STAT600 HW3

Zexian (Leo) Wang   936002904   leowang@tamu.edu

The objective function is

$$f(\beta) = -\sum_{i=1}^{n}\left\{\sum_{k=0}^{K-1}\mathbf{1}_{(y_i=k)}\log p_k(x_i;\beta)\right\} + \frac{\lambda}{2}\sum_{k=0}^{K-1}\sum_{j=1}^{p}\beta_{k,j}^2 \qquad p_k(x_i;\beta) = \frac{e^{x_i^\top\beta_k}}{\sum_{l=0}^{K-1}e^{x_i^\top\beta_l}}$$

$$\frac{\partial p_k(x_i;\beta)}{\partial\beta_k} = \frac{x_i e^{x_i^\top\beta_k}\sum_{l=0}^{K-1}e^{x_i^\top\beta_l} - x_i e^{x_i^\top\beta_k}e^{x_i^\top\beta_k}}{\left(\sum_{l=0}^{K-1}e^{x_i^\top\beta_l}\right)^2} = x_i\left(\frac{e^{x_i^\top\beta_k}}{\sum_{l=0}^{K-1}e^{x_i^\top\beta_l}} - \frac{\left(e^{x_i^\top\beta_k}\right)^2}{\left(\sum_{l=0}^{K-1}e^{x_i^\top\beta_l}\right)^2}\right)$$

$$= x_i\left(p_k(x_i;\beta) - (p_k(x_i;\beta))^2\right) = x_i p_k(x_i;\beta)(1 - p_k(x_i;\beta))$$

$$\frac{\partial p_j(x_i;\beta)}{\partial\beta_k} = \frac{-x_i e^{x_i^\top\beta_j}e^{x_i^\top\beta_k}}{\left(\sum_{l=0}^{K-1}e^{x_i^\top\beta_l}\right)^2} = -x_i p_j(x_i;\beta)p_k(x_i;\beta),\ j\neq k$$

$$\nabla f(\beta_k) = -\sum_{i=1}^{n}x_i\left(\mathbf{1}_{(y_i=k)} - p_k(x_i;\beta)\right) + \lambda\beta_k = \sum_{i=1}^{n}x_i\left(p_k(x_i;\beta) - \mathbf{1}_{(y_i=k)}\right) + \lambda\beta_k = X^\top\left(P_k - \mathbf{1}_{(Y=k)}\right) + \lambda\beta_k$$

$$[\nabla^2 f(\beta_k)]_{uu} = \sum_{i=1}^{n}x_i\left(\frac{\partial p_k(x_i;\beta)}{\partial\beta_k} - \mathbf{1}_{(y_i=k)}\right) + \frac{\partial(\lambda\beta_k)}{\partial\beta_k} = \sum_{i=1}^{n}x_{i,u}^2 p_k(x_i;\beta)(1 - p_k(x_i;\beta)) + \lambda,\ u = 1,\cdots,n$$

$$[\nabla^2 f(\beta_k)]_{uv} = \sum_{i=1}^{n}x_{i,u}x_{i,v}p_k(x_i;\beta)(1 - p_k(x_i;\beta)),\ \text{for}\ u\neq v,\ u = 1,\cdots,n, v = 1,\cdots,n$$

$$\nabla^2 f(\beta_k) = \frac{\partial\left(X^\top\left(P_k - \mathbf{1}_{(Y=k)}\right) + \lambda\beta_k\right)}{\partial\beta_k} = X^\top W_k X + \lambda I$$

where W_k is a $n\times n$ diagonal matrix with $W_{k,ii} = p_k(x_i;\beta)(1 - p_k(x_i;\beta))$. For $k = 0,\cdots,K-1$,

$$\beta_k^{(t+1)} = \beta_k^{(t)} - \eta\left(\nabla^2 f(\beta_k)\right)^{-1}\nabla f(\beta_k)$$

$$\beta_k^{(t+1)} = \beta_k^{(t)} - \eta\left(X^\top W_k X + \lambda I\right)^{-1}\left(X^\top\left(P_k - \mathbf{1}_{(Y=k)}\right) + \lambda\beta_k^{(t)}\right)$$