CSCI3230 / ESTR3108 2021-22 First Term Assignment 1

I declare that the assignment here submitted is original except for source material explicitly acknowledged, and that the same or closely related material has not been previously submitted for another course. I also acknowledge that I am aware of University policy and regulations on honesty in academic work, and of the disciplinary guidelines and procedures applicable to breaches of such policy and regulations, as contained in the following websites.

University Guideline on Academic Honesty:
`http://www.cuhk.edu.hk/policy/academichonesty/`

Faculty of Engineering Guidelines to Academic Honesty:
`http://www.erg.cuhk.edu.hk/erg-intra/upload/documents/ENGG_Discipline.pdf`

Student Name: ***Lai Man Hin***

Student ID : ***1155136167***

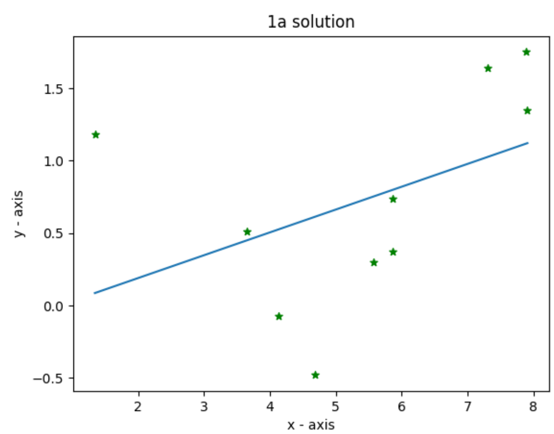
1a)

By $\Theta = (X^T X)^{-1} X^T Y$

$$\Theta = \begin{bmatrix} -0.12471382 \\ 0.15748272 \end{bmatrix}$$

which $\theta_0 = -0.12471382, \theta_1 = 0.15748272$

Figure is shown at right hand side.



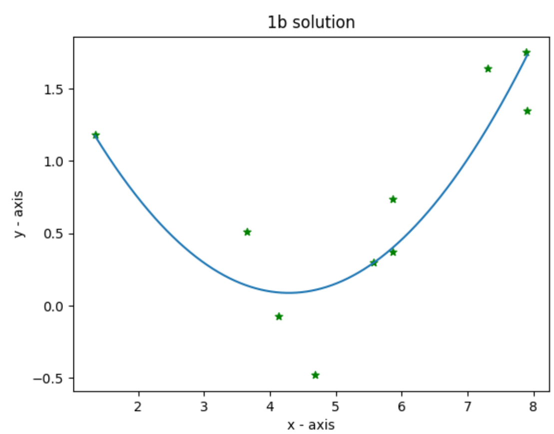
1b)

By $\Theta = (X^T X)^{-1} X^T Y$

$$\Theta = \begin{bmatrix} 2.38902063 \\ -1.07370721 \\ 0.1252789 \end{bmatrix}$$

which $\theta_0 = 2.38902063, \theta_1 = -1.07370721,$
$\theta_2 = 0.1252789$

Figure is shown at right hand side.



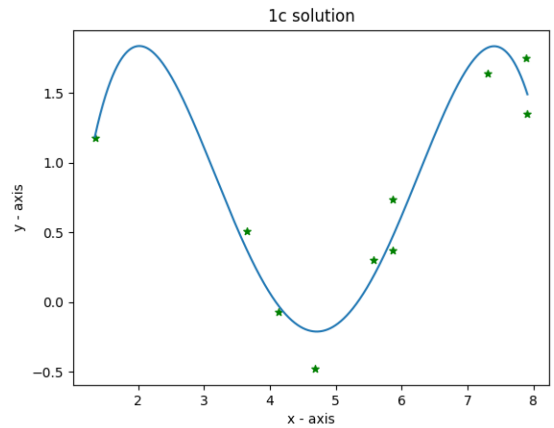
1c)

By $\Theta = (X^T X)^{-1} X^T Y$

$$\Theta = \begin{bmatrix} -6.80776406 \\ 10.91680274 \\ -4.60513794 \\ 0.73170903 \\ -0.03884038 \end{bmatrix}$$

which $\theta_0 = -6.80776406, \theta_1 = 10.91680274,$
$\theta_2 = -4.60513794, \theta_3 = 0.73170903,$
$\theta_4 = -0.03884038$

Figure is shown at right hand side.



1d)

By

$$EPE(X) = E((y - \hat{f}(X))^2)$$

With summarising values,
In 1a) , $EPE(X) = 9.415644$
In 1b) , $EPE(X) = 5.013401$
In 1c) , $EPE(X) = 19.504463$

Therefore,
1a) is underfitting,
1b) is relatively a good one,
1c) is overfitting.

2a)

By differentiating towards θ_0,

$$\begin{aligned}
\frac{\partial J(\Theta, \theta_0)}{\partial \theta_0} &= \sum_{i=1}^{m} 2(X^{(i)} - \bar{X})\Theta + \theta_0 + \bar{X}\Theta - Y^{(i)} \\
&= \sum_{i=1}^{m} 2X^{(i)}\Theta - \bar{X}\Theta + \theta_0 - Y^{(i)} \\
&= 2m\bar{X}\Theta - m\bar{X}\Theta + m\theta_0 - m\bar{Y} \\
&= m\bar{X}\Theta + m\theta_0 - m\bar{Y}
\end{aligned}$$

By letting $\dfrac{\partial J(\Theta, \theta_0)}{\partial \theta_0} = 0$,

$$\begin{aligned}
0 &= m\bar{X}\Theta + m\theta_0 - m\bar{Y} \\
0 &= \bar{X}\Theta + \theta_0 - \bar{Y} \\
\theta_0 &= \bar{Y} - \bar{X}\Theta
\end{aligned}$$

Therefore, when $\theta_0 = \bar{Y} - \bar{X}\Theta$, we will get the minimum value for $J(\Theta, \theta_0)$.

2b)

Considering $Y^{(i)} = Y_c^{(i)} + \bar{Y}$ and plugging in θ_0,

$$\begin{aligned} J(\Theta, \theta_0) = J_c(\Theta) &= \sum_{i=1}^{m} (X_c^{(i)}\Theta + \bar{Y} - \bar{X}\Theta + \bar{X}\Theta - Y_c^{(i)} - \bar{Y})^2 + \lambda \|\Theta\|_2^2 \\ &= \sum_{i=1}^{m} (X_c^{(i)}\Theta - Y_c^{(i)})^2 + \lambda \|\Theta\|_2^2 \end{aligned}$$

2c)

$$J_c(\Theta) = \sum_{i=1}^{m} (X_c^{(i)}\Theta - Y_c^{(i)})^2 + \lambda \|\Theta\|_2^2$$

By considering the minimum value of this formula can be obtained by the sum of each minimum part of summation, so I will remove the summation sign here and then take derivatives,

$$\begin{aligned} J_c'(\Theta) &= \frac{\partial}{\partial \Theta}((X_c\Theta - Y_c)^T (X_c\Theta - Y_c) + \lambda \Theta^T \Theta) \\ &= \frac{\partial}{\partial \Theta}(\Theta^T X_c^T X_c \Theta - \Theta^T X_c^T Y_c - Y_c^T X_c \Theta + Y_c^T Y_c + \lambda \Theta^T \Theta) \\ &= 2X_c^T X_c \Theta - X_c^T Y_c - X_c^T Y_c + 2\lambda\Theta \\ &= 2X_c^T X_c \Theta - 2X_c^T Y_c + 2\lambda\Theta \end{aligned}$$

By taking $J_c' = 0$,

$$\begin{aligned} 0 &= 2X_c^T X_c \Theta - 2X_c^T Y_c + 2\lambda\Theta \\ 2X_c^T Y_c &= 2X_c^T X_c \Theta + 2\lambda\Theta \\ X_c^T Y_c &= (X_c^T X_c + \lambda I)\Theta \\ \Theta &= (X_c^T X_c + \lambda I)^{-1} X_c^T Y_c \end{aligned}$$

Therefore, the analytic solution $\hat{\Theta} = (X_c^T X_c + \lambda I)^{-1} X_c^T Y_c$.

3a)

Logistic model

$$P(\hat{y}=1 \mid x_1, x_2) = \frac{1}{1+e^{-(\theta_0+\theta_1 x_1+\theta_2 x_2)}}$$

Cross-entropy function

$$-y^{(i)} \ln(P(\hat{y}=1 \mid x_1, x_2)) - (1-y^{(i)}) \ln(1-P(\hat{y}=1 \mid x_1, x_2))$$

Plug-in logistic model,

$$-y^{(i)} \ln(\frac{1}{1+e^{-(\theta_0+\theta_1 x_1+\theta_2 x_2)}}) - (1-y^{(i)}) \ln(1-\frac{1}{1+e^{-(\theta_0+\theta_1 x_1+\theta_2 x_2)}})$$

3b) New regression model $\hat{\Theta}$

$$\begin{aligned}
\hat{\Theta} &= \Theta - \alpha \sum_{i=1}^{m} \left(\frac{1}{e^{-X^{(i)T}\Theta}} - y^{(i)} \right) X^{(i)} \\
&= \begin{bmatrix} -1 \\ 1.5 \\ 0.5 \end{bmatrix} - 0.1 \Bigg(\begin{bmatrix} 0.47726569 \\ 0.16513393 \\ 0.37226724 \end{bmatrix} + \begin{bmatrix} 0.4192142 \\ 0.1270219 \\ 0.18403503 \end{bmatrix} + \begin{bmatrix} 0.47539489 \\ 0.17019137 \\ 0.34656288 \end{bmatrix} \\
&+ \begin{bmatrix} 1.39724159 \\ 0.84113944 \\ 1.20581949 \end{bmatrix} + \begin{bmatrix} 1.75330048 \\ 1.38510738 \\ 1.32023526 \end{bmatrix} + \begin{bmatrix} -1.01899756 \\ 1.53210518 \\ 0.51181202 \end{bmatrix} \Bigg) \\
&= \begin{bmatrix} -1.01899756 \\ 1.53210518 \\ 0.51181202 \end{bmatrix}
\end{aligned}$$

Therefore, the updated function is

$$P(\hat{y}=1 \mid x_1, x_2) = \frac{1}{1+e^{-(-1.01899756+1.53210518 x_1+0.51181202 x_2)}}$$

3c)

Index 1 (y = 0): $P(\hat{y}=1 \mid x_1, x_2) = 0.656064168$, FP
Index 2 (y = 0): $P(\hat{y}=1 \mid x_1, x_2) = 0.571187019$, FP
Index 3 (y = 0): $P(\hat{y}=1 \mid x_1, x_2) = 0.493790599$, TN
Index 4 (y = 1): $P(\hat{y}=1 \mid x_1, x_2) = 0.653157925$, TP
Index 5 (y = 1): $P(\hat{y}=1 \mid x_1, x_2) = 0.644341574$, TP
Index 6 (y = 1): $P(\hat{y}=1 \mid x_1, x_2) = 0.621692311$, TP
Therefore, we have 3 TP, 1 TN, 2 FP, 0 FN

Accuracy $= \frac{3+1}{3+1+2+0} = \frac{2}{3}$
Precision $= \frac{3}{3+2} = \frac{3}{5}$
Recall $= \frac{3}{3+0} = 1$

4a)
I will find the partial derivatives of

$$P(\hat{y}_i = 1 \mid X) = \frac{e^{X^T\Theta_i}}{\sum_{k=1}^{K} e^{X^T\Theta_k}}$$

by quotient rule.

Case 1: $i = j$

$$\begin{aligned}
\frac{\partial P(\hat{y}_i = 1 \mid X)}{\partial X^T\Theta_j} = \frac{\partial P(\hat{y}_i = 1 \mid X)}{\partial X^T\Theta_i} &= \frac{e^{X^T\Theta_i}\sum_{k=1}^{K} e^{X^T\Theta_k} - (e^{X^T\Theta_i})^2}{(\sum_{k=1}^{K} e^{X^T\Theta_k})^2} \\
&= \frac{e^{X^T\Theta_i}}{\sum_{k=1}^{K} e^{X^T\Theta_k}} - \left(\frac{e^{X^T\Theta_i}}{\sum_{k=1}^{K} e^{X^T\Theta_k}}\right)^2 \\
&= P(\hat{y}_i = 1 \mid X) - (P(\hat{y}_i = 1 \mid X))^2 \\
&= P(\hat{y}_i = 1 \mid X)(1 - P(\hat{y}_j = 1 \mid X))
\end{aligned}$$

Case 2: $i \neq j$

$$\begin{aligned}
\frac{\partial P(\hat{y}_i = 1 \mid X)}{\partial X^T\Theta_j} &= \frac{0 - (e^{X^T\Theta_i})(e^{X^T\Theta_j})}{(\sum_{k=1}^{K} e^{X^T\Theta_k})^2} \\
&= -\left(\frac{(e^{X^T\Theta_i})(e^{X^T\Theta_j})}{(\sum_{k=1}^{K} e^{X^T\Theta_k})(\sum_{k=1}^{K} e^{X^T\Theta_k})}\right) \\
&= -P(\hat{y}_i = 1 \mid X)P(\hat{y}_j = 1 \mid X)
\end{aligned}$$

Therefore,

$$\frac{\partial P(\hat{y}_i = 1 \mid X)}{\partial X^T\Theta_j} = \begin{cases} P(\hat{y}_i = 1 \mid X)(1 - P(\hat{y}_j = 1 \mid X)) & \text{if } x = y \\ -P(\hat{y}_i = 1 \mid X)P(\hat{y}_j = 1 \mid X) & \text{if } x \neq y \end{cases}$$

4b)
Considering

$$\begin{aligned}
\frac{\partial X^T\Theta_j}{\partial \Theta_j} &= (X^T)^T \\
&= X
\end{aligned}$$

Therefore, by using a) and above result,

$$\begin{aligned}
\frac{\partial P(\hat{y}_i = 1 \mid X)}{\partial \Theta_j} &= \frac{\partial P(\hat{y}_i = 1 \mid X)}{\partial X^T\Theta_j}\frac{\partial X^T\Theta_j}{\partial \Theta_j} \\
&= \begin{cases} P(\hat{y}_i = 1 \mid X)(1 - P(\hat{y}_j = 1 \mid X))X & \text{if } x = y \\ -P(\hat{y}_i = 1 \mid X)P(\hat{y}_j = 1 \mid X)X & \text{if } x \neq y \end{cases}
\end{aligned}$$

4c)
Considering

$$
\begin{aligned}
\frac{\partial E(\Theta)}{\partial P(\hat{y}_i = 1 \mid X)} &= \frac{\partial - \sum_{i=1}^{K} y_i \ln P(\hat{y}_i = 1 \mid X)}{\partial P(\hat{y}_i = 1 \mid X)} \\
&= - \sum_{i=1}^{K} \left(\frac{\partial y_i \ln P(\hat{y}_i = 1 \mid X)}{\partial P(\hat{y}_i = 1 \mid X)} \right) \\
&= - \sum_{i=1}^{K} \left(0 + y_i \frac{1}{P(\hat{y}_i = 1 \mid X)} \right) \\
&= - \sum_{i=1}^{K} \left(\frac{y_i}{P(\hat{y}_i = 1 \mid X)} \right) \\
&= \frac{-1}{P(\hat{y}_i = 1 \mid X)} && \text{Since only one of the } y_i \text{ will be 1}
\end{aligned}
$$

Therefore, combining above and b),

Case 1: $i = j$

$$
\begin{aligned}
\frac{\partial E(\Theta)}{\partial \Theta_j} &= \frac{\partial E(\Theta)}{\partial P(\hat{y}_i = 1 \mid X)} \frac{\partial P(\hat{y}_i = 1 \mid X)}{\partial X^T \Theta_j} \frac{\partial X^T \Theta_j}{\partial \Theta_j} \\
&= \frac{-1}{P(\hat{y}_i = 1 \mid X)} P(\hat{y}_i = 1 \mid X)(1 - P(\hat{y}_j = 1 \mid X))X \\
&= (P(\hat{y}_j = 1 \mid X) - 1)X \\
&= XP(\hat{y}_j = 1 \mid X) - X \\
&= XP(\hat{y}_j = 1 \mid X) - Xy_j && \text{By case condition, } i = j \text{ means } j \text{ is the true class, so } y_j \text{ must be 1}
\end{aligned}
$$

Case 2: $i \neq j$

$$
\begin{aligned}
\frac{\partial E(\Theta)}{\partial \Theta_j} &= \frac{\partial E(\Theta)}{\partial P(\hat{y}_i = 1 \mid X)} \frac{\partial P(\hat{y}_i = 1 \mid X)}{\partial X^T \Theta_j} \frac{\partial X^T \Theta_j}{\partial \Theta_j} \\
&= \frac{-1}{P(\hat{y}_i = 1 \mid X)} (-P(\hat{y}_i = 1 \mid X)(P(\hat{y}_j = 1 \mid X))X) \\
&= P(\hat{y}_i = 1 \mid X)X \\
&= XP(\hat{y}_j = 1 \mid X) - Xy_j && \text{By case condition, } i \neq j \text{ means } j \text{ is } \textbf{NOT} \text{ the true class, so } y_j \text{ must be 0}
\end{aligned}
$$

Combining 2 cases,

$$
\frac{\partial E(\Theta)}{\partial \Theta_j} = XP(\hat{y}_j = 1 \mid X) - Xy_j
$$