Assignment 3 ,Machine Learning,CS-GY 6923, Th 6:00 - 8:30 PM
Shantanu Tripathi

Homework 3

Part I

Ans 1) a)

x_1 be variable for Cancer Volume
x_2 be variable for Patient's age
x_3 be " " Cancer Type

Model 1:

$$\hat{y}^{(i)} = w_0 + w_1 x_1^{(i)}$$

$$\hat{y} = \mathbf{X}\mathbf{w}$$

where:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1^{(1)} \\ 1 & x_1^{(2)} \\ 1 & x_1^{(3)} \\ & \vdots \\ 1 & x_1^{(N)} \end{bmatrix}$$

& $\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$

Model 2:

$$\hat{y}^{(i)} = w_0 + w_1 x_1^{(i)} + w_2 x_2^{(i)}$$

or

$$\hat{y} = X w$$

where $X = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} \\ \vdots & & \\ 1 & x_1^{(N)} & x_2^{(N)} \end{bmatrix}$

$$W = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} \quad \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$

Ans 1)b) Model 3:

Here, we have taken a new column called x_3.

$$\boxed{\begin{array}{l} x_3 = 1 \\ x_3 = 0 \end{array}} \begin{array}{l} \rightarrow \text{type } 1 \\ \rightarrow \text{type } 2 \end{array}$$

The model for this is:

$$\cancel{\hat{y}^{(i)} = w_0 + x_1 . x}$$

$$\boxed{\begin{aligned}\hat{y}^{(i)} &= w_0 + x_1^{(i)} . x_3^{(i)} . w_1 \\ &+ x_1^{(i)} . (1 - x_3^{(i)}) . w_2 \\ &+ w_3 . x_2^{(i)}\end{aligned}}$$

↳ we have ~~replaced~~ removed features x_1 & x_3 ~~by feature~~ & have added two new feature $x_1 . x_3$ & $x_1 . (1 - x_3)$

So, $y = Xw$

where: $w = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{bmatrix}$

$$\& \quad X = \begin{bmatrix} 1 & x_1^{(1)} \cdot x_3^{(1)} & x_1^{(1)} \cdot (1 - x_3^{(1)}) & x_2^{(1)} \\ 1 & x_1^{(2)} \cdot x_3^{(2)} & x_1^{(2)} \cdot (1 - x_3^{(2)}) & x_2^{(2)} \\ 1 & x_1^{(3)} \cdot x_3^{(3)} & x_1^{(3)} \cdot (1 - x_3^{(3)}) & x_3^{(3)} \\ & \vdots & & \\ 1 & x_1^{(N)} \cdot x_3^{(N)} & x_1^{(N)} \cdot (1 - x_3^{(N)}) & x_3^{(N)} \end{bmatrix}$$

Ans 1) c)

~~Number of Params in Model 1 : 2~~
~~" " " " " 2 : 3~~

Model 1 param count = 2
Model 2 param count = 3

Out of models 1 & 2, 2 is more complex, due to more ~~Out of~~ number of params.

Out of models 1, 2 & 3, 3 is ~~most complex~~, most complex. This is because of the multiplication of two features to make a new one.

Ans 1) d)

Rows for model 1:

~~Cancer~~

$$X = \begin{bmatrix} 1 & 0.7 \\ 1 & 1.3 \\ 1 & 1.6 \end{bmatrix}$$

Rows for model 2:

$$X = \begin{array}{c} \begin{array}{ccc} & \text{Cancer vol.} & \text{Patient age} \end{array} \\ \begin{bmatrix} 1 & 0.7 & 55 \\ 1 & 1.3 & 65 \\ 1 & 1.6 & 70 \end{bmatrix} \end{array}$$

Rows for model 3 will be as:

Let cancer type be replaced with 0s & 1s.

If type is I, replace with 1
" " is II " " 0

(a) Cancer type	(b) Cancer Vol.	(a×b)	(1-a)×b
1	0.7	0.7	0
0	1.3	0	1.3
0	1.6	0	1.6

new column

new value of feature

Patient's age

$$X = \begin{bmatrix} 1 & 0.7 & 0 & 55 \\ 1 & 0 & 1.3 & 65 \\ 1 & 0 & 1.6 & 70 \end{bmatrix}$$

Ans 1) e) The model having the minimum value of (validation MSE) should be used:

∴ Model 2, with validation MSE of 0.72 is used.

Ans 2)

~~We can have a~~

The data collected over the time looks as follows:

~~Rainfall Fertilizer Avg. temp. Mold~~

Ans 2) Suppose we have the following data:

Amount of Rainfall	Amount of fertilizer	Average Temperature	No. of Sunny days	Crop Yield.
~~a~~	~~b~~	~~c~~	~~d~~	
~~e~~	~~f~~	~~g~~	~~h~~	
28.0	39.5	50.00	37	1600
39.2	48.7	98.00	50	1800

Here, we can treat 4 independent variables:

Rainfall : x_1
fertilizer : x_2
Temperature : x_3
No. of Sunny day : x_4

} 4 independent variables.

These 4 independent variables are used to get a linear function which

determines the value of the dependent variable "y" ~~w~~. Here, the dependent variable is:

~~the crop yield~~, the crop yield.

Here:, the crop yield (y) depends linearly on:

- ~~Rai~~ Amount of Rainfall (x_1)
- Amount of Fertilizers (x_2)
- ~~Amount of~~
- ~~Amount of~~
- Average Temperature (x_3)
- No. of Sunny days (x_4)

$$\Rightarrow \boxed{\hat{y}^{(i)} = w_0 + w_1 x_1^{(i)} + w_2 x_2^{(i)} + w_3 x_3^{(i)} + w_4 x_4^{(i)}}$$

Now, this becomes a typical linear regression problem, where we have to ~~get the~~ train the value of parameters

$$W = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}$$

so that ~~we~~ the value of crop yield (y) can be predicted with least RSE.

Ans 3)

a)

$$X = \begin{bmatrix} 1 & 28540 \\ 1 & 40133 \\ 1 & 39900 \\ 1 & 0 \\ 1 & 0 \\ 1 & 42050 \\ 1 & 43220 \\ 1 & 39565 \\ 1 & 40400 \\ 1 & 54506 \end{bmatrix}$$

b) $W_{lin} = X^{+} y$

Where X^{+} = pseudo inverse

i.e. $X^{+} = (X^{T}X)^{-1} X^{T}$

Multiplication ~~on next page~~:

$$X^{T} \cdot X = \begin{bmatrix} 10 & 328314 \\ 328314 & 13821823450 \end{bmatrix}$$

~~det = 1.38×10^{10}~~

Det =

$13821\,8234500 - (328314)^2$

$= 30428151900$

=> Inv.

$$= \frac{1}{\det.} \cdot \begin{bmatrix} 13821823450 & -328314 \\ -328314 & 10 \end{bmatrix}$$

$$= \begin{bmatrix} 0.45424 & -1.08 \times 10^{-5} \\ -1.08 \times 10^{-5} & 3.286 \times 10^{-10} \end{bmatrix}$$

=> Pseudo-inverse

$$\begin{bmatrix} 0.45424 & -1.08 \times 10^{-5} \\ -1.08 \times 10^{-5} & 3.286 \times 10^{-10} \end{bmatrix} \cdot$$

$$* \begin{bmatrix} 1 & 1 & 1 & \cdots \\ 28540 & 40133 & \cdots & \end{bmatrix}$$

$$\cancel{= \begin{bmatrix} -4.86 \times 10^{-6} \end{bmatrix}}$$

$$= \begin{bmatrix} 0.148862 \end{bmatrix}$$

$$\begin{bmatrix} 0.146008 \\ -1.42 \times 10^{-0} \end{bmatrix}$$

Pseudo inverse X

$$\begin{bmatrix} 0.45424 & -1.08\times10^{-5} \\ -1.08\times10^{-5} & 3.286\times10^{-10} \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 28540 & 40133 & & 54506 \end{bmatrix}$$

$\therefore X^{+} =$ Pseudo inverse

1.46×10

	c1	c2	c3	c4	c5
r1	1.46×10^{-1}	2.12×10^{-2}	2.37×10^{-2}	4.54×10^{-1}	4.54×10^{-1}
r2	-1.41×10^{-6}	2.40×10^{-6}	2.32×10^{-6}	-1.08×10^{-5}	-1.08×10^{-5}

	c6	c7	c8	c9	c10
r1	5.33×10^{-4}	-1.21×10^{-2}	2.73×10^{-2}	1.83×10^{-2}	-1.34×10^{-1}
r2	3.03×10^{-6}	3.41×10^{-6}	2.21×10^{-6}	2.49×10^{-6}	7.12×10^{-6}

This above is a 2×10 matrix.

r1: row 1
r2: row 2

{ The first 2 lines, represent first 5 columns }

{ The last 2 lines, represent last 5 columns }

$$W = X^{+} \cdot y$$

$$= \begin{bmatrix} 1.46\times10^{-1} & \cdots & -1.34\times10^{-1} \\ -1.41\times10^{-6} & \cdots & 7.12\times10^{-6} \end{bmatrix} \begin{bmatrix} 137 \\ 135 \\ \vdots \\ 117 \\ 114 \end{bmatrix}$$

$$\Rightarrow W = \begin{bmatrix} 124.56 \\ -6.30 \times 10^{-5} \end{bmatrix}$$

$$\therefore g(x) = 1.25 \times 10^{2}$$

$$\therefore g(x) = (-6.30 \times 10^{-5}) \times x + w_0$$

$$\therefore g(x) = \underbrace{(-6.30\times10^{-5})}_{w_1} \times x + \underbrace{124.56}_{w_0}$$

~~Ans 3) c) Given~~

Ans 3) c) Graph given along

Ans 3) d)

$$R^2 = \frac{ess}{tss}$$

$$ess = \sum_{i=1}^{N} (\hat{y}^{(i)} - \bar{y})^2$$

$$tss = \sum_{i=1}^{N} (y^{(i)} - \bar{y})^2$$

~~ess =~~

~~$\bar{y} = 125$~~

~~$\bar{y} = 122.5$~~

~~diff =~~

~~$\hat{y} = [122.7$~~

$$\hat{y} = [122.77 \quad 122.039 \quad 122.054 \quad 124.569 \quad 124.569 \quad 121.918 \quad 121.845 \quad 122.075 \quad 122.022 \quad 121.134]$$

$\cancel{\hat{y} - y =}$

$$ess = \sum_{i=1}^{N} (\hat{y}^{(i)} - \bar{y})^2$$

$$\bar{y} = 122.5$$

$$ess \Rightarrow (122.77 - 122.5)^2 + (122.039 - 122.5)^2 + \dots (121.134 - 122.5)^2$$

$$\boxed{ess = 12.087}$$

$\cancel{tss \Rightarrow \sum y - \bar{y}}$

$$tss = \sum_{i=1}^{N} (y^{(i)} - \bar{y})^2$$

$$= (137 - 122.5)^2 + (135 - 122.5)^2 + \dots (114 - 122.5)^2$$

$$\boxed{tss = 566.5}$$

$$\boxed{R^2 = \frac{ess}{tss} = \frac{12.087}{566.5} = 0.0213}$$

Ans 3) e)

$$y = (-6.3 \times 10^{-5}) . x + 1204.56$$

$$= -6.3 \times 10^{-5} \times 40000 + 1204.56$$

$$= \cancel{122.048} = \boxed{122.048}$$

Salary would be ~~[illegible]~~ 122.048 thousand

Ans 3) f) After removing the outliers

~~The equation achieve~~

Using closed norm form value of the ~~[illegible]~~ matrix :)

$$w = \begin{bmatrix} 1.6 \times 10^{2} \\ -9.05 \times 10^{-4} \end{bmatrix}$$

=> i.e. equation is:

$$\boxed{g(x) = \underbrace{(-9.05 \times 10^{-4})}_{w_1} \times x + \underbrace{1.6 \times 10^{2}}_{w_0}}$$

=> Graph will be provided alongwith.

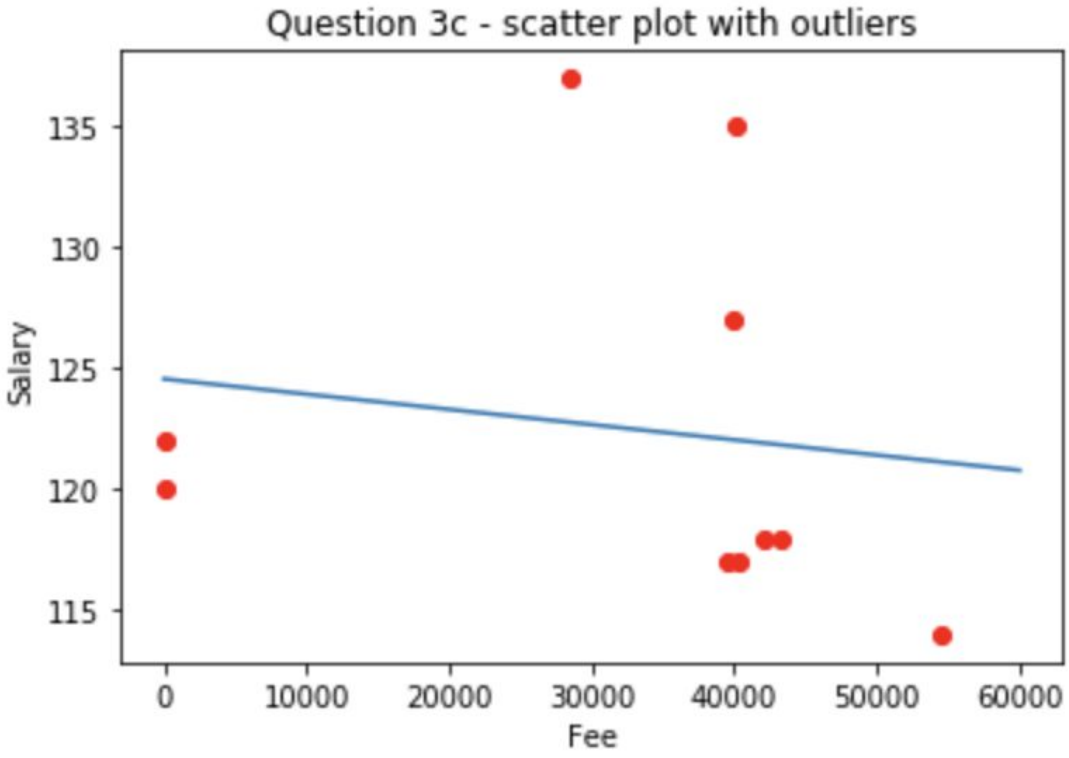
=> $\boxed{R^2 = \frac{ess}{tss}}$, $\boxed{ess = 284.894}$ $\boxed{tss = 558.875}$

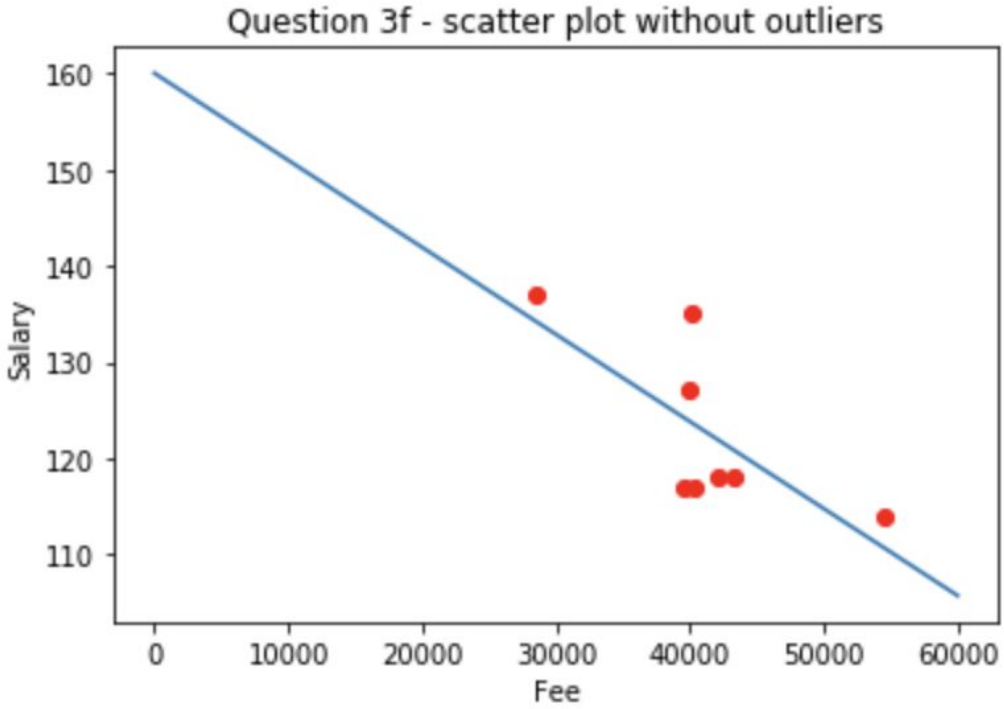
$$\boxed{R^2 = \frac{284.894}{558.875} = 0.50976}$$

=> For 40000, value is:

$$y = (-9.05 \times 10^{-4}) \times x + 160$$
$$= (-9.05 \times 10^{-4} \times 40000) + 160$$
$$= 123.8$$

$$\therefore \boxed{y = 123.8}$$


Question 3c - scatter plot with outliers
Salary
135
130
125
120
115
0
10000
20000
30000
40000
50000
60000
Fee


Question 3f - scatter plot without outliers
Salary
160
150
140
130
120
110
0
10000
20000
30000
40000
50000
60000
Fee

Ans 4) a) $x^{(i)} \in \mathbb{R}^n$

This has been given as x_i in the question

RSS cost for a single $x^{(i)}$ is:

$$= \left(y^{(i)} - \hat{y}^{(i)}\right)^2 \text{ or } \left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

This has been given as y_i in the question

$$\text{Cost} = \left(y^{(i)} - w x^{(i)}\right)^2$$

where $w \in \mathbb{R}^{1 \times n}$

Ans 4) b) Let $x^{(i)} \in \mathbb{R}^n$ & $w \in \mathbb{R}^{1 \times n}$

$$\hat{y}^{(i)} = w x^{(i)}$$

$$\text{cost} = \sum_{i=1}^{N} \left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

$$= \sum_{i=1}^{N} \left(w . x^{(i)} - y^{(i)}\right)^2$$

$$\cancel{= \sum_{i=1}^{N} \left[(w_0 x_0^{(i)} + \right.}$$

$\therefore$ We do not have an intercept.

$$w x^{(i)} = w_1 x_1^{(i)} + w_2 x_2^{(i)} + w_3 x_3^{(i)} + \dots + w_n x_n^{(i)}$$

$\Rightarrow$ RSS =

$$\sum_{i=1}^{N} \left[(w_1 x_1^{(i)} + w_2 x_2^{(i)} + \dots w_n x_n^{(i)}) - y^{(i)} \right]^2$$

$$\frac{\delta RSS}{\delta w_1} = 2 \sum_{i=1}^{N} \left[(w_1 x_1^{(i)} + \dots w_n x_n^{(i)}) - y^{(i)} \right] \left[x_1^{(i)} \right]$$

Put $\frac{\delta RSS}{\delta w_1} = 0$

$\Rightarrow$

$$\sum_{i=1}^{N} \left[(w_1 x_1^{(i)} + \dots w_n x_n^{(i)}) - y^{(i)} \right] . (x_1^{(i)}) = 0$$

$$\sum_{i=1}^{N} w_1 . x_1^{(i)} . x_1^{(i)} + \sum_{i=1}^{N} \left[(w_2 x_2^{(i)} + \dots w_n x_n^{(i)}) - y^{(i)} \right] =$$

$$= 0$$

(2i)

$$w_1 \sum_{i=1}^{N} \left(x_1^{(i)}\right)^2 = -\sum_{i=1}^{N} \left[w_2 x_2^{(i)} + \dots w_n x_n^{(i)} - y^{(i)}\right] \cdot x_1^{(i)}$$

$$w_1 = \frac{-\sum_{i=1}^{N} \left[w_2 x_2^{(i)} + \dots w_n x_n^{(i)} - y^{(i)}\right] \cdot x_1^{(i)}}{\sum_{i=1}^{N} \left(x_1^{(i)}\right)^2}$$

$$\cancel{w_j = -\sum_{i=1}^{N} \sum_{k=0, k \neq j}}$$

$$\boxed{w_j = -\frac{\sum_{i=1}^{N} \sum_{k=0, k \neq j}^{N} w_k \cdot x_k^{(i)} \cdot x_j^{(i)}}{\sum_{i=1}^{N} \left(x_j^{(i)}\right)^2}}$$

Ans 5)

Some formulae for matrix differentiation (calculus)

$$\boxed{\begin{aligned} y &= Ax \\ \frac{\delta y}{\delta x} &= A^T \end{aligned}}$$

$$\boxed{\begin{aligned} y &= x^T A \\ \frac{\delta y}{\delta x} &= A \end{aligned}}$$

$$\boxed{\begin{aligned} y &= x^T x \\ \frac{\delta y}{\delta x} &= 2x \end{aligned}}$$

$$\boxed{\begin{aligned} y &= x^T A x \\ \frac{\delta y}{\delta x} &= (A + A^T)x \end{aligned}}$$

$$\Rightarrow \hat{W}_{WLS} = \underset{w \in \mathbb{R}^d}{\arg\min} \left((y - Xw)^T \Omega (y - Xw) \right)$$

Simplifying this equation

$$\Rightarrow$$

$$= \left(-w^T X^T + y^T \right) . \Omega . (y - Xw)$$

$$\cancel{= -w^T x}$$

$$= \left(-w^T X^T \Omega + y^T \Omega\right).(y - Xw)$$

$$\hat{w}_{WLS} = -w^T X^T \Omega y + w^T \underbrace{X^T \Omega X}_{A} w + y^T \Omega y - y^T \Omega X w$$

where

$$\frac{\delta \hat{w}_{WLS}}{\delta w} = -X^T \Omega y + (A + A^T) w + 0 - (y^T \Omega X)^T$$

where $A = X^T \Omega X$

<u>Equating $\delta \hat{w}_{WLS} = 0$</u>

$$\Rightarrow -X^T \Omega y + \left(X^T \Omega X + (X^T \Omega X)^T\right) w - (y^T \Omega X)^T = 0$$

$$= -X^T \Omega y + \left(X^T \Omega X + X^T \Omega^T X\right) w - X^T \Omega^T y$$

$\therefore \Omega$ is a ~~identity~~ Diagonal matrix

$$\Omega^T = \Omega$$

$$\Rightarrow -X^T\Omega y + 2X^T\Omega X w - X^T\Omega y = 0$$

$$= \cancel{2}X^T\Omega X w = \cancel{2}X^T\Omega y$$

$$\Rightarrow \boxed{w = (X^T\Omega X)^{-1}(X^T\Omega) \cdot y}$$