Homework 2

$\boxed{\text{DATA 558}}$

Carlin Charpentier

4 a. $Y = \beta_0 + \beta_1 X_1 + \epsilon \implies Y = -165.1 + 4.8(64) = \boxed{142}$

b. OLS: $\hat{\beta} = (x^T x)^{-1} x^T y$ If each height in the training set is divided by 12

$\boxed{\begin{array}{l} \beta_0 : -165.1 \\ \beta_1 : 4.8(12) = 57.6 \end{array}}$

since β_0, the y-intercept, is the value of the predictor when x is 0, it will remain unchanged as x is scaled

As x is scaled down, its coefficient, β_1, will be scaled up to compensate

c. Given $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$

Suppose $x \in R^{n \times 2}$ and $x_{i_2} = 12 x_{i_1}$ i.e. $x = \begin{bmatrix} x_{11} & 12 \cdot x_{11} \\ x_{21} & 12 \cdot x_{21} \\ \vdots & \vdots \\ x_{n1} & 12 x_{n1} \end{bmatrix}$

$y \in R^{n \times 1}$

$x^T y \to (2 \times n) \cdot (n \times 1) = 2 \times 1$

$\boxed{\begin{array}{l} \beta_0 = -165.1 \\ \beta_1 = 4.8 \\ \beta_2 = 57.6 \end{array}}$ (one possible solution)

β_1 and β_2 are perfectly correlated, so the addition of either doesn't add any new information to the model

$$\hat{\beta} = (x^T x)^{-1} x^T y$$

Not that the input matrix will not be linearly independent, so a pseudo inverse will be required, to find a single solution.

rank 1

d. Training error will be the same for all of them since they are all equivalent models.

---

5 a. $p = 1$ $X \in \mathbb{R}$ $Y \in \{1, 2\}$

Class $1 \sim N(\mu, \sigma^2)$

class $2 \sim \text{Unif}[-2, 2]$

Derive the expression of the Bayes decision boundary:

x s.t. $P(Y=1 | X=x) = P(Y=2 | X=x)$

$$P(X=x | Y=1) = f_1(x)$$

$$P(Y=1 | X=x) = P(Y=2 | X=x) \implies \frac{\pi_1 f_1(x)}{\cancel{\pi_1 f_1(x) + \pi_2 f_2(x)}} = \frac{\pi_2 f_2(x)}{\cancel{\pi_1 f_1(x) + \pi_2 f_2(x)}}$$

$$\implies \pi_1 f_1(x) = \pi_2 f_2(x)$$

$$-2 \le x \le 2 : \pi_1 \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(\frac{-1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}\right) = \frac{1}{4}\pi_2$$

$$\implies \exp\left(\frac{-1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}\right) = \frac{\pi_2 \sigma \sqrt{2\pi}}{4\pi_1}$$

↓ [Next page] ↓

$$\Rightarrow -\frac{1}{2}\cdot\frac{(x-\mu)^2}{\sigma^2} = \ln\left(\frac{\pi_2\sigma\sqrt{2\pi}}{4\pi_1}\right)$$

$$\Rightarrow x = \sqrt{-2\sigma\ln\left(\frac{\pi_2\sigma\sqrt{2\pi}}{4\pi_1}\right)} + \mu$$

$x<-2, x>2$: undefined since we'd have $\ln(0)$

b. Assume $\mu=0, \sigma=1, \pi_1=0.45 \Rightarrow \pi_2=0.55$

$$x = \sqrt{-2(1)\cdot\ln\left(\frac{0.55(1)(\sqrt{2\pi})}{4(0.45)}\right)} + 0 \approx 1.93$$

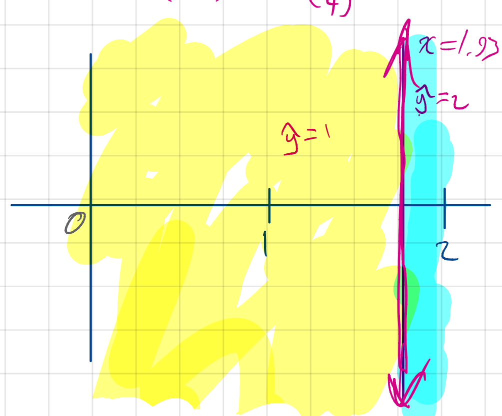
Now lets see which class yields a higher probability for $x=1$

$$f_1(1) = \frac{1}{1\sqrt{2\pi}}\cdot\exp\left(\frac{(1-0)^2}{2}\right) = \frac{1}{\sqrt{2\pi}}\cdot\exp\left(\frac{1}{2}\right) \approx 0.66$$

$$P(Y=1|X=1) = \frac{\pi_1 f_1(1)}{\pi_1 f_1(1)+\pi_2 f_2(1)} = \frac{0.45(0.66)}{0.45(0.66)+0.55\left(\frac{1}{4}\right)} = 0.68$$

$$p(Y=2|X=1) = \frac{\pi_2 f_2(1)}{\pi_1 f_1(1)+\pi_2 f_2(1)} = \frac{0.55\left(\frac{1}{4}\right)}{0.45(0.66)+0.55\left(\frac{1}{4}\right)} = 0.316$$

$$\Rightarrow \boxed{\begin{array}{l} 0\le x<1.93 \rightarrow 1=\hat{y} \\ 1.93\le x\le 2 \rightarrow 2=\hat{y}\end{array}}$$



C. Observe n training observations $(x_1,y_1),\dots,(x_n,y_n)$

Q: How to use these values to estimate μ, σ, π_1?

Let $S=\{(x_1,y_1),\dots,(x_n,y_n)\}$, $T_k=\{(x_i,y_i): y_i=k, (x_i,y_i)\in S\}$ $k\in\{1,2\}$
card(·) be the cardinality of a set

$$\pi_1 = \frac{\text{card}(T_1)}{\text{card}(S)} \qquad \sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2}, \quad x_i \text{ from } (x,y)\in T_1$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad x \text{ from } (x,y)\in T_1$$

d. Given $X = x_0$, estimate $P(Y=1 | X=x_0)$

Let $f_1(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}\right)$

$\pi_1, \mu, \sigma \equiv$ their values from problem (5c)

$\pi_2 = 1 - \pi_1$

$$\boxed{P(Y=1 | X=x_0) = \frac{\pi_1 f_1(x_0)}{\pi_1 f_1(x_0) + \pi_2 f_2(x_0)}}$$

(no unknowns)

6

a. Logistic Regression: observe $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$ log-odds $= 0.7$

Find $P(Y=1 | X=x)$

$P(Y=1 | X=x)$

$\log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 X_1 + \dots \beta_p X_p = 0.7$

$\Rightarrow \frac{p(x)}{1-p(x)} = \exp(0.7)$

$\Rightarrow p(x) = \frac{\exp(0.7)}{1+\exp(0.7)}$

$$\Rightarrow \boxed{P(Y=1 | X=x) = \frac{\exp(0.7)}{1+\exp(0.7)} \text{ for } x = (x_1, \dots, x_p)^T}$$

b. $x^* = (x_1+1, x_2-1, x_3+2, x_4 \dots, x_p)^T$ Find $P(Y=1 | X=x^*)$

$\log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1(x_1+1) + \beta_2(x_2-1) + \beta_3(x_3+2) + \dots + \beta_p x_p - (\beta_1 - \beta_2 + 2\beta_3) = 0.7$

$\Rightarrow \log\left(\frac{P(x^*)}{1-P(x^*)}\right) = 0.7 + \beta_1 - \beta_2 + 2\beta_3$

$$\Rightarrow \boxed{P(Y=1 | X=x^*) = \frac{\exp(0.7+\beta_1-\beta_2+2\beta_3)}{1+\exp(0.7+\beta_1-\beta_2+2\beta_3}}$$

For problems 7 and 8, see code

(7) b. Calculate the Bayes' decision boundaries

$$P(Y=k|X=x) = \frac{P(X=x|Y=k)\,P(Y)}{P(X)} \qquad P(Y)=\frac{1}{3} \qquad P(x)=?$$

Let $\Sigma_k = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}$ for $k=1,2,3$

$\mu_1 = (-2,-2)^T$, $\mu_2 = (0,0)^T$, $\mu_3 = (2,2)^T$

k=1, k=2 Decision Boundary:

$$\frac{P(X=x|Y=1)\cdot\cancel{\frac{1}{3}}}{\cancel{P(x)}} = \frac{P(X=x|Y=2)\cdot\cancel{\frac{1}{3}}}{\cancel{P(x)}}$$

$$\Rightarrow P(X=x|Y=1) = P(X=x|Y=2)$$

$$\Rightarrow \cancel{\det(2\pi\Sigma)^{-\frac{1}{2}}}\exp\left(-\tfrac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right) = \cancel{\det(2\pi\Sigma)^{-\frac{1}{2}}}\exp\left(-\tfrac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)\right)$$

$$\Rightarrow \cancel{-\tfrac{1}{2}}\overset{1\times2}{(x-\mu_1)^T}\,\overset{2\times2}{\cancel{\Sigma^{-1}}}\,\overset{2\times1}{(x-\mu_1)} = \cancel{-\tfrac{1}{2}}(x-\mu_2)^T\cancel{\Sigma^{-1}}(x-\mu_2)$$

$$\Rightarrow (x-\mu_1)^T(x-\mu_1) = (x-\mu_2)^T(x-\mu_2)$$

$$(a \quad b)\begin{pmatrix} c \\ d \end{pmatrix} = ac+bd$$

$$(x_1-\mu_{11} \quad x_2-\mu_{12})\begin{pmatrix} x_1-\mu_{12} \\ x_2-\mu_{12} \end{pmatrix}$$

$$\Rightarrow (x_1+2)^2 + (x_2+2)^2 = (x_1-0)^2 + (x_2-0)^2$$

$$\Rightarrow \cancel{x_1^2}+4x_1+4+\cancel{x_2^2}+4x_2+4 = \cancel{x_1^2}+\cancel{x_2^2}$$

$$\Rightarrow 4x_1+4x_2 = -8$$

$$\Rightarrow \boxed{x_2 = -x_1-2}$$

k=1, k=3 Boundary

we can start here

$$(x-\mu_1)^T(x-\mu_1) = (x-\mu_3)^T(x-\mu_3)$$

$$\Rightarrow (x_1+2)^2+(x_2+2)^2 = (x_1-2)^2+(x_2-2)^2$$

$$\Rightarrow \cancel{x_1^2}+4x_1+\cancel{4}+\cancel{x_2^2}+4x_2+\cancel{4} = \cancel{x_1^2}-4x_1+\cancel{4}+\cancel{x_2^2}-4x_2+\cancel{4}$$

$$\Rightarrow \boxed{x_2 = -x_1}$$

k=2, k=3 Boundary

$$(x-\mu_2)^T(x-\mu_2) = (x-\mu_3)^T(x-\mu_3)$$

$$\Rightarrow (x_1+0)^2+(x_2+0)^2 = (x_1-2)^2+(x_2-2)^2$$

$$\Rightarrow \cancel{x_1^2}+\cancel{x_2^2} = \cancel{x_1^2}-4x_1+4+\cancel{x_2^2}-4x_2+4$$

$$\Rightarrow -4x_1-4x_2+8=0$$

$$\Rightarrow \boxed{x_2 = -x_1+2}$$

4) Derive an expression for the ridge regression estimator

Let $y \in \mathbb{R}^{n\times 1}$, $x \in \mathbb{R}^{n\times p}$, $\beta \in \mathbb{R}^{p\times 1}$, $R(\beta, y, x)$ be the ridge regression estimator

$$\mathrm{Loss}(x, y, \beta) = \underbrace{\|y - x\beta\|^2}_{\text{2-norm}} + \underbrace{\lambda \beta^T \beta}_{\text{this is a constant}}$$

$$= (y - x\beta)^T (y - x\beta) + \lambda \beta^T \beta$$

$$= y^T y - y^T x \beta - \beta^T x^T y + \beta^T x^T x \beta + \lambda \beta^T \beta$$

$$\frac{\partial \mathrm{Loss}}{\partial \beta} = 0 - y^T x - y^T x + 2 x^T x \beta + 2 \lambda \beta$$

$$= -2 y^T x + 2 x^T x \beta + 2 \lambda \beta$$

Now set to 0 and solve for β:

$$2 x^T x \beta + 2 \lambda \beta = 2 y^T x$$

$$\Rightarrow (\cancel{2} x^T x + \cancel{2} \lambda) \beta = \cancel{2} y^T x$$

$$\Rightarrow \boxed{\beta = (x^T x + \lambda)^{-1} y^T x} \equiv R(\beta, y, x)$$