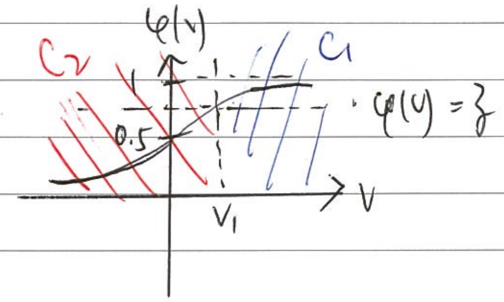
ME5404 ChuaHongWei A0074741E

No.: Homework 1 Q1 Date:

Q1 i) Logistic function $\varphi(v) = \frac{1}{1+e^{-v}}$

Case 1



From the diagram, only one decision boundary is observed at $v = v_1$.

$$\frac{1}{1+e^{-v_1}} = \xi$$

$$1+e^{-v_1} = \frac{1}{\xi}$$

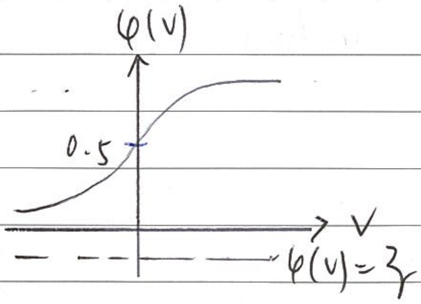
$$\frac{1}{e^{v_1}} = \frac{1}{\xi} + 1$$

$$e^{v_1} = \frac{\xi}{1-\xi}$$

$$v_1 = \ln \frac{\xi}{1-\xi} \quad \text{(constant)}$$

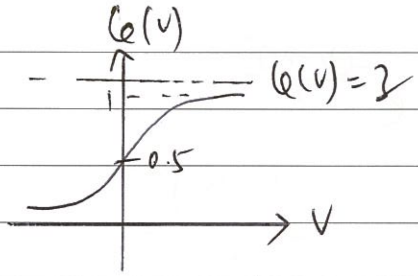
Since v is a constant, there exist a hyperplane on v_1. Any observation vector that ~~want~~ is ~~at~~ more than the hyperplane v_1 is classified as C_1 ~~and~~. Any observation vector that ~~is~~ is less than hyperplane v_1 is classified as C_2.

case 2



if $\xi < 0$, $\varphi(v)$ is always larger than ξ threshold, and will always be classified as C_1. ~~the~~ There is no decision boundary for this case

Case 3

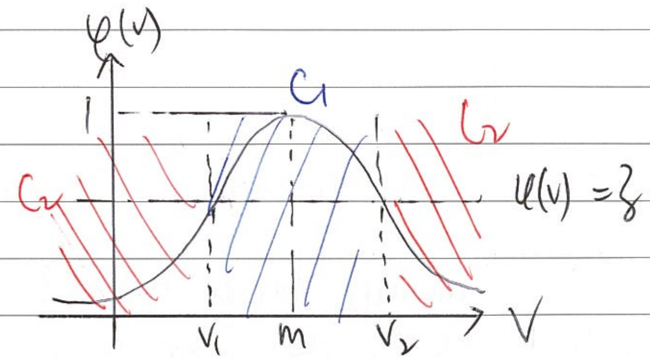


if ~~$\xi > 0$~~ $\xi > 1$, $\varphi(v)$ is always smaller than ξ threshold, and will be classified as C_2. There is no decision boundary for this case.

No.: .... Date: ....

Q1 2) Bell shaped Gaussian function $\varphi(v) = e^{-\frac{(v-m)^2}{2}}$

Case 1



From the diagram, it is observed to have 2 decision boundary at $v = v_1$ and $v = v_2$.

$$\xi = e^{-\frac{(v-m)^2}{2}}$$

$$\ln \xi = -\frac{(v-m)^2}{2}$$

$$(v-m)^2 = -2\ln \xi$$

$$v = m \pm \sqrt{-2\ln \xi}$$

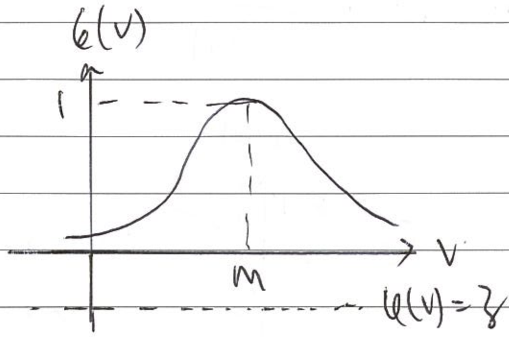
$$v_2 = m + \sqrt{-2\ln \xi} \qquad v_1 = m - \sqrt{-2\ln \xi}$$

~~the~~ The 2 hyperplane exist at v_1 and v_2.

Any observation vector that results in $v_1 < v < v_2$ is classified as C_1, ~~and~~ while observation vector that results in $v > v_2$ or $v < v_1$ are classified as C_2.

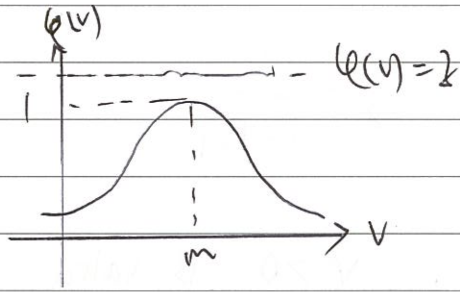
No.: .... Date: ....

Case ~~e~~ 2



if $\zeta < 0$, $\varphi(v)$ is ~~ca~~ always larger than ζ, and will be classified as C1. No decision boundary for this case.

Case 3

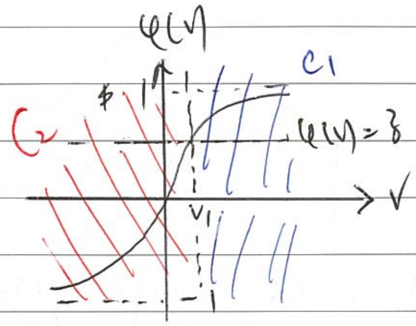


if ~~$\zeta > 0$,~~ $\zeta > 1$, $\varphi(v)$ is always smaller than ζ, and will be classified as C2. No decision boundary ~~t~~ for this case

No.: Date:

Q1 & 3) Softsign function $\varphi(v) = \dfrac{v}{1+|v|}$

case 1



From the diagram, it is observed that one decision boundary exist at $v = v_1$.

So $$\xi = \frac{v}{1+|v|}$$

For $0 < \xi < 1$, $v > 0$ is valid

$$\xi = \frac{v}{1+v}$$

$$\xi(1+v) = v$$

$$\xi + \xi v = v$$

$$\xi = v - \xi v$$

$$v = \frac{\xi}{1-\xi}$$

For $-1 < \xi < 0$, $v < 0$ is valid

$$\xi = \frac{v}{1-v}$$

$$\xi(1-v) = v$$

$$\xi - \xi v = v$$

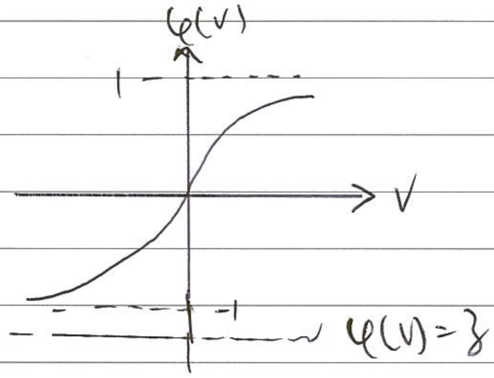
$$\xi = v + \xi v$$

$$v = \frac{\xi}{1+\xi}$$

At each condition of v, there exist only one hyperplane. If $v > 0$, $v_1 = \frac{\xi}{1-\xi}$, while for $v < 0$, $v_1 = \frac{\xi}{1+\xi}$

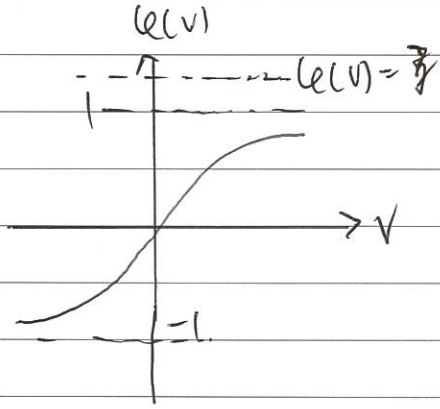
Any observation vector that result in $v > v_1$ is classified as C_1, while any observation vector that result in $v < v_1$ is classified as ~~Case 2~~ C_2

Case 2



if $\xi < -1$, $\varphi(v)$ is always larger than ξ, and will be classified as C_1. There is no decision boundary for this case.

Case 3



if $\xi > 1$, $\varphi(v)$ is always less than ξ, and will be classified as C_2. There is no decision boundary for this case.

Chua HongWei (A0074741E)
Homework 1
Q2)

The truth table for XOR is as follows.

Truth Table of XOR

x_1	0	1	0	1
x_2	0	0	1	1
y	0	1	1	0

By proof by contradiction, assume XOR is linearly separable. For a linearly separable problem, there must be a decision boundary with the following equation

$$w_1 x_1 + w_2 x_2 + b = 0$$

From the truth table, the following four inequalities is derived

$$w_1(0) + w_2(0) + b < 0$$
$$w_1(1) + w_2(0) + b > 0$$
$$w_1(0) + w_2(1) + b > 0$$
$$w_1(1) + w_2(1) + b < 0$$

Which can further be simplified to the following

$$b < 0 \qquad (1)$$
$$w_1 + b > 0 \qquad (2)$$
$$w_2 + b > 0 \qquad (3)$$
$$w_1 + w_2 + b < 0 \qquad (4)$$

Summing up (1) and (4) give us

$$w_1 + w_2 + 2b < 0$$

Summing up (2) and (3) give us

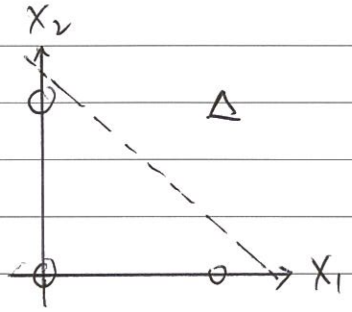
$$w_1 + w_2 + 2b > 0$$

Both inequalities are contradicting and thus prove that XOR is linearly inseparable.

No.: .... Date: ....

Q3 a)

AND

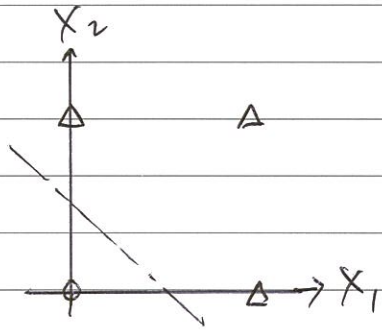


$$X_2 = -X_1 + 1.5$$

$$X_1 + X_2 - 1.5 = 0$$

$$W = [-1.5 \quad 1 \quad 1]$$

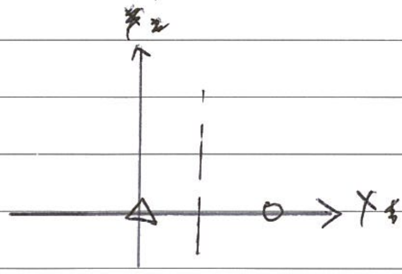
OR



$$X_2 = -X_1 + 0.5$$
$$X_1 + X_2 - 0.5 = 0$$

$$W = [-0.5 \quad 1 \quad 1]$$

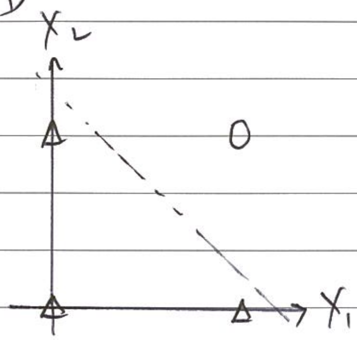
COMPLEMENT



$$X = 0.5$$
$$-X + 0.5 = 0$$

$$W = [0.5 \quad -1]$$

NAND

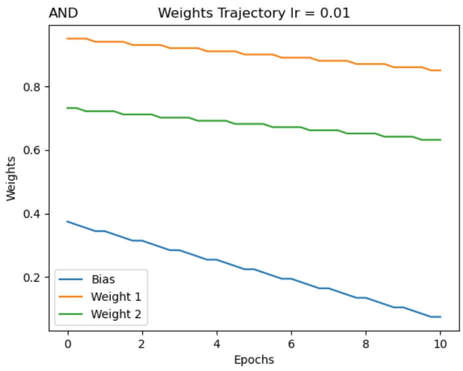


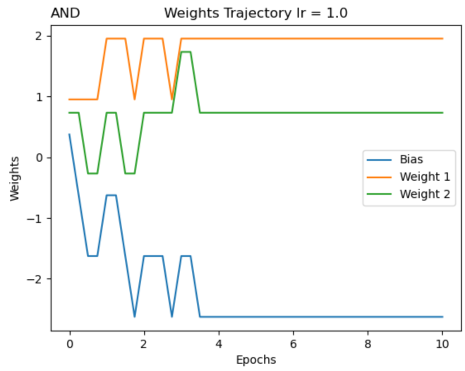
$$X_2 = -X_1 + 1.5$$
$$\cancel{X_1 + X_2 - 1.5 = 0}$$
$$-X_1 - X_2 + 1.5 = 0$$
$$W = [1.5 \quad -1 \quad -1]$$

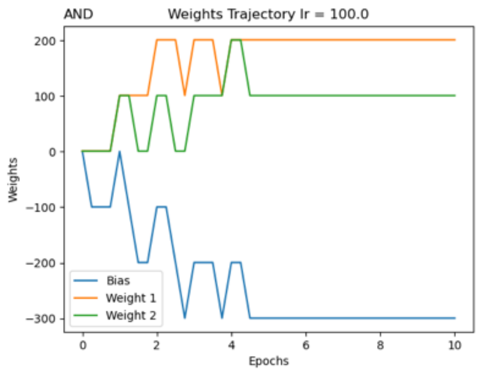
Q3b)

For AND, OR, COMPLEMENT and NAND logic functions, learning rate of 1 is the most suitable for calculating the weights. Small learning rate like 0.01 will slow down the learning and produces weights that are small as seen from the above graph where 10 epoch is insufficient for the weights vector converge. A large learning rate like 100 will speed up the learning but also contributed to large weights vector. All the simulation results and weights vector for AND, OR, COMPLEMENT and NAND logic functions is shown below.

AND



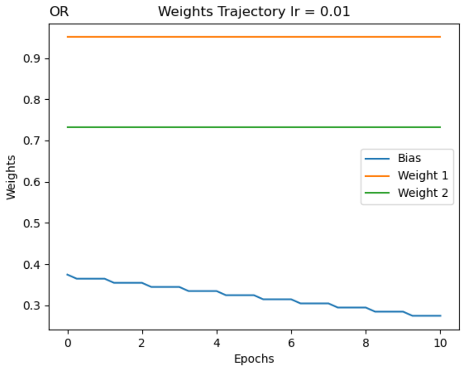


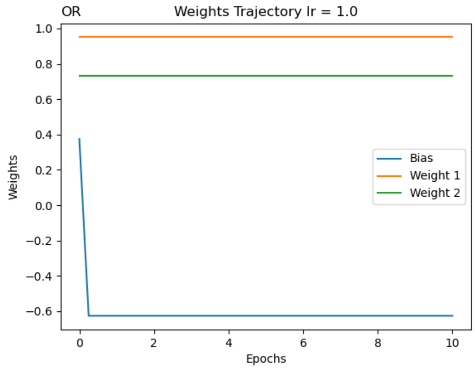


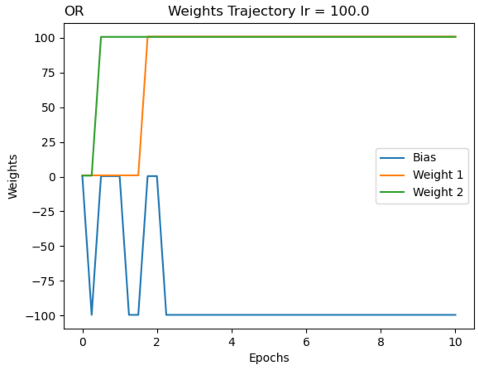
The weights vector after 10 epochs for learning rate of 1 is [-2.6254598811526373, 1.9507143064099162, 0.731993941811405]^T.

For offline calculation, the weights vector is [-1.5 1 1].

OR



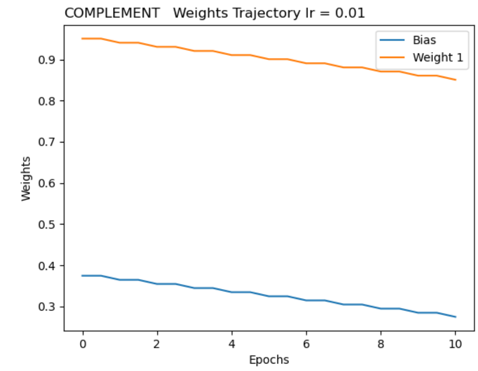


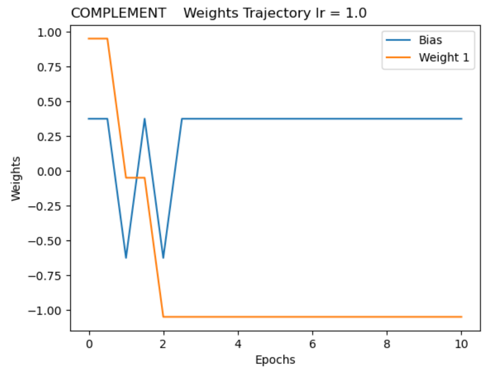


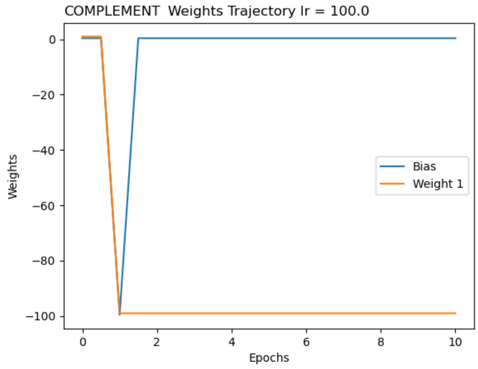
The weights vector after 10 epochs for learning rate of 1 is [-0.6254598811526375, 0.9507143064099162, 0.7319939418114051]^T.

For offline calculation, the weights vector is [-0.5 1 1].

COMPLEMENT



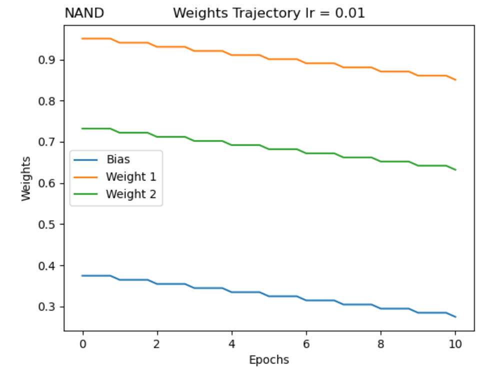


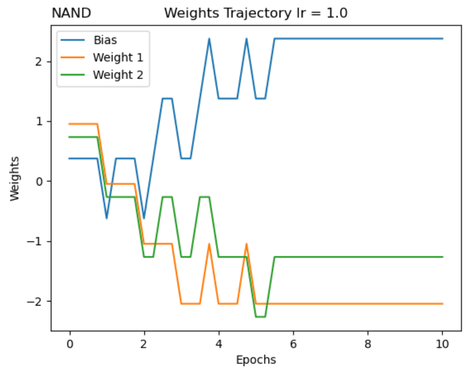


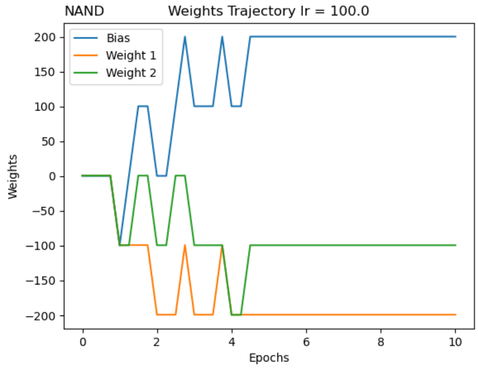
The weights vector after 10 epochs for learning rate of 1 is $[0.3745401188473625, -1.0492856935900838]^T$.

For offline calculation, the weights vector is [0.5 1].

NAND



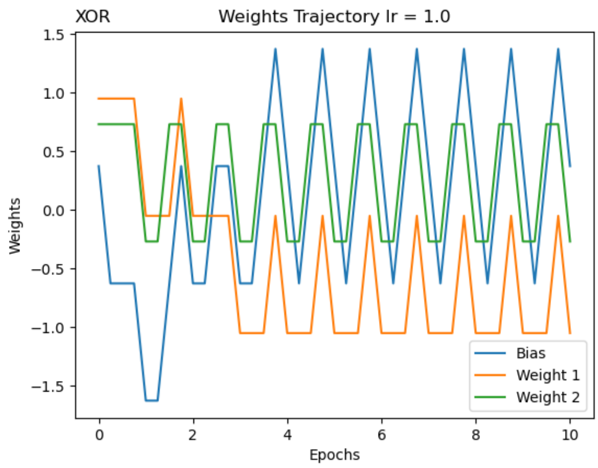




The weights vector after 10 epochs for learning rate of 1 is [2.3745401188473627, -2.049285693590084, -1.268006058188595]^T.

For offline calculation, the weights vector is [1.5 -1 -1].

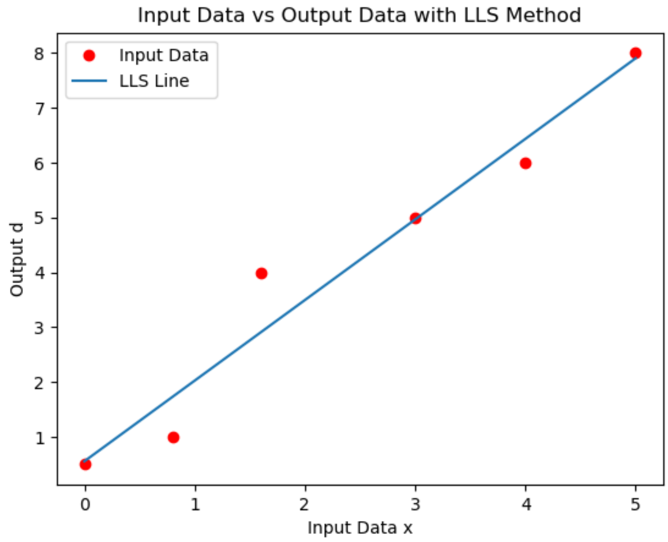
Q3c)



As proven in Question 2, XOR is not linearly separable. As seen in the above figure, the weights vector is unable to converge into a stable value as the weights fluctuates thru each epoch. This prove that perceptron is unable to solve functions that are not linearly separable.

Q4(a)

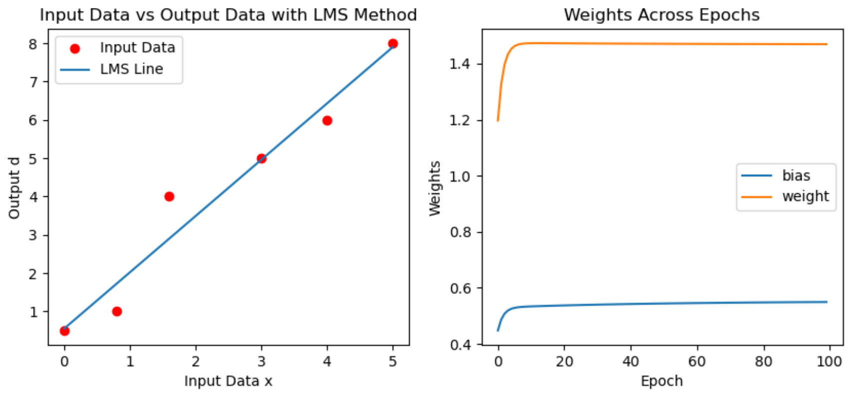
By using the standard linear least squares (LLS), the following plot is obtained.



The bias, b of the LLS is 0.555 and the weight, w of the LLS is 1.47.

Q4(b)

For LMS method with learning rate $\eta = 0.01$, and a randomly chosen weight, the weights converge after around 20 epochs which can be seen from the flatting weights vs epoch curve after 20 epochs.



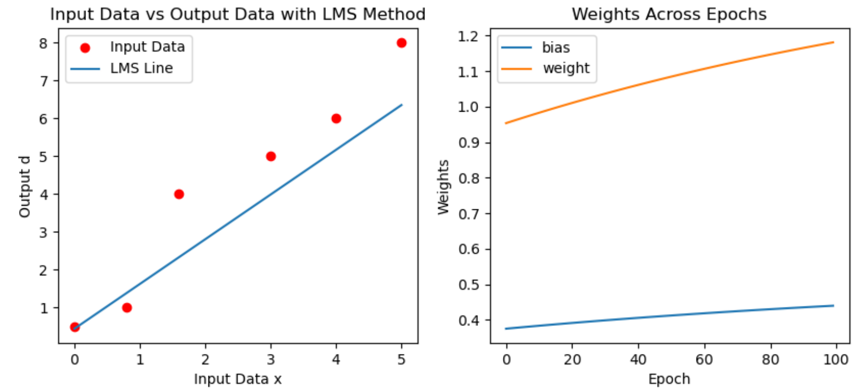
Q4(c)

The weights and bias obtained from LLS method is [0.555 1.47], while the weights and bias obtained from LMS method is [0.549 1.469]. Both methods converge to similar weights and bias.

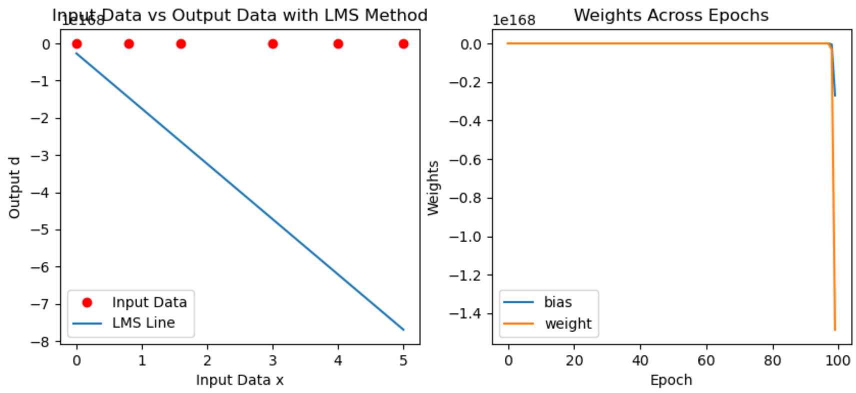
Q4(d)

For this question, the learning rates of 0.0001 and 0.5 were plotted. The findings are as follows.

Learning rate = 0.0001



Learning rate = 0.5



With low learning rate, 100 epochs are insufficient for the weights to converge. The LMS line produced is not a good fit to the data. The weights would require more epoch to converge, and the process might be slow.

With high learning rate, the weights show a weird behavior with the regression line diverging away from the data points.

Q5

$$J = \frac{1}{2} R^T R e^T e + \frac{1}{2} \lambda w^T w$$

$$y = w^T x = x^T w$$

Let X be the regression matrix where

$$X = \begin{bmatrix} x(1)^T \\ \vdots \\ x(n)^T \end{bmatrix}$$

$$e = d - Xw$$

First let's calculate $\frac{\partial e}{\partial \mathrm{w}}$

$$\frac{\partial e}{\partial w} = -X$$

Now let's calculate $\frac{\partial J}{\partial w}$ for first term

By chain rule,

$$\frac{\partial J}{\partial w} = \frac{\partial J}{\partial e}\frac{\partial e}{\partial w} = ((\frac{1}{2})R^T R 2e^T)(-X) = -R^T R e^T X$$

Now let's calculate $\frac{\partial J}{\partial w}$ for second term

$$\frac{\partial J}{\partial w} = \lambda w^T$$

This brings us to the following

$$\frac{\partial J}{\partial w} = -R^T R e^T X + \lambda w^T$$

To minimize J(w)

$$\frac{\partial J}{\partial w} = -R^T R (d - Xw)^T X + \lambda w^T = 0$$

$$-R^T R d^T X + R^T R X^T X w^T + \lambda w^T = 0$$

$$w^T (R^T R X^T X + \lambda I) = R^T R d^T X$$

$$w (R^T R X^T X + \lambda I) = R^T R X^T d$$

$$w = (R^T R X^T X + \lambda I)^{-1} R^T R X^T d$$