3) = 2\sigma^3 E(Z) = 0$$
$$\mu_4 = \sigma^4 E(Z^4) = 3\sigma^4 E(Z^2) = 3\sigma^4$$

11.11 Consider the model $Y_i = \beta_1 + \beta_2 X_{i,2} + \cdots + \beta_p X_{i,p} + e_i = \mathbf{x_i^T}\boldsymbol{\beta} + e_i$. The least squares estimator $\hat{\boldsymbol{\beta}}$ minimizes

$$Q_{OLS}(\boldsymbol{\eta}) = \sum_{i=1}^{n} (Y_i - \mathbf{x_i^T}\boldsymbol{\eta})^2$$

and the weighted least squares estimator minimizes

$$Q_{WLS}(\boldsymbol{\eta}) = \sum_{i=1}^{n} w_i (Y_i - \mathbf{x_i^T}\boldsymbol{\eta})^2$$

where the w_i, Y_i and $\mathbf{x_i}$ are known quantities. Show that

$$\sum_{i=1}^{n} w_i (Y_i - \mathbf{x_i}^T\boldsymbol{\eta})^2 = \sum_{i=1}^{n} (\tilde{Y}_i - \tilde{\mathbf{x}}_i^T\boldsymbol{\eta})^2$$

by identifying $\tilde{Y}_i$, and $\tilde{\mathbf{x}}_i$. (Hence the WLS estimator is obtained from the least squares regression of $\tilde{Y}_i$ on $\tilde{\mathbf{x}}_i$ without an intercept.)

Let $\tilde{Y}_i = \sqrt{w_i} Y_i$, $\tilde{\mathbf{x}}_i = \sqrt{w_i}\mathbf{x}_i$, then we have

$$\sum_{i=1}^{n} w_i (Y_i - \mathbf{x_i}^T\boldsymbol{\eta})^2 = \sum_{i=1}^{n} (\tilde{Y}_i - \tilde{\mathbf{x}}_i^T\boldsymbol{\eta})^2$$