Sibei

Sibei Liu sl4660

2020/3/19

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & ... & x_{1p} \\ 1 & x_{21} & ... & x_{2P} \\ ... & ... & ... & ... \\ 1 & x_{n1} & ... & x_{np} \end{pmatrix}_{n\times(1+p)} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ ... \\ \mathbf{x}_n \end{pmatrix} \qquad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{pmatrix}_{n\times 1} \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ ... \\ \beta_p \end{pmatrix}_{(1+p)\times 1}$$

Note: $\mathbf{x}_n = \begin{pmatrix} 1 & x_{n1} & ... & x_{np} \end{pmatrix}_{1\times(p+1)}$ By using the logit link, we can get the logistic regression:

$$\log(\frac{\pi_i}{1-\pi_i}) = \mathbf{x}_i\boldsymbol{\beta} = \theta_i\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ ... \\ \theta_n \end{pmatrix}$$

The density function is :

$$f_Y(y_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i}$$

Log likelihood function:

$$l(\mathbf{Y},\boldsymbol{\pi}) = \sum_{i=1}^{n} l(y_i,\pi_i) = \sum_{i=1}^{n}(y_i log\frac{\pi_i}{1-\pi_i} + log(1-\pi_i)) = \sum_{i=1}^{n}(y_i log\frac{\pi_i}{1-\pi_i} + log(1-\pi_i))$$

Take θ_i into above equation :

$$l(\mathbf{Y},\boldsymbol{\theta}) = \sum_{i=1}^{n}(y_i\theta_i - log(1+e^{\theta_i})) = \mathbf{Y}^T\boldsymbol{\theta} - \sum_{i=1}^{n} log(1+e^{\theta_i})$$

Due to the relationship:

$$\mathbf{x}_i\boldsymbol{\beta} = \theta_i \qquad \mathbf{X}\boldsymbol{\beta} = \boldsymbol{\theta}$$

Take β into log likelihood function:

$$l(\mathbf{Y},\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i\mathbf{x}_i\boldsymbol{\beta} - log(1+e^{\mathbf{x}_i\boldsymbol{\beta}})) = Y^T\mathbf{X}\boldsymbol{\beta} - \sum_{i=1}^{n} log(1+e^{\mathbf{x}_i\boldsymbol{\beta}}))$$

Take the first derivative of log likelihood function about $\boldsymbol{\beta}$ to get the gradient:

$$\nabla f(\mathbf{Y},\boldsymbol{\theta}) = (\frac{\partial\boldsymbol{\theta}}{\partial\beta})^T \times \frac{\partial l(\mathbf{Y},\boldsymbol{\beta})}{\partial\boldsymbol{\theta}} = \mathbf{X}^T(\mathbf{Y} - \frac{e^{\boldsymbol{\theta}}}{1+e^{\boldsymbol{\theta}}})$$

Take derivative of score function to get the Hessian:

$$\nabla^2 f(\mathbf{Y},\boldsymbol{\theta}) = -\mathbf{X}^T diag(\pi_i(1-\pi_i))\mathbf{X}$$