HW1

106064701

蔡昀霖

1.

(a) 令 $X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}$, $\bar{x} = \frac{x_1+x_2}{2}$.

$\Rightarrow (X^T X)^{-1} = \frac{1}{(x_2-x_1)^2} \cdot \begin{bmatrix} x_1^2+x_2^2 & x_1+x_2 \\ x_1+x_2 & 2 \end{bmatrix}$ 左下右上都少一個負號

$\Rightarrow SE(\hat{\beta}_j)^2 = \left[(X^T X)^{-1} \cdot \sigma^2 \right]_{jj}$

$$SE(\hat{\beta}_1)^2 = \frac{2\sigma^2}{(x_2-x_1)^2} = \frac{\sigma^2}{\frac{2(x_1-x_2)^2}{4}} = \frac{\sigma^2}{\frac{(x_1-x_2)^2}{4} + \frac{(x_2-x_1)^2}{4}}$$

$$= \frac{\sigma^2}{(x_1-\bar{x})^2+(x_2-\bar{x})^2}$$

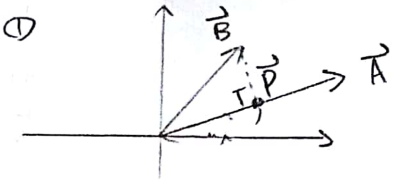
$$= \frac{\sigma^2}{\sum_{i=1}^{2} (x_i-\bar{x})^2} \quad \#$$

$$SE(\hat{\beta}_0)^2 = \frac{\sigma^2(x_1^2+x_2^2)}{(x_2-x_1)^2}$$

$$= \sigma^2 \left[\frac{\frac{(x_1+x_2)^2}{2} + \frac{(x_1-x_2)^2}{2}}{(x_1-x_2)^2} \right]$$

$$= \sigma^2 \cdot \left[\frac{1}{2} + \frac{\bar{x}}{\sum_{i=1}^{2}(x_i-\bar{x})^2} \right] \quad \#$$

1(b) ①



$$A^T\cdot(b - A\hat{B}) = 0$$
$$A^TA\hat{B} = A^Tb$$
$$\hat{B} = (A^TA)^{-1}A^Tb$$

$$p = A\hat{B} = A(A^TA)^{-1}A^Tb$$

projection matrix: $A(A^TA)^{-1}A^T$

where b is data Y, A is data X.

then we can find projection matrix H

$$H = X(X^TX)^{-1}X^T \quad \#$$

② proof that

$HZ = Z$, where Z is in subspace of X.

First, we know that it exists a vector V whose projection on X is Z, i.e. $HV = Z$

$$\Rightarrow HZ = H(HV)$$
$$= \left[X(X^TX)^{-1}X^T\right]\left[X(X^TX)^{-1}X^T V\right]$$
$$= X\cdot I\cdot(X^TX)^{-1}X^TV$$
$$= X(X^TX)^{-1}X^TV = HV = Z \quad \#$$

2.(a) MLE 由題 $\varepsilon \sim N(0,\sigma^2)$，那麼 y 為 Normal Distribution 的線性組合

所以 $Y \sim N(X^TB, \sigma^2)$

$\rightarrow$ likelihood function $L(B) = (2\pi\sigma^2)^{-\frac{N}{2}}\cdot\exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{N}(y_i - x_i^TB)^2\right) = \prod_{i=1}^{N} f_Y(y_i|X;B)$

$\rightarrow$ 取 $\ln$ $\rightarrow$ $\ln(L(B)) = \sum_{i=1}^{N}\ln(f_Y(y_i|X;B)) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\cdot\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - x_i^TB)^2$

找 max $\rightarrow$ $\frac{\partial \ln(L(B))}{\partial B} = \frac{1}{\sigma^2}\sum_{i=1}^{N}x_i^T(y_i - x_iB) = \frac{1}{\sigma^2}\left(\sum_{i=1}^{N}x_i^Ty_i - \sum_{i=1}^{N}x_i^Tx_iB\right) = 0$

$\rightarrow B = \left(\sum_{i=1}^{N}x_i^Tx_i\right)^{-1}\cdot\left(\sum_{i=1}^{N}x_i^Ty_i\right) = (X^TX)^{-1}X^Ty$ #

(b)

$$P(B|y;X) = \frac{P(y|B;X)P(B)}{P(y)} \Rightarrow \underset{B}{\operatorname{argmax}}\frac{\prod_{i=1}^{N}P(y_i|B;X)P(B)}{P(y)}$$

$\xrightarrow[\text{且 } P(y) = \text{Constant}]{\text{取 } \ln}$ $\underset{B}{\operatorname{argmax}}\frac{\sum_{i=1}^{N}\left[\ln(P(y_i|B;X)) + \ln(P(B))\right] + \ln P(B)}{C}$, where $P(B) = (2\pi\gamma^2)^{-0.5}\cdot\exp\left(\frac{-1}{2}\cdot\left(\frac{B}{\gamma}\right)^2\right)$

$P(y_i|B;x) = (2\pi\sigma^2)^{-0.5}\cdot\exp\left(\frac{-1}{2}\left(\frac{y_i - x_i^TB}{\sigma}\right)^2\right)$

$$\frac{d}{dB} \rightarrow \sum_{i=1}^{N}\left(\frac{y_i - x_i^TB}{\sigma^2}\right) - \left(\frac{B}{\gamma^2}\right) = 0$$

$$\rightarrow \sum_{i=1}^{N}\left(\frac{y_i}{\sigma^2}\right) = \frac{B}{\gamma^2} + \frac{B}{\sigma^2}\cdot\sum_{i=1}^{N}x_i^T$$

$$\rightarrow B = \frac{\sum_{i=1}^{N}\left(\frac{y_i}{\sigma^2}\right)}{\frac{1}{\gamma^2} + \frac{1}{\sigma^2}\cdot\sum_{i=1}^{N}x_i^T} \quad \#$$

(c) 分點論述.

① 最小平方法是利用最小化 L_2-norm

② MLE 是由機率觀點出發，求取一組參數 B 使 $P(D|B)$ 最大.

③ 當觀察值來自指數族且滿足特定條件時，最小平方法和 MLE 相同.

④ MLE 本身認定參數的機率分布是均勻的 (常數)

⑤ MLE 是求取 B 使 $P(D|B)$ max，而 MAP 是求 B 使 $P(D|B)P(B)$ max，求得 B 不單單使 likelihood function max，也使 $P(B)$ 最大，這有點像正規化的概念，只不過一般正規化是加 P_λ，而這裡使用乘法 (正規化可用於避免 overfitting) #

(a) ① 依題目要求，分別家取的佔比為：$\frac{31}{200}, \frac{33}{200}, \frac{68}{200}, \frac{121}{200}, \frac{144}{200}$ 數量的樣本，並把樣本合為一了 $X = 396\times1$，$y = 396\times1$ 的矩陣。

② $\hat{\beta} = (X^TX)^{-1}X^Ty = \begin{bmatrix} 2.2817 \\ 0.032 \end{bmatrix}$ ← β_0, ← β_1 #

$\begin{bmatrix} 1 & \text{折扣} \\ \vdots & \vdots \\ 1 & \vdots \end{bmatrix}_{396\times2}$ $\begin{bmatrix} 1 \\ 0 \\ 1 \\ \vdots \end{bmatrix}$

③ $25\% \Rightarrow 0.09 + 1.71\times0.25 = 0.502$

④ $10\% \Rightarrow 0.26$，$15\% \Rightarrow 0.37$，$20\% \Rightarrow 0.48$，$30\% \Rightarrow 0.71$

→ 各折扣所對應的兌換率 #

(b) ① 第一步如 (a) -① 所述。

② 迭代所有 Data：$B(k+1) = B(k) - \eta \cdot \nabla J(B(k))$，where. $\nabla J(B(k)) = -\sum_{i=1}^{N} \frac{(2y_i-1)x_i}{1+e^{(2y_i-1)B(k)^Tx_i}}$

$\eta = $ 定 $= 0.01$

③ 當 $B(k+1)$ 不再改變，即可停止。⇒ 即可估測 25% 的兌換率 #

(c) 令 $k=1 \Rightarrow y_i=1$，$k=2 \Rightarrow y_i=0$

$\hat{\mu}_1 = \frac{1}{396}\times(0.05\times31 + 0.1\times53 + 0.15\times68 + 0.2\times121 + 0.3\times144) = 0.202$

$\hat{\mu}_2 = \frac{1}{604}\times(0.05\times169 + 0.1\times148 + 0.15\times132 + 0.2\times99 + 0.3\times56) = 0.132$

$\hat{\sigma} = \frac{1}{1000-2}\times\left(\sum_{i:y_i=1}(x_i-0.202)^2 + \sum_{i:y_i=0}(x_i-0.132)^2\right) = 0.006$

$\hat{\pi}_1 = \frac{396}{1000}$

$\hat{\pi}_2 = \frac{604}{1000}$

$$\delta_k(x) = x\cdot\frac{\hat{\mu}_k}{\hat{\sigma}^2} - \frac{\hat{\mu}_k^2}{2\hat{\sigma}^2} + \log(\hat{\pi}_k)$$

$10\% = \delta_1(0.1) = 0.235$

$15\% = \delta_1(0.15) = 0.35$

$20\% = \delta_1(0.2) = 0.48$

$25\% = \delta_1(0.25) = 0.62$

$30\% = \delta_1(0.3) = 0.74$

→ 各折扣所對應的兌換率 #

date: / /

4. $R^2 = (TSS - RSS)/TSS$ "in the case of simple linear Regression"

$$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i)^2$$

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{N}(y_i - (B_0 + B_1 x_i))^2$$

$B_0 = 0$

$$= \sum_{i=1}^{N}\left(y_i - \underbrace{\left(\frac{\sum_{j=1}^{n} x_j y_j}{\sum_{k=1}^{n} x_k^2}\right)}_{B_1} x_i\right)^2$$

this not zero

$$TSS - RSS = \sum_{i=1}^{n}(y_i)^2 - \sum_{i=1}^{N}(y_i^2 - 2y_i B_1 x_i + (B_1 x_i)^2)$$

$$= \sum_{i=1}^{N} 2x_i y_i B_1 - \frac{(\sum_{j=1}^{n} x_j y_j)^2}{\sum_{k=1}^{n} x_k^2} = \frac{(\sum_{j=1}^{n} x_j y_j)^2}{\sum_{k=1}^{n} x_k^2}$$

$$\frac{TSS - RSS}{TSS} = \frac{(\sum_{j=1}^{n} x_j y_j)^2}{\sum_{i=1}^{n}(y_i)^2 \times \sum_{k=1}^{n}(x_k)^2} = (\mathrm{Cor}(X,Y))^2 \#$$

5. (a) uniform $X: [0,1]$, $f_X(x) = \frac{1}{1-0}$

$\Rightarrow (0.65 - 0.55) \times \frac{1}{1-0} = 0.1$ #

(b) X_1 : uniform. $[0,1]$
X_2 : uniform. $[0,1]$ $\Rightarrow (0.1)^2$ #

(c) 0.1^{100} 極小. #

(e) $p=1 \Rightarrow 0.1$

$p=2 \Rightarrow (0.1)^{\frac{1}{2}} \approx 0.316$

$p=100 \Rightarrow (0.1)^{\frac{1}{100}} \approx 0.977$

也就是說當我特徵數越高時,在使用固定(10%)的觀察量時,
我們更越需要包含每個特徵的所有範圍 #