STAT 600 Homework 3

Jingkun Lin UIN: 436001541

September 24, 2024

$$f(\beta) = \left[-\sum_{i=1}^{n} \left\{ \sum_{k=1}^{K-1} 1(y_i = k) \log p_k(x_i; \beta) \right\} + \frac{\lambda}{2} \sum_{k=0}^{K-1} \sum_{j=1}^{P} \beta_{k,j}^2 \right]$$

$$p_k(x_i, \beta) = \frac{\exp(x_i^T \beta_k)}{\sum_{l=0}^{K-1} \exp(x_i^T \beta_l)}$$

$$\frac{\partial}{\partial \beta_j} f(\beta) = \left[-\sum_{i=1}^{n} \left\{ \sum_{k=1}^{K-1} 1(y_i = k) \frac{\frac{\partial}{\partial \beta_j} p_k(x_i; \beta)}{p_k(x_i; \beta)} \right\} + \lambda \beta_k \right]$$

If $k = j$

$$\frac{\partial}{\partial \beta_k} p_k(x_i, \beta) = \frac{1}{(\sum_{l=0}^{K-1} \exp(x_i^T \beta_l))^2} (\sum_{l=0}^{K-1} \exp(x_i^T \beta_l) - \exp(x_i^T \beta_k)) x_i \exp(x_i^T \beta_k) = x_i p_k(x_i, \beta)(1 - p_k(x_i, \beta))$$

else:

$$\frac{\partial}{\partial \beta_j} p_k(x_i, \beta) = -\frac{1}{(\sum_{l=0}^{K-1} \exp(x_i^T \beta_l))^2} (\exp(x_i^T \beta_j) x_i \exp(x_i^T \beta_k)) = -x_i p_j(x_i, \beta) p_k(x_i, \beta)$$

$$\frac{\partial}{\partial \beta_j} f(\beta) = \left[-\sum_{i=1}^{n} \left\{ \sum_{k=1}^{K-1} 1(y_i = k) 1(k = j) x_i (1 - p_k(x_i, \beta)) - (1 - 1(j = k)) x_i p_j(x_i, \beta) \right\} + \lambda \beta_j \right]$$

Compact expression:

$$\frac{\partial}{\partial \beta_j} f(\beta) = X^\top \{P_j - 1(Y = k)\} + \lambda \beta_j^{(t)}$$

Then we calculate each element of Hessian. For each term:

$$\frac{\partial^2}{\partial \beta_j^2} f(\beta)_i = x_i \frac{\partial}{\partial \beta_k} p_k(x_i, \beta) + \lambda$$

$$\frac{\partial^2}{\partial \beta_j^2} f(\beta)_i = x_i^2 p_k(x_i, \beta)(1 - p_k(x_i, \beta))$$

$$\frac{\partial^2}{\partial \beta_j^2} f(\beta) = X^T \{W_k\} X + \lambda$$

where $W_k = \text{diag}(w_{11}, w_{22}, \ldots, w_{n,n})$, $w_{kii} = p_k(x_i, \beta)(1 - p_k(x_i, \beta))$

Therefore, we plug the results in the formula of damped Newton's method, we got:

$$\beta_k^{(t+1)} = \beta_k^{(t)} - \eta (X^\top W_k X + \lambda I)^{-1} \left[X^\top \{P_k - 1(Y = k)\} + \lambda \beta_k^{(t)} \right], \quad k = 0, \ldots, K-1;$$