MATH 7243
HW 3 Part a

1/ $$P(Y|X) = \frac{e^{\beta_k^T \vec{x}}}{\sum_{j=1}^{K} e^{\beta_k^T \vec{x}}} \quad , \quad y \in \{1, 0\}$$

$$P(y|\vec{x};\beta) = h_\beta(\vec{x})^y \left(1 - h_\beta(\vec{x})\right)^{1-y}$$

$$P(y=1|\vec{x};\beta) = h_\beta(\vec{x})$$

$$P(y=0|\vec{x};\beta) = 1 - h_\beta(\vec{x})$$

$$L(\beta) = \prod_{i=1}^{n} h(x_i)^{y_i} \left(1 - h(x_i)\right)^{1-y_i}$$

(1) Minimize :- Negative log likelihood:

$$\ell(\beta_k) = -\ln(L(\beta)) = -\sum_{i=1}^{n} \left(y_i \cdot \ln h(x_i) + (1 - y_i) \left(\ln\left(1 - h(x_i)\right)\right)\right)$$

$$\ell(\beta_k) = \sum_{i=1}^{n} \sum_{j=1}^{K} y_{ij} \left(\beta_k^T x_i + \ln\left(\sum_{j=1}^{K} e^{-\beta_j^T x_i}\right)\right)$$

(2) $$J(\beta) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K} y_{ij} \left(\beta_k^T x_i + \ln\left(\sum_{j=1}^{K} e^{-\beta_j^T x_i}\right)\right)$$

$$\frac{dJ(\beta)}{d\beta_k} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K} y_{ij} \left(\beta_k^T x_i + \ln\left(\sum_{j=1}^{K} e^{-\beta_j^T x_i}\right)\right) + \lambda \sum_{k=1}^{K} \|\beta_k\|^2$$

$$= \frac{1}{n} \sum_{i=1}^{N} y_{ij} \left(1 - \frac{y_{ij}\, e^{-\beta_j^T x_i}}{\sum_{j=1}^{K} e^{-\beta_j^T x_i}}\right) x_i + 2\lambda\beta_k$$

(3) Stochastic GD: Update $\beta = \beta_k$

$$\frac{dJ_i}{d\beta_k} = y_{ij}\left(1 - \frac{y_{ij}\, e^{-\beta_j^T x_i}}{\sum_{j=1}^{k} e^{-\beta_j^T x_i}}\right) x_i + 2\lambda\beta_k$$

β for each epoch: $\beta = \beta - \eta\left(\frac{dJ_i}{d\beta_k}\right)$

Batch GD: Divide data in batches

$d_1, \ldots, d_n$ with corresponding $J_1, \ldots, J_n$

and then $\beta = \beta - \eta\left(\frac{dJ_T}{d\beta_k}\right)$; $1 \leq T \leq n$

2 Gaussian Discriminant Function: Code attached

a) $\Sigma = \begin{pmatrix} 2.73532 & -1.968208 \\ -1.968208 & 3.759784 \end{pmatrix}$

$\mu_1 = \begin{pmatrix} 2.088 \\ 2.186 \end{pmatrix}$, $\mu_2 = \begin{pmatrix} -2.156 \\ -1.72 \end{pmatrix}$

$g_i(x) = \frac{-1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln p(\omega_i)$

Label 1:

$g_1(x_1) = \frac{\mu_1}{\sigma_1^2} - \frac{\mu_1^2}{2\sigma_1^2} + \ln(p(\omega_1))$

$= \frac{2.088}{1.932816} - \frac{(2.088)^2}{2(1.932816)} + \ln(0.5) = -0.7407$

$g_2(x_1) = \frac{2.186}{2.224104} - \frac{(2.186)^2}{2(2.224104)} + \ln(0.5) = -0.784554$

Label 2:

$g_1(x_2) = \frac{-2.156}{0.802504} - \frac{(-2.156)^2}{2(0.802504)} + \ln(0.5) = -6.27588$

$g_2(x_2) = \frac{-1.72}{1.53568} - \frac{(-1.72)^2}{2(1.53568)} + \ln(0.5) = -2.7764$

b) formula for the line:

$\log\frac{\pi_1}{\pi_2} - \frac{1}{2}(\mu_1+\mu_2)^T \Sigma^{-1}(\mu_1-\mu_2) + x^T\Sigma^{-1}(\mu_1-\mu_2) = 0$

Plug in values: $3.688x_1 + 2.970x_2 - 0.567$