HW #4

Thomas Chan

SID 222155698

Problem 1

① $\mathcal{L}(\beta) = \lambda \|\beta\|_2^2 - \sum_{i=1}^{n} [y_i \log \mu_i + (1-y_i)\log(1-\mu_i)]$

We know $\mu_i = \frac{1}{1+e^{-\beta^T x_i}}$

$$\mathcal{L}(\beta) = \lambda\|\beta\|_2^2 - \sum_{i=1}^{n} [y_i \log\left(\frac{1}{1+e^{-\beta^T x_i}}\right) + (1-y_i)\log\left(1 - \frac{1}{1+e^{-\beta^T x}}\right)$$

$$\mathcal{L}(\beta) = \lambda\|\beta\|_2^2 - \sum_{i=1}^{n} \left[y_i \log\left(\frac{e^{\beta^T x}}{1+e^{\beta^T x}}\right) + (1-y_i)\log\left(\frac{e^{-\beta^T x}}{1+e^{-\beta^T x}}\right)\right]$$

$$\mathcal{L}(\beta) = \lambda\|\beta\|_2^2 - \sum_{i=1}^{n} y_i \log\left(\frac{e^{\beta^T x}}{1+e^{\beta^T x}}\right) + (1-y_i)\log\left(\frac{1}{1+e^{\beta^T x}}\right)$$

$$\mathcal{L}(\beta) = \lambda\|\beta\|_2^2 - \sum_{i=1}^{n} y_i \left[\log e^{\beta^T x} - \log(1+e^{\beta^T x})\right]$$

$$\mathcal{L}(\beta) = \lambda\|\beta\|_2^2 - \sum_{i=1}^{n} y_i \left[\beta^T x - \log(1+e^{\beta^T x})\right]$$

$$\frac{\partial \mathcal{L}(\beta)}{\partial \beta} = 2\lambda\beta - \sum_{i=1}^{n} y_i \left[x_i - \frac{x_i e^{\beta^T x_i}}{1+e^{\beta^T x}}\right]$$

$$\boxed{\frac{\partial \mathcal{L}(\beta)}{\partial \beta} = 2\lambda\beta = \sum x_i (y_i - \mu_i)}$$

② $$\frac{\partial \nabla_f \beta}{\partial \beta} = 2\lambda - \sum -x_i^2 \left[\frac{e^{-\beta^T x_i}}{1+e^{-\beta^T x_i}}\right]\mu_i$$

$$= 2\lambda + \sum_{i=1}^{n} x_i^2 (1-\mu_i)\mu_i$$

$$\boxed{\frac{\partial \nabla_f \beta}{\partial \beta} = 2\lambda + \sum x_i^2 (1-\mu_i)\mu_i}$$

(3) Newton's: $\beta_{k+1} = \beta_k - H^{-1} \nabla f$

$$\boxed{\beta_{k+1} = \beta_k - \left[2\lambda + \sum_i^N x_i^2 (1-\mu_i)\,\mu_i\right]^{-1} \left[2\lambda\beta - \sum_i x_i (y_i - \mu_i)\right]}$$

(4) $\lambda = 0.07$, $\beta_0 = [2, 1, 0]^T$

(a). State value of μ_0.

$$\mu_0 = \frac{1}{1+e^{-\beta^T x_i}} = \boxed{[0.4526,\ 0.7311,\ 0.7311,\ 0.2689]}$$

calculated via MATLAB.

(b). β_1.

$$\beta_1 = \beta_0 - H^{-1} \nabla f(\beta)$$

$$\boxed{\beta_1 = [-0.3868,\ 1.4043,\ -2.2842]}$$

(c) μ_1.

$$\boxed{\mu_1 = [0.8731,\ 0.8258,\ 0.2932,\ 0.2198]}$$

(d). β_2

$$\boxed{\beta_2 = [-0.5122,\ 1.4527,\ -2.1627]}$$