Homework 4
CS-4033
Roman Munoz

1) $Pr(y_i=0|x_i) = \frac{1}{1+\exp(-x_i^T\beta)}$, $Pr(y_i=1|x_i) = \frac{\exp(-x_i^T\beta)}{1+\exp(-x_i^T\beta)}$

$$L(\beta) = \sum_{i=1}^{n} \log Pr(y_i|x_i)$$

$$\lg P(y_i|x_i) = y_i \cdot \lg(y_i=1|x_i) + (1-y_i)\cdot \lg(P[y_i=0|x_i])$$
$$\lg P(y_i=1|x_i) = -\lg[\exp(-x_i^T\beta)]$$
$$\lg P(y_i=0|x_i) = \lg \exp(x_i^T\beta) - \lg[\exp(x_i^T\beta)+1]$$
$$\Rightarrow y_i(-\lg[1+\exp_i]) + (1-y_i)\cdot[x_i^T\beta - \lg(\exp(x_i^T\beta)+1]$$
$$= y_i \lg[\cancel{1+\exp}(x_i^T\beta)] + (x_i^T\beta) - \lg(\exp(x_i^T\beta)) - y_i\cdot(x_i^T\beta) + y\cdot\lg(\cancel{\exp(x_i^T\beta)+1})$$
$$= (1-y_i)(x_i^T\beta) - \lg(\exp(x_i^T\beta+1)), \text{ Thus,}$$
$$L(\beta) = \sum_{i=1}^{n} (1-y_i)x_i^T\beta - \log[1+\exp(x_i^T\beta)]$$

2) $\lg P(y_i|x_i) = (1-y_i)x_i^T\beta - \log[1+\exp(x_i^T\beta)]$

$$\frac{\partial \lg P(y_i|x_i)}{\partial\beta} = (1-y_i)x_i - \frac{\exp(x_i^T\beta)\cdot x_i}{1+\exp(x_i^T\beta)} = \left[(1-y_i) - \frac{\exp(x_i^T\beta)}{1+\exp(x_i^T\beta)}\right]\cdot x_i$$
$$\Rightarrow [(1-y_i) - Pr(y_i=0|x_i)]\cdot x_i = [-y_i - [1-Pr(y_i=0|x_i)]]\cdot x_i$$
$$\Rightarrow -(y_i - Pr(y_i=1|x_i))\cdot x_i, \text{ Therefore,}$$
$$\frac{\partial L(\beta)}{\partial\beta} = -\sum_{i=1}^{n}(y_i - Pr(y_i=1|x_i))\cdot x_i$$
, which is equals to the matrix form: $-X^T(Y-P_1)$, where P_1 is an n-dimensional vector with the ith element being $Pr(y_i=1|x_i)$

4. $J(\beta) = L(\beta*) + \left(\frac{\partial L(\beta*)}{\partial \beta}\right)^T (\beta - \beta*) + \frac{1}{2}(\beta - \beta*)^T \left(\frac{\partial L(\beta*)}{\partial \beta \partial \beta^T}\right)(\beta - \beta*)$

Now to optimize β:

$$\frac{\partial J(\beta)}{\partial(\beta - \beta*)} = 0 + \left(\frac{\partial L(\beta*)}{\partial \beta}\right)^T + \frac{1}{2} \cdot 2 \left[\frac{\partial L(\beta*)}{\partial \beta \partial \beta^T}\right](\beta - \beta*)$$

$$= \left(\frac{\partial L(\beta*)}{\partial \beta}\right) + \left(\frac{\partial L(\beta*)}{\partial \beta \partial \beta^T}\right)(\beta - \beta*)$$

$$\Rightarrow \left(\frac{\partial L(\beta*)}{\partial \beta \partial \beta^T}\right)(\beta - \beta*) = -\left(\frac{\partial L(\beta*)}{\partial \beta}\right)$$

$$\beta - \beta* = -\left(\frac{\partial L(\beta*)}{\partial \beta \partial \beta^T}\right)^{-1} \cdot \left(\frac{\partial L(\beta*)}{\partial \beta}\right)$$

Thus,

$$\beta = \beta* - \left(\frac{\partial L(\beta*}{\partial \beta \partial \beta^T}\right)^{-1} \cdot \left(\frac{\partial L(\beta*)}{\partial \beta}\right)$$