Problem Set - Logistic Regression

September 25, 2021

1 The Gradient and Hessian Matrix of Logistic Regression

Given n observation $y_1, y_2 ..., y_n$ that follows Bernoulli Distribution and the corresponding predictor vectors $\vec{x}_1$, $\vec{x}_2$, ..., $\vec{x}_n$, here $\vec{x}_i$ has m components $\vec{x}_i = [x_{i1}, x_{i2}, ...x_{ij}, ...x_{im}]$. The total likelihood is therefore $L(y_1, y_2, ...y_n, p_1, p_2, ..., p_n) = \prod_{i=1}^{n} {p_i}^{y_i}(1-p_i)^{1-y_i}$. Assume p_i is estimated by a logistic function $p_i = \frac{1}{1+\exp(-\vec{\beta}\cdot\vec{x}_i)}$

1. Prove that the log-likelihood function $\ell = \sum_{i=1}^{n}\left(y_i(\vec{x}_i \cdot \vec{\beta}) - \log(1+\exp(\vec{x}_i \cdot \vec{\beta}))\right)$?
2. Prove that $\frac{\partial \vec{x}_i \cdot \vec{\beta}}{\partial \beta_j} = x_{ij}$.
3. Prove that

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n}\left(y^i - \frac{1}{1+\exp(-\vec{\beta}\cdot\vec{x}^i)}\right)x_j^i \text{ , } j = 1, 2, 3, ..., m$$

.

4. Use the results from (3) to calculate the Hessian matrix and prove that

$$\frac{\partial^2 \ell}{\partial \beta_a \partial \beta_b} = -\sum_{i=1}^{n} x_b^i p_i(1-p_i)x_a^i$$

5. Prove that the jth component of the column matrix $\nabla \ell = X^T(y-p)$ is the same as $\frac{\partial \ell}{\partial \beta_j}$, here$y$ and p are column vectors $y = [y_1, y_2, ..., y_n]^T$, $p = [p_1, p_2, ..., p_n]^T$. X is the predictor matrix

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2m} \\ & & \ddots & & & \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{nm} \end{bmatrix}$$

6. Prove that ath row and bth column of $\mathbf{H}$, $\mathbf{H}_{ab} = -\sum_{i=1}^{n} x_b^i p_i(1-p_i)x_a^i$, where

$$\mathbf{H} = -X^TWX, W = \begin{bmatrix} p_1(1-p_1) & & \\ & \ddots & \\ & & p_n(1-p_n) \end{bmatrix}$$