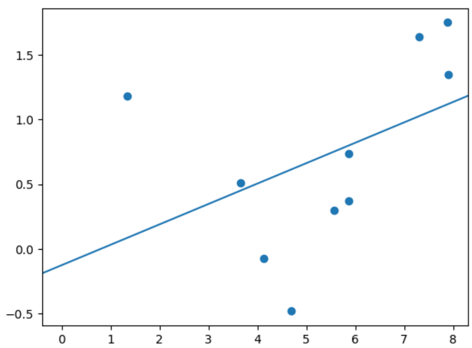
CSCI3230 / ESTR3108 2023-24 First Term Assignment 1

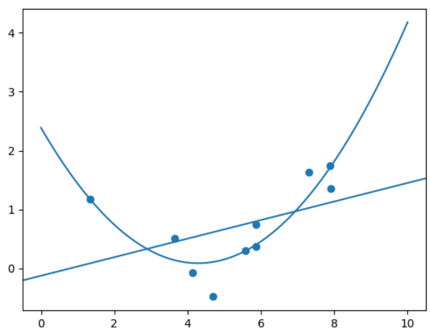
I declare that the assignment here submitted is original except for source material explicitly acknowledged, and that the same or closely related material has not been previously submitted for another course. I also acknowledge that I am aware of University policy and regulations on honesty in academic 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: Fong Long Wai

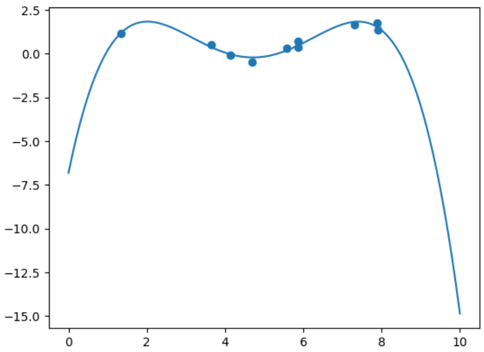
Student ID : 1155177220



1a's figure.



1b's figure.



1c’s figure.

CSCI 3230 hw 1

1.a. $\hat{\theta} = (X^T X)^{-1} X^T Y$ since X is invertible

$$= \left[\begin{pmatrix} 5.86 & \cdots & 7.89 \end{pmatrix} \begin{pmatrix} 5.86 \\ \vdots \\ 7.89 \end{pmatrix} \right]^{-1} \begin{pmatrix} 5.86 & \cdots & 7.89 \end{pmatrix} \begin{pmatrix} 0.74 \\ \vdots \\ 1.75 \end{pmatrix}$$

$$= \left[\begin{pmatrix} 1 & \cdots & 1 \\ 5.86 & \cdots & 7.89 \end{pmatrix} \begin{pmatrix} 1 & 5.86 \\ \vdots & \vdots \\ 1 & 7.89 \end{pmatrix} \right]^{-1} \begin{pmatrix} 1 & \cdots & 1 \\ 5.86 & \cdots & 7.89 \end{pmatrix} \begin{pmatrix} 0.74 \\ \vdots \\ 1.75 \end{pmatrix}$$

$$\simeq \begin{pmatrix} -0.125 \\ 0.1576 \end{pmatrix}$$

b. Answer $\simeq \begin{pmatrix} 2.389 \\ -1.0704 \\ 0.125 \end{pmatrix}$

c. Answer $\simeq \begin{pmatrix} -6.808 \\ 10.917 \\ -4.605 \\ 0.732 \\ -0.039 \end{pmatrix}$

No.

Date.

1d. Q1a's may have underfitting
Q1c's may have overfitting
Q1b's may be a relatively good one

1a's RSS ~~prediction error~~ ≈ 9.4156
1b's RSS ≈ 5.0134
1c's RSS ≈ 19.5045
1a's expected prediction error ≈ 0.94156
1b's ≈ 0.50134
1c's ≈ 1.95045

2b. $\theta_0 = \bar{Y} - \bar{X}\theta$

$$J(\theta, \theta_0) = \sum_{i=1}^{m} \left(\quad \right)^2 + \lambda \|\theta\|_2^2$$

$$\left[(X^{(i)} - \bar{X})\theta + \theta_0 + \bar{X}\theta - Y^{(i)} \right]$$

$$= \sum_{i=1}^{m} \left(X_c^{(i)}\theta + \bar{Y} - Y^{(i)} \right)^2 + \lambda \|\theta\|_2^2$$

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

$$= J_c(\theta)$$

No.

Date.

3 a. $P(\hat{y}=1 \mid x_1, x_2)$

$$= \frac{1}{1+e^{-(1.5x_1+0.5x_2-1)}}$$

$$= \frac{1}{1+e^{(1-1.5x_1-0.5x_2)}}$$

Cross-entropy error function:

$$= \sum_{i=1}^{m} -y^{(i)} \ln\left(\frac{1}{1+e^{(1-1.5x_1^{(i)}-0.5x_2^{(i)})}}\right) - (1-y^{(i)}) \ln\left(1-\frac{1}{1+e^{(1-1.5x_1^{(i)}-0.5x_2^{(i)})}}\right),$$

$y \in \{0,1\}$

b. $E(\theta)$ = cross-entropy error function. $\theta = \begin{pmatrix} -1 \\ 1.5 \\ 0.5 \end{pmatrix}$

$$\nabla E(\theta) = \sum_{i=1}^{m} \left(\frac{1}{1+e^{(1-1.5x_1^{(i)}-0.5x_2^{(i)})}} - y^{(i)}\right) \begin{pmatrix} 1 \\ x_1^{(i)} \\ x_2^{(i)} \end{pmatrix}$$

new $\theta = \theta - 0.1 \nabla E(\theta)$

$$\approx \theta - 0.1 \cancel{\begin{pmatrix} 1.386 \\ -1.09 \\ -0.757 \end{pmatrix}} \begin{pmatrix} 0.18998 \\ -0.32105 \\ -0.1161 \end{pmatrix}$$

$$\approx \cancel{\begin{pmatrix} -1.136 \\ 1.609 \\ 0.5757 \end{pmatrix}} \begin{pmatrix} -1.0019 \\ 1.5321 \\ 0.50118 \end{pmatrix}$$

c. True positive = 3   False Positive = 2   Accuracy = $\frac{2}{3}$ Recall = 1

True Negative = 1   False negative = 0   Precision = $\frac{3}{5}$

~~4a. for j=i,~~ $\cancel{\dfrac{\partial P(\hat{y}_i=1|X)}{\partial X^T\theta_j} = \dfrac{e^{X^T\theta_i}}{\sum_{k=1}^{K} e^{X^T\theta_k}} \times \left(1-\dfrac{e^{X^T\theta_i}}{\sum_{k=1}^{K} e^{X^T\theta_k}}\right)}$

$$\cancel{= \left(\frac{e^{X^T\theta_i}}{\sum_{k=1}^{K} e^{X^T\theta_k}}\right) - \left(\frac{e^{X^T\theta_i}}{\sum_{k=1}^{K} e^{X^T\theta_k}}\right)^2}$$

4a. for $j=i$,

$$\frac{\partial}{\partial X^T\theta_j}\left(\frac{e^{X^T\theta_i}}{\sum_{k=1}^{K} e^{X^T\theta_k}}\right) = \frac{\partial}{\partial X^T\theta_j}\left[\left(e^{X^T\theta_i}\right)\left(\sum_{k=1}^{K} e^{X^T\theta_k}\right)^{-1}\right]$$

$$= \frac{e^{X^T\theta_i}}{\sum_{k=1}^{K} e^{X^T\theta_k}} + e^{X^T\theta_i} \times \frac{1}{\sum_{k=1}^{K} e^{X^T\theta_k}} e^{X^T\theta_i} \times (-1)$$

$$= P(\hat{y}_i=1|X)(1-P(\hat{y}_j=1|X))$$

for $j \neq i$,

$$\frac{\partial}{\partial X^T\theta_j}\left(\frac{e^{X^T\theta_i}}{\sum_{k=1}^{K} e^{X^T\theta_k}}\right) = 0\left(\sum_{k=1}^{K} e^{X^T\theta_k}\right)^{-1} + \frac{e^{X^T\theta_i} \times e^{X^T\theta_j} \times (-1)}{\sum_{k=1}^{K} e^{X^T\theta_k}}$$

$$= -P(\hat{y}_i=1|X)P(\hat{y}_j=1|X)$$

4b $$\frac{\partial P(\hat{y}_i=1|X)}{\partial \theta_j} = \frac{\partial P(\hat{y}_i=1|X)}{\partial X^T\theta_j}\frac{\partial X^T\theta_j}{\partial \theta_j}$$

$$= -P(\hat{y}_i=1|X)P(\hat{y}_j=1|X)X \quad \text{if } i \neq j$$

$$= P(\hat{y}_i=1|X)(1-P(\hat{y}_j=1|X))X \quad \text{if } i=j$$

No.

Date.

4(. $\frac{\partial E(\theta)}{\partial \theta_j} = \frac{\partial E(\theta)}{\partial P(\hat{Y}_i=1|X)} \frac{\partial P(\hat{Y}_i=1|X)}{\partial X^T\theta_j} \frac{\partial X^T\theta_j}{\partial \theta_j}$

$\cancel{= -Y_j \frac{1}{P(\hat{Y}_i=1|X)} \frac{\partial P(\hat{Y}_i=1|X)}{\partial X^T\theta_j} \frac{\partial X^T}{\partial \theta}}$

$= \frac{\partial}{\partial \theta_j}\left(-\sum_{i=j} Y_i \ln P(\hat{Y}_i=1|X)\right)$

$+ \frac{\partial}{\partial \theta_j}\left(-\sum_{i\neq j} Y_i \ln P(\hat{Y}_i=1|X)\right)$

$= -\sum_{i=j} Y_i \frac{1}{P(\hat{Y}_i=1|X)} \times P(\hat{Y}_i=1|X)(1-P(\hat{Y}_j=1|X))X$

$+ \sum_{i\neq j} Y_i \frac{1}{P(\hat{Y}_i=1|X)} \times P(\hat{Y}_i=1|X)(P(\hat{Y}_j=1|X))X$

$= \sum_{i=j} (P(\hat{Y}_i=1|X)-1)Y_i X$

$+ \sum_{i\neq j} P(\hat{Y}_j=1|X)Y_i X$

$= P(\hat{Y}_j=1|X)Y_j X - Y_j X^T + \sum_{i\neq j} P(\hat{Y}_j=1|X)Y_i X$

$= -Y_j X + \sum_{k=1}^{K} P(\hat{Y}_j=1|X)Y_i X$

$= -Y_j X + 1\times P(\hat{Y}_j=1|X) X$

$= XP(\hat{Y}_j=1|X) - XY_j$

No.

Date.

2c. $\cancel{\frac{d J_c(\theta)}{d\theta} = \sum_{i=1}^{m} X_c^{(i)} + 2\lambda \|\theta\|_2}$

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

$$= \|x_c \theta - y_c\|^2 + \lambda \|\theta\|_2^2, \quad x_c = \begin{pmatrix} X_c^{(1)} \\ \vdots \\ X_c^{(m)} \end{pmatrix} \quad y_c = \begin{pmatrix} Y_c^{(1)} \\ \vdots \\ Y_c^{(m)} \end{pmatrix}$$

$$= (x_c \theta - y_c)^T (x_c \theta - y_c) + \lambda \theta^T \theta$$

$$= \theta^T x_c^T x_c \theta - y_c^T x_c \theta - \theta^T x_c^T y_c - y_c^T y + \lambda \theta^T \theta$$

$$\frac{\partial J_c(\theta)}{\partial \theta} = 2 x_c^T x_c \theta - x_c^T y_c - x_c^T y_c + 2 \lambda I \theta = 0$$

$$(x_c^T x_c + \lambda I)\theta - x_c^T y_c = 0$$

$$\theta = (x_c^T x_c + \lambda I)^{-1} (x_c^T y_c)$$

$$\frac{\partial^2 J_c(\theta)}{\partial \theta} = (x_c^T x_c + \lambda I) > 0 \text{ and } J_c(\theta) \text{ is convex}$$

$\therefore J_c(\theta)$ is in global min when $\frac{\partial J_c(\theta)}{\partial \theta} = 0$,

$$\min_{\theta}(J_c(\theta)) = (x_c^T x_c + \lambda I)^{-1} (x_c^T y_c)$$

2a. $\frac{\partial J(\theta, \theta_0)}{\partial \theta_0} = \sum_{i=1}^{m} 2\left[(X^{(i)} - \bar{x})\theta + \theta_0 + \bar{x}\theta - Y^{(i)}\right]$

$$= 2\sum_{i=1}^{m}\left[X^{(i)}\theta + \theta_0 - Y^{(i)}\right] = 0$$

$$m\theta_0 + \sum_{i=1}^{m} X^{(i)}\theta - \sum_{i=1}^{m} Y^{(i)} = 0$$

$$\theta_0 = \bar{Y} - \bar{x}\theta$$

Since $J(\theta, \theta_0)$ is convex, $\theta_0 = \bar{Y} - \bar{x}\theta$,

~~can make~~ $J(\theta, \theta_0)$ is in global minimum.

and $\frac{\partial^2 J(\theta, \theta_0)}{\partial \theta_0^2} = m \geq 0$.

So when

Q1

```
import numpy as np;
import matplotlib.pyplot as plt;

def OLS(XT,y):
    temp=np.matmul(XT,XT.T);
    temp=np.linalg.inv(temp);
    temp=np.matmul(temp,XT);
    temp=np.matmul(temp,y);
    print("coefficient is: ");
    print(temp);
    return temp;

def render(x1,y,Coefficient):
    plt.scatter(x1,y);
    plotx= np.linspace(0, 10, 100);
    ploty = np.polyval(Coefficient, plotx)
    plt.plot(plotx,ploty);
    plt.show();

def test(testx,testy,CoefficientInReverse):
    predictArray=np.zeros(sample_size,dtype=float);
    multiple=np.ones(sample_size,dtype=float);
    for coefficient in CoefficientInReverse:
        predictArray+=coefficient*multiple;
        multiple*=testx;
    print("predict is: ");
    print(predictArray)
    predictArray=predictArray-testy;
    predictArray=predictArray*predictArray;
    error=predictArray.sum();
    print("error is: ");
    print(error);

sample_size=10;
x1 = np.array([5.86,1.34,3.65,4.69,4.13,5.87,7.91,5.57,7.3,7.89]);
y = np.array([0.74,1.18,0.51,-0.48,-0.07,0.37,1.35,0.3,1.64,1.75]);
```

```
x2= x1*x1;
x3=x2*x1;
x4=x3*x1;
XT=np.vstack(([1,1,1,1,1,1,1,1,1,1],x1,x2,x3,x4));
print(XT);
render(x1,y,OLS(XT,y)[::-1])


testx=([5.8,0.57,4.3,6.55,0.82,3.72,5.8,3.26,6.75,4.77]);
testy=([0.93,1.87,-0.06,1.6,1.22,0.9,0.93,1.53,1.73,-0.51]);
test(testx,testy,OLS(XT,y));
```

Q3

```
import numpy as np
import math
import matplotlib.pyplot as plt;
def sig(X):
    return 1/ (1+  math.e**(-X));
def df(theta,trainX,trainY):
    total=np.zeros(dim);
    for X, Y in zip(trainX,trainY):
        temp=sig((np.dot(X,theta)))-Y;
        total+=temp*X;
    print("df: ");
    print(total);
    return total;
def test(theta,testX,testY):
    print("Beginning Testing: ")
    TP=0;TN=0;FP=0;FN=0;
    testResult=sig((np.dot(testX,theta)));
    testResult=np.where(testResult>=0.5,1,0);
    print(testResult);
    for X, Y in zip(testResult,testY):
        if X==1 and Y==1:
            TP=TP+1;
        if X==0 and Y==0:
            TN=TN+1;
        if X==1 and Y==0:
            FP=FP+1;
        if X==0 and Y==1:
            FN=FN+1;
    print("TP %d TN %d FP %d FN %d"%(TP,TN,FP,FN))
dim=3;
theta=np.array([-1,1.5,0.5]);
trainX=np.array([[1,0.346,0.78],[1,0.303,0.439],[1,0.358,0.729],[1,0.60
2,0.863],[1,0.790,0.753],[1,0.611,0.965]]);
trainY=np.array([0,0,0,1,1,1]);
theta=theta-0.1*df(theta,trainX,trainY);
```

```
testX=np.array([[1,0.959,0.382],[1,0.75,0.306],[1,0.395,0.76],[1,0.823,
0.764],[1,0.761,0.874],[1,0.844,0.435]])
testY=np.array([0,0,0,1,1,1]);
test(theta,testX,testY);
```