Machine Learning
Assignment 3

Name: Girish Chandar G
Roll No: 16110057
Department: Electrical Engineering

Codes for all questions has been coded in single jupyter notebook and headings given accordingly
https://gist.github.com/girish1511/8c2455e0504d2186d0d4f6dc71908e22#file-assignment3-ipynb

Q1. Attached as a continuation to the report after Q6

Q2.
(b) Non-invertibility in the normal equation occurs due to multi-collinearity of the feature matrix X and therefore X^TX is singular and non-invertible. The sci-kit learn uses pseudo-inverse for calculating inverse of singular matrix in the normal equation.

2 (b)

$$X = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 4 \\ 1 & 3 & 6 \\ 1 & 4 & 8 \end{bmatrix}$$

$$X^TX = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 4 \\ 1 & 3 & 6 \\ 1 & 4 & 8 \end{bmatrix}$$

$$X^TX = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 4 & 10 & 20 \\ 10 & 30 & 60 \\ 20 & 60 & 120 \end{bmatrix} \end{array}$$

Col 2 and Col 3 are linearly dependent

Col 3 = 2 x col 2.

$$\Rightarrow \det(X^TX) = 0.$$

Q3.
(b) It is incorrect to assume that larger coefficients means more important feature since the samples are not normalized across features. The magnitude of coefficients is non-normalized data is inversely proportional to range and magnitude of the values of corresponding features.

(c) It can be inferred from the coefficients that the most important feature is latitude of the house. But in general the regression coefficients is not a good way to quantify the importance of corresponding features.

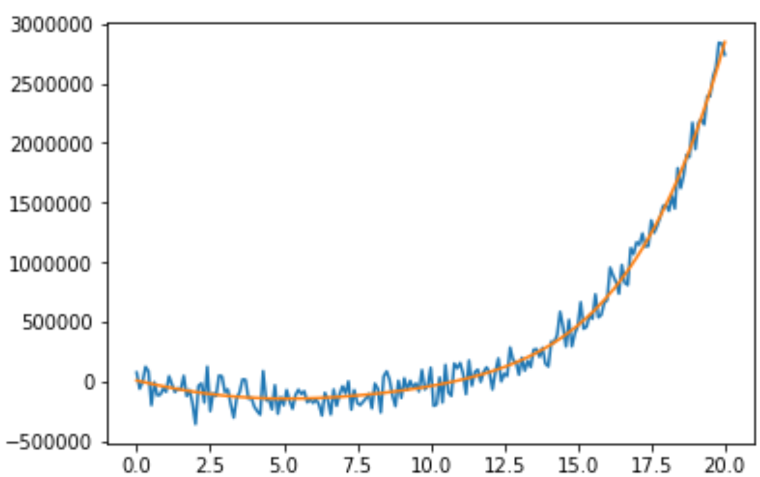
(d)The residuals have a normal distribution.

(e)

1. Exhaustive search:-
   The best subset of features are:
   ['X2 house age', 'X3 distance to the nearest MRT station',
   'X4 number of convenience stores', 'X5 latitude']
   Root Mean Square Error(sub-set of features): 0.0034637565013380237
   Root Mean Square Error(all features): 0.003371103552633352
2. Step-Forward Greedy Selection:-
   The best subset of features are:
   ['X5 latitude', 'X6 longitude', 'X2 house age', 'X3 distance to the nearest MRT station', 'X1 transaction date']
   Root Mean Square Error(sub-set of features): 0.0037800912340117915
   Root Mean Square Error(all features): 0.003297385133896944

Q4.
(a)
Runtime of Normal Equation Regression = 0.01576399803161621
Runtime of Gradient Descent Regression = 0.00998687744140625
Runtime of Autograd Regression = 3.415041923522949
Runtime of PyTorch Regression = 0.41961121559143066
Runtime of Sci-kit learn = 0.565683126449585

Root Mean Square Error (Normal Equation Regression): 0.014125051797358566
Root Mean Square Error (Gradient Descent Regression): 0.012213889542436504
Root Mean Square Error (Autograd Regression): 0.012213889542436506
Root Mean Square Error (PyTorch Regression): 0.019719493118208094
Root Mean Square Error (Sci-kit learn): 0.0033531887752362736

Q5.


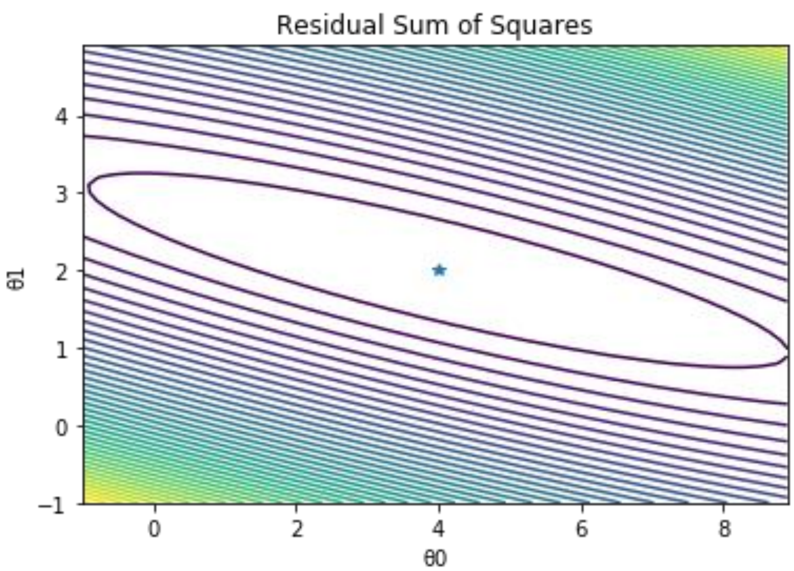
Root Mean Square Error: 96145.33829873

The high deviation of learnt coefficients from the original values is due to the high magnitude of noise added (order of 1e5).

Coefficients	Learnt Coefficients	Original Coefficients
Θ0	8823.456856525532	-10000
Θ1	-57553.651720250404	-300
Θ2	3579.6458352547997	8
Θ3	753.7606786243987	-100
Θ4	-94.74990553647069	3
Θ5	3.652045541414192	1

Q6.

(a)



① $y = \theta_0 + \theta_1 x.$

$$X = \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix} \qquad Y = \begin{bmatrix} 6 \\ 10 \\ 16 \end{bmatrix}$$

(a) $$X_n = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{bmatrix} \qquad Y = \begin{bmatrix} 6 \\ 10 \\ 16 \end{bmatrix}$$

$$(X_n^T \cdot X_n)^{-1} X_n^T \cdot Y = \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

$$(X_n^T \cdot X_n)^{-1} = \left(\begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 6 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{bmatrix} \right)^{-1} = \begin{bmatrix} 3 & 10 \\ 10 & 46 \end{bmatrix}^{-1}$$

$$= \frac{1}{38} \begin{bmatrix} +46 & -10 \\ -10 & +3 \end{bmatrix}$$

$$\cancel{\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} -46 & 10 \\ 10 & -3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{bmatrix}}$$

$$\begin{bmatrix}\theta_0\\ \theta_1\end{bmatrix} = \begin{bmatrix}+46 & -10\\ -10 & +3\end{bmatrix}\begin{bmatrix}1 & 1 & 1\\ 1 & 3 & 6\end{bmatrix}\begin{bmatrix}6\\ 10\\ 16\end{bmatrix}$$

$$= \frac{1}{38}\begin{bmatrix}+46 & -10\\ -10 & +3\end{bmatrix}\begin{bmatrix}32\\ 132\end{bmatrix}$$

$$\boxed{\begin{bmatrix}\theta_0\\ \theta_1\end{bmatrix} = \begin{bmatrix}4\\ 2\end{bmatrix}}$$

(c) $\theta_1 = \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)}$

$\theta_0 = \bar{y} - \theta_1 \bar{x}$

X	Y	XY
1	6	6
3	10	30
6	16	96

$\mathrm{cov}(X,Y) = E[XY] - E[X]E[Y]$

$E[X] = \frac{1+3+6}{3} = 10/3$

$E[Y] = \frac{6+10+16}{3} = 32/3$

$E[XY] = \frac{6+30+96}{3} = 132/3$

$\mathrm{cov}(X,Y) = \frac{132}{3} - \frac{320}{9} = 76/9.$

Scanned by CamScanner

$$Var(x) = E[x^2] - (E[x])^2$$

$$= \frac{1+9+36}{3} - \left(\frac{10}{3}\right)^2$$

$$= \frac{46}{3} - \frac{100}{9}$$

$$Var(x) = \frac{38}{9}$$

$$\Rightarrow \theta_1 = \frac{Cov(x,y)}{Var(x)} = \frac{76/9}{38/9}$$

$$\Rightarrow \boxed{\theta_1 = 2}$$

$$\theta_0 = E[y] - \theta_1 E[x]$$

$$= \frac{32}{3} - \frac{2 \cdot 10}{3}$$

$$\boxed{\theta_0 = 4}$$

$$\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

(b)

Let $\Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_d \end{bmatrix}$ $d \rightarrow$ Number of features

Let $J(\theta) = \frac{\sum \varepsilon_i^2}{N} = \frac{2}{N} \sum (y_i - (\theta_0 + \theta_1 x_1 + \dots \theta_d x_d))^2$

$N \rightarrow$ No of samples

Let X be the feature matrix
y be the ~~ground truth~~ actual value

Let $g = \nabla_\Theta J(\Theta)$

$$\Rightarrow g = -\frac{2}{N} \left[X^T . (y - X\Theta) \right]$$

Updating θ :-
$$\Theta^{(k+1)} = \Theta^{(k)} - \alpha g^{(k)}$$

Given :-

x	y
1	6
3	10
6	16

$\theta_{init} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $\alpha = 0.1$

$X = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{bmatrix}$ $y = \begin{bmatrix} 6 \\ 10 \\ 16 \end{bmatrix}$ $\Theta_{init} = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$N = 3$

$\Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$

Iteration 1:-

$$g^{(0)} = -\frac{2}{3}\left[\begin{bmatrix}1 & 1\\1 & 3\\1 & 6\end{bmatrix}^T \left(\begin{bmatrix}6\\10\\16\end{bmatrix} - \begin{bmatrix}1 & 1\\1 & 3\\1 & 6\end{bmatrix}\begin{bmatrix}0\\0\end{bmatrix}\right)\right]$$

$$g^{(0)} = \begin{bmatrix}-21.3333\\-88\end{bmatrix}$$

$$\theta^{(1)} = \theta^{(0)} - \alpha\, g^{(0)}$$

$$\begin{bmatrix}\theta_0^{(1)}\\\theta_1^{(1)}\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} - 0.1\begin{bmatrix}-21.3333\\-88\end{bmatrix}$$

$$\theta^{(1)} = \begin{bmatrix}\theta_0^{(1)}\\\theta_1^{(1)}\end{bmatrix} = \begin{bmatrix}2.13333\\8.8\end{bmatrix}$$

Iteration 2:-

$$g^{(1)} = -\frac{2}{3}\left[\begin{pmatrix}1 & 1\\1 & 3\\1 & 6\end{pmatrix}^T \left[\begin{pmatrix}6\\10\\16\end{pmatrix} - \begin{pmatrix}1 & 1\\1 & 3\\1 & 6\end{pmatrix}\begin{pmatrix}2.1333\\8.8\end{pmatrix}\right]\right]$$

$$g^{(1)} = \begin{pmatrix}41.6\\196.0889\end{pmatrix}$$

$$\theta^{(2)} = \theta^{(1)} - \alpha\, g^{(1)}$$

$$\begin{bmatrix}\theta_0^{(2)}\\\theta_1^{(2)}\end{bmatrix} = \begin{bmatrix}2.13333\\8.8\end{bmatrix} - 0.1\begin{bmatrix}41.6\\196.0889\end{bmatrix}$$

$$\Theta^{(2)} = \begin{bmatrix} \theta_0^{(2)} \\ \theta_1^{(1)} \end{bmatrix} = \begin{bmatrix} -2.026667 \\ -10.8089. \end{bmatrix}$$

<u>Iteration 3 :-</u>

$$\begin{pmatrix} -2.026667 \\ -10.8089 \end{pmatrix}$$

$$g^{(2)} = -\frac{2}{3}\left[\begin{pmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{pmatrix}^T \left[\begin{pmatrix} 6 \\ 10 \\ 16 \end{pmatrix} - \begin{pmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{pmatrix} \cancel{\begin{pmatrix} 41.6 \\ 196.0892 \end{pmatrix}}\right]\right]$$

$$g^{(2)} = \begin{pmatrix} -97.44593 \\ -432.984 \end{pmatrix}$$

$$\Theta^{(3)} = \Theta^{(2)} - \alpha g^{(2)}$$

$$\begin{bmatrix} \theta_0^{(3)} \\ \theta_1^{(3)} \end{bmatrix} = \begin{bmatrix} -2.026667 \\ -10.8089 \end{bmatrix} - 0.1 \begin{bmatrix} -97.44593 \\ -432.984 \end{bmatrix}$$

$$\Theta^{(3)} = \begin{bmatrix} \theta_0^{(3)} \\ \theta_1^{(3)} \end{bmatrix} = \begin{bmatrix} 7.71793 \\ 32.4895 \end{bmatrix}$$

<u>Iteration 4 :-</u>

$$g^{(3)} = -\frac{2}{3}\left[\begin{pmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{pmatrix}^T \left[\begin{pmatrix} 6 \\ 10 \\ 16 \end{pmatrix} - \begin{pmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{pmatrix}\begin{pmatrix} 7.71793 \\ 32.4895 \end{pmatrix}\right]\right]$$

$$g^{(3)} = \begin{pmatrix} 210.7 \\ 959.797 \end{pmatrix}$$

$$\Theta^{(4)} = \Theta^{(3)} - \alpha g^{(3)}$$

$$\begin{bmatrix} \theta_0^{(4)} \\ \theta_1^{(4)} \end{bmatrix} = \begin{bmatrix} 7.71793 \\ 32.4895 \end{bmatrix} - 0.1 \begin{pmatrix} 210.7 \\ 959.797 \end{pmatrix}$$

$$\Theta^{(4)} = \begin{bmatrix} \theta_0^{(4)} \\ \theta_1^{(4)} \end{bmatrix} = \begin{bmatrix} -13.35198 \\ -63.49 \end{bmatrix}$$

<u>Iteration 5:-</u>

$$g^{(4)} = \frac{-2}{3}\left[\begin{pmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{pmatrix}^T \left[\begin{pmatrix} 6 \\ 10 \\ 16 \end{pmatrix} - \begin{pmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{pmatrix}\begin{pmatrix} -13.35198 \\ -63.49 \end{pmatrix}\right]\right]$$

$$g^{(4)} = \begin{pmatrix} -471.3054 \\ -2124.0464 \end{pmatrix}$$

$$\Theta^{(5)} = \Theta^{(4)} - \alpha g^{(4)} = \begin{pmatrix} -13.35198 \\ -63.49 \end{pmatrix} - 0.1 \begin{pmatrix} -471.3054 \\ -2124.0464 \end{pmatrix}$$

$$\Theta^{(5)} = \begin{bmatrix} \theta_0^{(5)} \\ \theta_1^{(5)} \end{bmatrix} = \begin{pmatrix} 33.7786 \\ 148.914 \end{pmatrix}$$