Assignment Sheet 9

$T = \{(0.3, 1), (1.8, 1), (1.5, 1), (4.8, 2), (2.6, 2)\}$

~~For the class label 1,~~

~~$T^1 = \{(0.3, 1), (1.8, 1), (1.5, 1)\}$~~.

Let us build a training set where the output is $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ if the class label is 1 and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ when the class label is 2.

$$T' = \left\{ \left(0.3, \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right), \left(1.8, \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right), \left(1.5, \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right), \left(4.8, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right), \left(2.6, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) \right\}.$$

Then, for the class label 1, we have the following training data,

$$T^1 = \{(0.3, 1), (1.8, 1), (1.5, 1), (4.8, 0), (2.6, 0)\}.$$

Applying linear regression,

$$X = \begin{bmatrix} 1 & 0.3 \\ 1 & 1.8 \\ 1 & 1.5 \\ 1 & 4.8 \\ 1 & 2.6 \end{bmatrix} \qquad y = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$X^T X \beta = X^T y$, we obtain,

$$\beta = \begin{bmatrix} 1.1903 \\ -0.2683 \end{bmatrix}$$

Model :- $1.1903 - 0.2683x$.

For class label 2, we have the following training data,

$T^2 = \{(0.3, 0), (1.8, 0), (1.5, 0), (4.8, 1), (2.6, 1)\}$.

~~X =~~ Here, $y = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$, $X = \begin{bmatrix} 1 & 0.3 \\ 1 & 1.8 \\ 1 & 1.5 \\ 1 & 4.8 \\ 1 & 2.6 \end{bmatrix}$

$X^TX\beta = X^Ty$, we obtain,

$\beta = \begin{bmatrix} -0.1903 \\ 0.2683 \end{bmatrix}$,

Model :- $-0.1903 + 0.2683x$

for $x_1 = 2.4$,

$1.1903 - 0.2683x = 0.54638$

$-0.1903 + 0.2683x = 0.45362$. $(0.54683 > 0.456362)$

So, it belongs to class '1'.

for, $x_2 = 6.2$

$1.1903 - 0.2683x = -0.47361$

$-0.1903 + 0.2693x = 0.47316$.

$0.47816 > -0.47361$

Hence, it belongs to class '2'.

for the training error:-

$f^1(x) = 1.1903 - 0.2683\,x$

$f^2(x) = -1.1903 + 0.2683\,x$

for, $x = 0.3$,

$f^1(x) = 1.10981$, $f^2(x) = -1.10981$

$f(0.3) = 1$.

for, $x = 1.8$

$f^1(x) = \cancel{1.1903 - 0.26}\ 0.70$

$f^2(x) = -0.70736$

$f(1.8) = 1$.

for $x = 1.5$

$f^1(x) = 0.787$

$f^2(x) = -0.787$

$f(1.5) = 1$

for $x = 4.8$

$f^1(x) = \cancel{0.6537}\ -0.09754$

$f^2(x) = \cancel{-0.6537}\ 0.09754$

$\cancel{f(2) =}\ f(4.8) = 2$

for $x = 2.6$

$f^1(x) = 0.49272$

$f^2(x) = -0.49272$ $f(2.6) = 1$.

Here, only 2·6 is classified incorrectly.
Hence training error $= 0+0+0+0+1 = 1.$ //

Excercise 2. (Classification by linear discriminant analysis)

We again start from the training data set

$$\mathcal{T}_{train} = \{(0.3, 1), (1.8, 1), (1.5, 1), (4.8, 2), (2.6, 2)\}$$

for a paper and pencil classification task. In addition, you are given the validation set

$$\mathcal{T}_{val} = \{(1.6, 1), (1.9, 2), (2.5, 2)\} \ .$$

a) Use linear discriminant analysis to build a classifier based on the training data.

b) Evaluate the generalization error for the just constructed predictor using the 0-1 loss and the validation set approach.

(4 Points)

$N_1 = \{(0.3, 1), (1.8, 1), (1.5, 1)\}.$

$N_2 = \{(4.8, 2), (2.6, 2)\}.$

$\hat{P}_G(1) = \frac{3}{5}. \qquad \hat{P}_G(2) = \frac{2}{3}$

$\hat{\mu}_1 = \frac{0.3 + 1.8 + 1.3}{3} = 1.2.$

$\hat{\mu}_2 = \frac{4.8 + 2.6}{2} = 3.7.$

$$\sigma = \frac{1}{3}\left[(0.3-1.2)^2 + (1.8-1.2)^2 + (1.5-1.2)^2 + (4.8-3.7)^2 + (2.6-3.7)^2\right]$$

$$= \frac{1}{3}\left[0.81 + 0.36 + 0.09 + 1.21 + 1.21\right]$$

$$= \frac{1}{3}\left[3.68\right] = 1.2267.$$

So,

$$\hat{G}(x) = \underset{g \in \{0,1\}}{\operatorname{argmax}} \left(\log \hat{P}_G(g) + x^T \Sigma^{-1} \hat{\mu}_g - \frac{1}{2} \hat{\mu}_g^T \Sigma^{-1} \hat{\mu}_g \right)$$

$$\underset{g \in \{0,1\}}{\operatorname{argmax}} \left(\log \hat{P}_G(g) + x^T \cdot \frac{1}{1.2267} \hat{\mu}_g - 0.40759 . \hat{\mu}_g^T \hat{\mu}_g \right)$$

Now, on the validation set..

For $x = 1.9$

$$\hat{G}(1.6) = \underset{g \in \{1,2\}}{\arg\max} \left(\log \hat{P}_G(g) + 1.6 \cdot \frac{1}{1.2267} \cdot \hat{\mu}_g - 0.40759\, \hat{\mu}_g^T \hat{\mu}_g \right)$$

so, for $g = 1 \implies 0.46741$

for $g = 2 \implies -1.6702$

$\therefore \hat{G}(1.6) = 1$

For $x = 1.9$

$$\hat{G}(1.9) = \underset{g \in \{1,2\}}{\arg\max} \left(\log \hat{P}_G(g) + \frac{1.9}{1.2267} \hat{\mu}_g - 0.40759\, \hat{\mu}_g^T \hat{\mu}_g \right)$$

for $g = 1 \implies 0.76088$

$g = 2 \implies -0.76088$

$\therefore \hat{G}(1.9) = 1$

For $x = 2.5$.

$$\hat{G}(2.5) = \underset{g \in \{1,2\}}{\operatorname{argmax}} \left(\log \hat{p}_G(g) + \frac{2.5}{1.2267} \hat{\mu}_g - 0.40759\, \hat{\mu}_g^{\top} \hat{\mu}_g \right)$$

for $g = 1 \Rightarrow 1.34783$.

$g = 2 \Rightarrow 1.04435$

$\therefore \hat{G}(2.5) = 1$

So,

$$GE = \frac{1}{|\mathcal{T}val|} \sum_{(x,y) \in \mathcal{T}val} L_{0\text{-}1}(y, \hat{G}(x))$$

$$= \frac{1}{3}[0 + 1 + 1]$$

$$= 2/3.$$

n ... with nec...

(4 Points)

Excercise 3. (Training of logistic regression)

Prove Lemma 8.2 from the lecture.

(4 Points)

=

Now, the partial derivative of J_β with respect to the q^{th} coefficient of the p^{th} class posterior is.

The original $J_\beta(\beta)$ is.

$$J_\beta(\beta) = -\sum_{i=1}^{N}\left[\sum_{g=1}^{r} p(g|x_i)\,\delta_g(x_i) - \log\left(\sum_{n=1}^{r}\exp(\delta_n(x_i))\right)\right]$$

Instead of this, we can use.

$$J_\beta(\beta) = -\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{g=1}^{r} p(g|x_i)\,\delta_g(x_i) - \log\left(\sum_{n=1}^{r}\exp(\delta_n(x_i))\right)\right].$$

This functi . has no effect on the outcome of the gradient decent. as it only changes the learning rate, which can be easily recoered by using a higher η

$$\frac{dJ_B}{dB_q^p} = -\frac{dJ_B}{dB_q^p}\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{g=1}^{r} p(g|x_i)\cdot s_g(x_i) - \log\left(\sum_{n=1} \exp(s_n(x_i))\right)\right]$$

$$= -\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{g=1}^{r}\frac{dJ_B}{dB_q^p} p(g|x_i)\cdot s_g(x_i) - \frac{1}{\sum_{n=1}^{r}\exp(s_n(x_i))}\cdot\sum_{n=1}^{r}\frac{d}{dB_q^p}\exp(s_n(x_i))\right]$$

$$= -\frac{1}{N}\sum_{i=1}^{N}\left[p(p|x_i)\cdot x_{iq} - \frac{1}{\sum_{n=1}^{r}\exp(s_n(x_i))}\cdot\sum_{n=1}^{r}\exp(s_n(x_i))\cdot\frac{d}{dB_q^p}s_n(x_i)\right]$$

$$= -\frac{1}{N}\sum_{i=1}^{N}\left[p(p|x_i)\,x_{iq} - \frac{1}{\sum_{n=1}^{r}\exp(s_n(x_i))}\exp(s_p(x_i))\cdot x_{iq}\right].$$

$$= -\frac{1}{N}\sum_{i=1}^{N} p(p|x_i)\cdot x_{iq} - p_B(p|x_i)\cdot x_{iq}$$

$$= \frac{1}{N}\sum_{i=1}^{N} p_B(p|x_i)\cdot x_{iq} - p(p|x_i)\cdot x_{iq}$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_{iq}\left(p_B(p|x_i) - p(p|x_i)\right)$$

so,

$$\nabla_g J(\beta) = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} x_{i0}\left(P_B(g|x_i) - p(g|x_i)\right) \\ \frac{1}{N}\sum_{i=1}^{N} x_{i1}\left(P_B(g|x_i) - p(g|x_i)\right) \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N} x_{iD}\left(P_B(g|x_i) - p(g|x_i)\right) \end{bmatrix}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left[\left(P_B(g|x_i) - p(g|x_i)\right) \cdot \begin{pmatrix} x_{i0} \\ x_{i1} \\ \vdots \\ x_{iD} \end{pmatrix}\right]$$

$\because$ we know $x_{i0} = 1$

$$\nabla_g J(\beta) = \frac{1}{N}\sum_{i=1}^{N}\left[\left(P_B(g|x_i) - p(g|x_i)\right)\begin{bmatrix} 1 \\ x_i \end{bmatrix}\right]$$