Shucen Liu
HW3

1.

1.1 (a) $p_1(x) = P(x|Y=1) = \frac{1}{2}\frac{1}{\sqrt{2\pi}}\left[e^{-(x+5)^2/2} + e^{-(x-5)^2/2}\right]$

$p_0(x) = P(x|Y=0) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

Bayes classifier.

$$h^*(x) = \begin{cases} 1 & \text{if } \frac{p_1(x)}{p_0(x)} > \frac{1-\frac{1}{2}}{\frac{1}{2}} \Rightarrow p_1(x) > p_0(x) \\ 0 & \text{otherwise.} \end{cases}$$

$R^* = P(Y \neq h^*(x))$

$= P(h^*(x)=1|Y=0)P(Y=0) + P(h^*(x)=0|Y=1)P(Y=1)$

$= \frac{1}{2}\left[P(h^*(x)=1|Y=0) + P(h^*(x)=0|Y=1)\right]$

$= \frac{1}{2}\left[P(X=h^{*-1}(1)|Y=0) + P(X=h^{*-1}(0)|Y=1)\right]$

where $X|Y=0 \sim N(0,1)$

$X|Y=1 \sim \frac{1}{2}N(-5,1) + \frac{1}{2}N(5,1)$.

(b) Solving the equation $p_1(x) > p_0(x)$,

we get $x < -2.64$ or $x > 2.64$.

Thus, the best linear classifier is

$$h^*(x) = \begin{cases} 1 & \text{if } x \in (-\infty, -2.64) \cup (2.64, \infty) \\ 0 & \text{if } x \in (-2.64, 2.64). \end{cases}$$

Using the property of normal distribution, or the CDF / pdf function of $X|Y=1$, $X|Y=0$, we get the minimized risk:

$$R^* = \int_{-\infty}^{-2.64} f_1(x)dx + \int_{2.64}^{\infty} f_1(x)dx + \frac{1}{2}\int_{-2.64}^{2.64} f_2(x)dx + \frac{1}{2}\int_{-2.64}^{2.64} f_3(x)dx$$

$f_1(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$, $f_2(x) = \frac{e^{-(x+5)^2/2}}{\sqrt{2\pi}}$, $f_3(x) = \frac{e^{-(x-5)^2/2}}{\sqrt{2\pi}}$.

$R^* = 0.0082 + 0.0091 = 0.0173$.

1.2 (a) $$p_1(x) = P(X|Y=1) = \begin{cases} \frac{1}{15} & \text{if } x \in (-5, 10) \\ 0 & \text{otherwise.} \end{cases}$$

$$p_{-1}(x) = P(X|Y=-1) = \begin{cases} \frac{1}{15} & \text{if } x \in (-10, 5) \\ 0 & \text{otherwise.} \end{cases}$$

$R(h^*) = \frac{1}{2}\left[P(X=h^{-1}(-1)|Y=1) + P(X=h^{-1}(1)|Y=-1)\right]$

(b) $h_\alpha(x) = \text{sign}(x-\alpha)$

$R(h^*) = \frac{1}{2}\left[P(X<\alpha|Y=1) + P(X>\alpha|Y=-1)\right]$

by $p_1(x)$, $p_2(x)$, $P(X<\alpha|Y=1) = 0$ for $\alpha < -5$ ~~and α ≥ 0~~

$P(X>\alpha|Y=-1) = 0 \Rightarrow \alpha > 5$

(b) $R(h) = \frac{1}{2}\cdot\left[\frac{\alpha+5}{15} + \frac{5-\alpha}{15}\right]$ if $-5 < \alpha < 5$

$= \frac{1}{2}\cdot\frac{5-\alpha}{15}$ if $-10 < \alpha < -5$

$= \frac{1}{2}$ if $\alpha < -10$

$= \frac{1}{2}\cdot\frac{\alpha+5}{15}$ if $5 < \alpha < 10$

$= \frac{1}{2}$ if $\alpha > 10$

Thus, $R(h)$ is minimized when $\alpha = \pm 5$

$R(h) = \frac{1}{3}$

(c) $E(1-Y\beta X) = E[E(1-Y\beta X|Y)]$

$\cancel{= E[(1-\beta X)]}$

$= E[(1-Y\beta X)|Y=1]\cdot P(Y=1) + E[(1-Y\beta X)|Y=-1]\cdot P[Y=-1]$

$= \frac{1}{2}\left[E((1-\beta X)|Y=1) + E[(1+\beta X)|Y=-1]\right]$

$= \frac{1}{2}[1-\beta\ E(X|Y=1) + 1 + \beta\cdot E(X|Y=-1)]$

$= \frac{1}{2}\cdot[1-2.5\beta + 1 + \beta\cdot(-2.5)]$

$= 1 - 5\beta/2 \quad = 1 - 2.5\cdot\beta$

the positive part is 1

2. (a) In the logistic regression model, $y_i \in \{0,1\}$.

$\cancel{p_1(x_i;\theta) = p(x}$ $\quad p_1 + p_0 = 1$

the log likelihood:

$$\ell(\beta) = \sum_{i=1}^{N}\{y_i \ln p(x_i;\beta) + (1-y_i)\ln(1-p(x_i;\beta))\}$$

$$= \sum_{i=1}^{N}\{y_i\beta^T x_i - \ln(1+\exp(\beta^T x_i))\}$$

Solve the equation using Newton's method.

$$\beta \leftarrow \beta - \left(\frac{\partial^2\ell(\beta)}{\partial\beta\partial\beta^T}\right)^{-1}\cdot\frac{\partial\ell(\beta)}{\partial\beta}.$$

For log likelihood of logistic regression,

$$\frac{\partial^2\ell(\beta)}{\partial\beta\partial\beta^T} = -\sum_{i=1}^{N} x_i x_i^T p(x_i;\beta)(1-p(x_i;\beta))$$

let $\vec{p}$ be the vector of our probabilities,

$\beta \leftarrow \beta - (X^TWX)^{-1}X^T(\vec{y}-\vec{p})$.

with W be a diagnol weight matrix.

$W_i = \pi_i(x_i)(1-\pi_i(x_i))$.

then we can rearrange the Newton step as a weighted least squares step: $\beta^{new} = (X^TWX)^{-1}X^TW\vec{z}$, with $\vec{z} = X\beta^{old} + W^{-1}(\vec{y}-\vec{p})$.

Shueen Liu
homework 3

2.(b) $\max_{\beta} \sum_{i=1}^{n} \log p(y_i | x_i, \beta)$

when the traing data is perfectly separable, there are infinitely many MLE's. as any step function that is in the gap between the two classes is an MLE.

The IRLS algorithm will not converge.

2.(c) $p(x_i) = \exp(x_i\beta)/\{1+\exp(x_i\beta)\}$.

the log likelihood is

$$\ell(\beta) = \sum_i [y_i \log p(x_i) + (1-y_i)\cdot\log(1-p(x_i))]$$

with penalty:

$$\ell^{\lambda}(\beta) = \ell(\beta) - \lambda\|\beta\|^2$$

$$\frac{\partial \ell^{\lambda}(\beta)}{\partial \beta} = \frac{\partial \ell(\beta)}{\partial \beta} - 2\lambda\beta$$

$$H(t,\beta)^{-1} = \frac{\partial \ell^{\lambda}(\beta)}{\partial\beta\partial\beta^T} = \frac{\partial \ell(\beta)}{\partial\beta\partial\beta^T} - 2\lambda[1,1,\ldots,1]$$

$$\beta \leftarrow \beta - \left(\frac{\partial^2 \ell(\beta)}{\partial\beta\partial\beta^T} - 2\lambda\right)^{-1}\left(\frac{\partial \ell(\beta)}{\partial\beta} - 2\lambda\beta\right)$$

where $\frac{\partial^2 \ell(\beta)}{\partial\beta\partial\beta^T} = -\sum_{i=1}^{N} x_i x_i^T p(x_i)(1-p(x_i))$

and $\frac{\partial \ell(\beta)}{\partial\beta} = \sum_{i=1}^{N}\left[y_i x_i - \frac{x_i \exp(\beta^T x_i)}{1+\exp(\beta^T x_i)}\right]$