HW1

Q1

1. $\varphi(v) = (v-a)^2 + \text{c}$

The decision boundary can be defined by:

$$\xi = (v-a)^2 + c$$
$$v = a \pm \sqrt{\xi - c}$$

v is the induced field: $v = \sum_{i-1}^{m} w_i x_i + b$

$$a \pm \sqrt{\xi - c} = \sum_{i-1}^{m} w_i x_i + b$$

Not a hyper plane. This is a non-linear decision boundary because $\sqrt{\xi - c}$ isdepends on ξ and c.

2. $\varphi(v) = \frac{1-e^{-v}}{1+e^{-v}}$

$$\xi = \frac{1-e^{-v}}{1+e^{-v}}$$
$$v = -ln(\frac{1=\xi}{1+\xi})$$

$$-ln(\frac{1=\xi}{1+\xi}) = \sum_{i-1}^{m} w_i x_i + b$$

This is a hyper plane.

3. $\varphi(v) = e^{-\frac{(v-m)^2}{2\sigma^2}}$

$$\xi = e^{-\frac{(v-m)^2}{2\sigma^2}}$$
$$v = m \pm \sqrt{-2\sigma^2 ln(\xi)}$$

$$\sum_{i-1}^{m} w_i x_i + b = m \pm \sqrt{-2\sigma^2 ln(\xi)}$$

Not a hyper plane. Because the square root results in a non-linear relationship between the input x_i.

Q2

$$\mathrm{XOR}\left(x_1, x_2\right)=\left\{\begin{array}{ll} 0 & \text{if } x_1=x_2 \\ 1 & \text{if } x_1 \neq x_2 \end{array}\right.$$

Assume there is a linear function:

$$y = w_1 x_1 + w_2 x_2 + b,$$

which satisfies:

$$\begin{array}{ll} w_1(0)+w_2(0)+b<0 & \text{for } (0,0) \rightarrow 0 \\ w_1(0)+w_2(1)+b>0 & \text{for } (0,1) \rightarrow 1 \\ w_1(1)+w_2(0)+b>0 & \text{for } (1,0) \rightarrow 1 \\ w_1(1)+w_2(1)+b<0 & \text{for } (1,1) \rightarrow 0 \end{array}$$

Simplify:

$$\begin{gathered} b<0 \\ w_2+b>0 \\ w_1+b>0 \\ w_1+w_2+b<0 \end{gathered}$$

From the second and third inequation, we have

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

which contradicts forth inequation.

Therefore, the assumption is false. There is no linear function that can separate XOR.

Q3

a)

1. AND

$$
\begin{gathered}
y = \varphi(w_1 x_1 + w_2 x_2 + b) \\
w_1 = 1 \\
w_2 = 1 \\
b = -1.5
\end{gathered}
$$

2. OR

$$
\begin{gathered}
y = \varphi(w_1 x_1 + w_2 x_2 + b) \\
w_1 = 1 \\
w_2 = 1 \\
b = -0.5
\end{gathered}
$$

3. COMPLEMENT

$$
\begin{gathered}
y = \varphi(w_1 x_1 + b) \\
w_1 = -1 \\
b = 0.5
\end{gathered}
$$

4. NAND

$$
\begin{gathered}
y = \varphi(w_1 x_1 + w_2 x_2 + b) \\
w_1 = -1 \\
w_2 = -1 \\
b = 1.5
\end{gathered}
$$

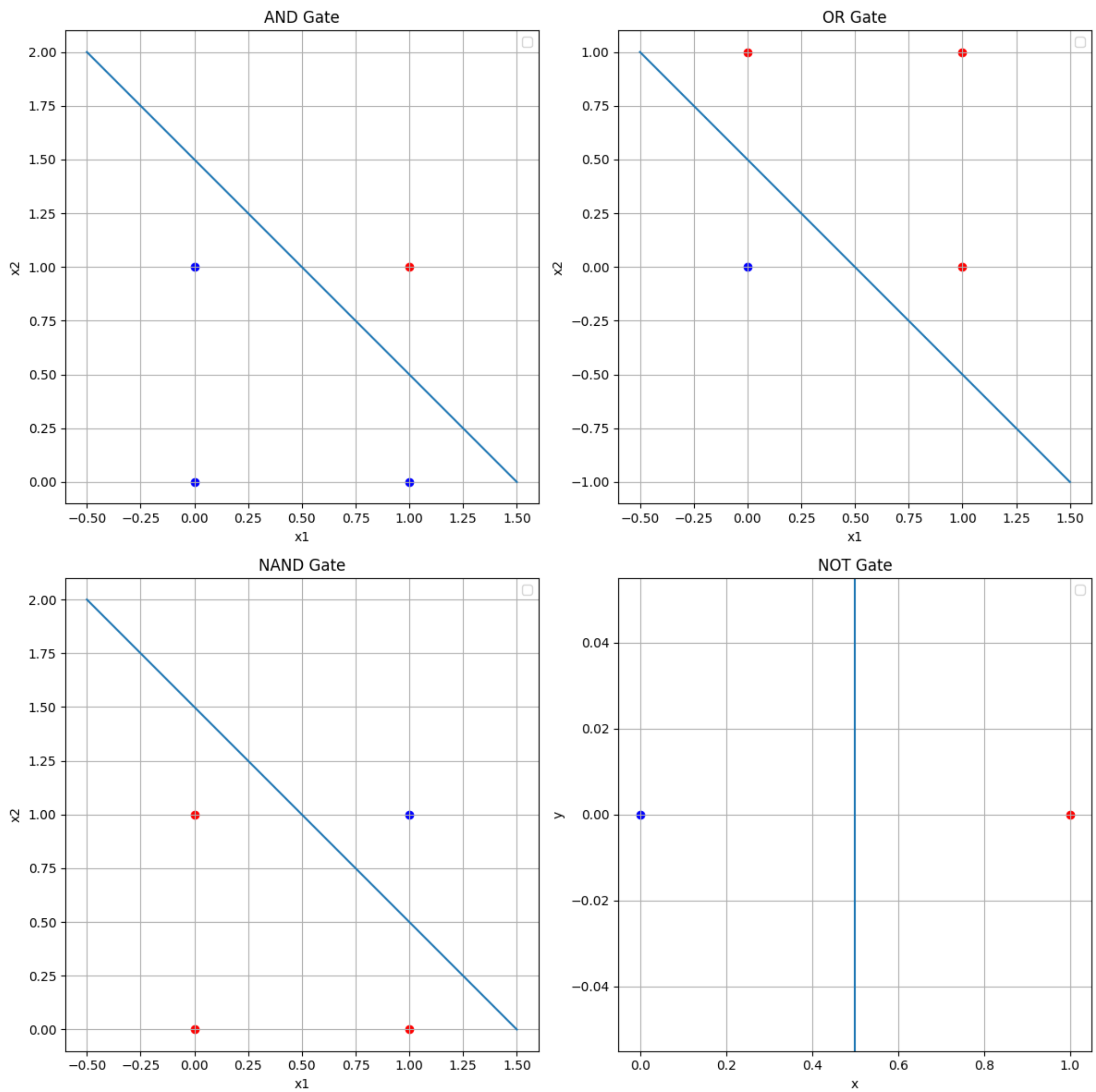
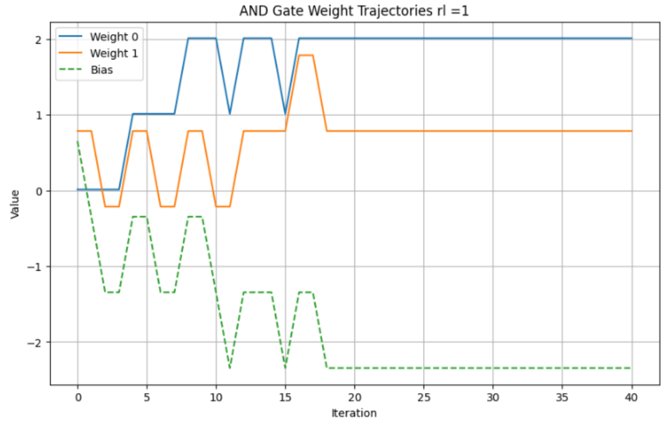
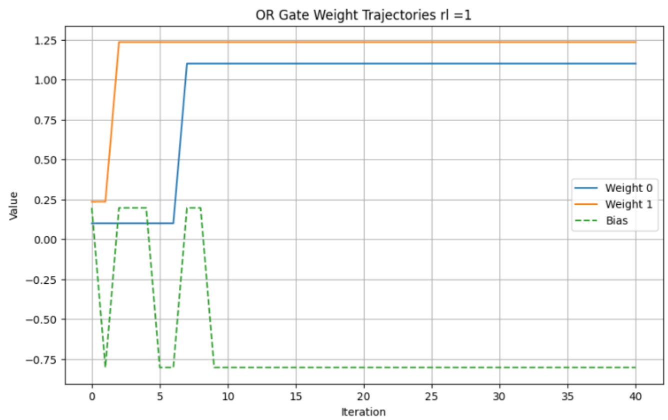


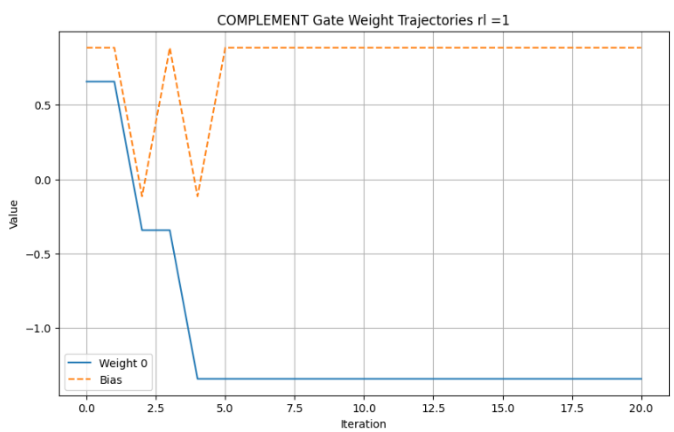
Fig. 1 Off-line Calculations (blue line is the boundary, red and blue dots represent 1 and 1 respectively)

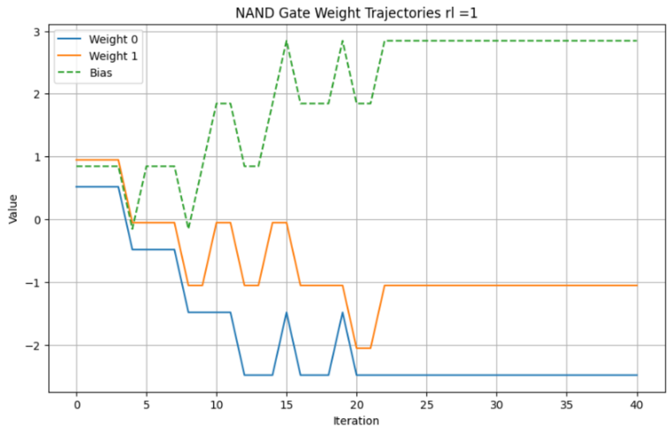
b)

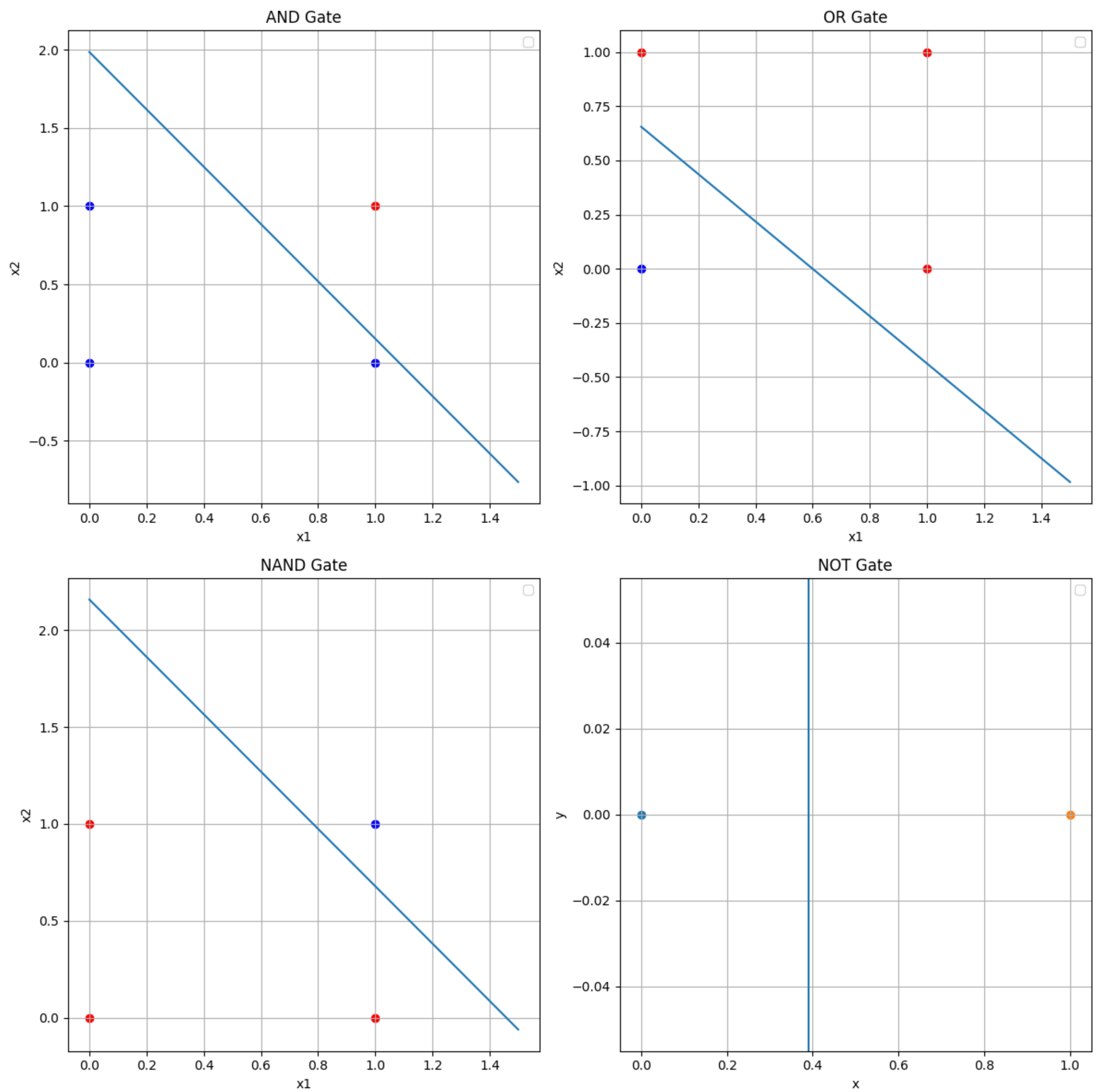
The figures below show the trajectories of parameters with learning rate = 1.


AND Gate Weight Trajectories rl =1
Weight 0
Weight 1
Bias
Value
Iteration

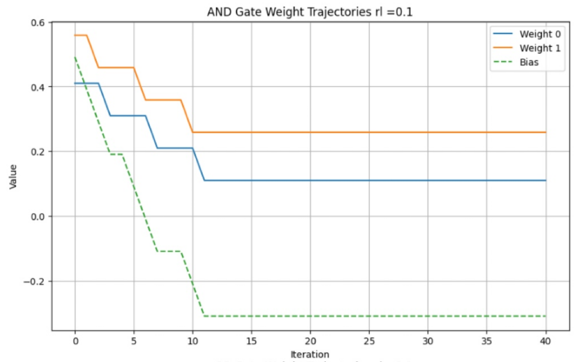

OR Gate Weight Trajectories rl =1
Weight 0
Weight 1
Bias
Value
Iteration

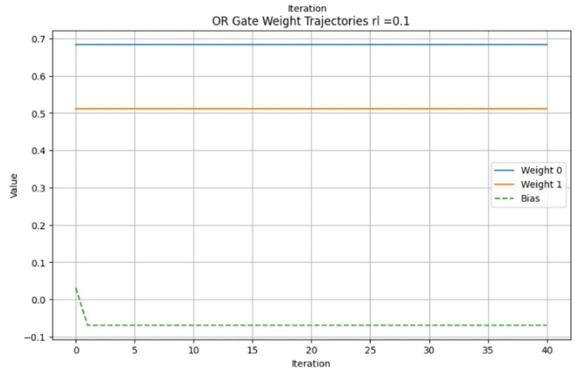

COMPLEMENT Gate Weight Trajectories rl =1
Weight 0
Bias
Value
Iteration

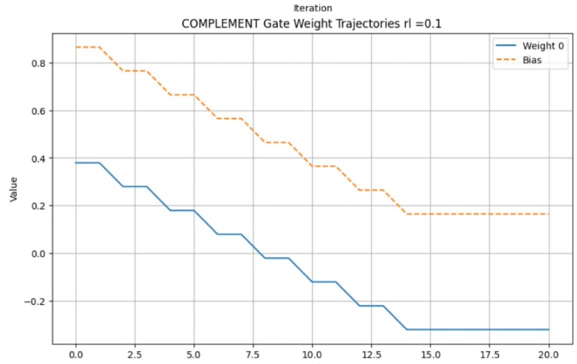

NAND Gate Weight Trajectories rl =1
Weight 0
Weight 1
Bias
Value
Iteration
3
2
1
0
−1
−2
0
5
10
15
20
25
30
35
40

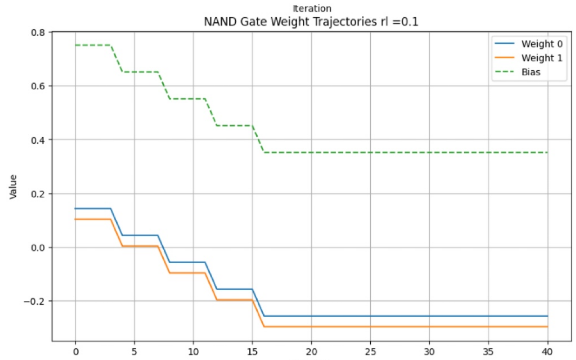


When learning rate = 0.1

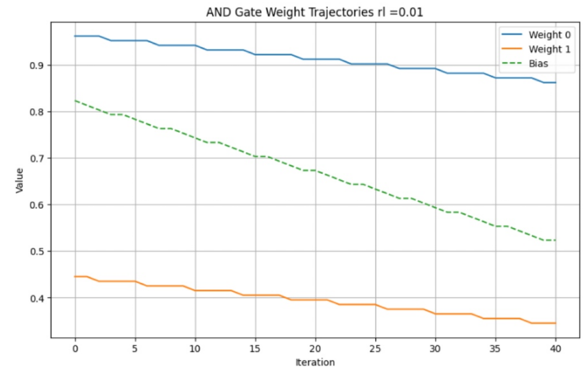


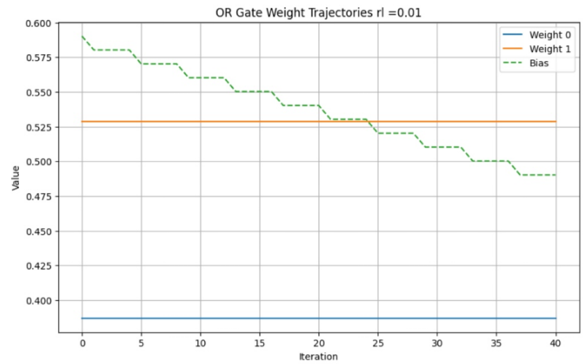


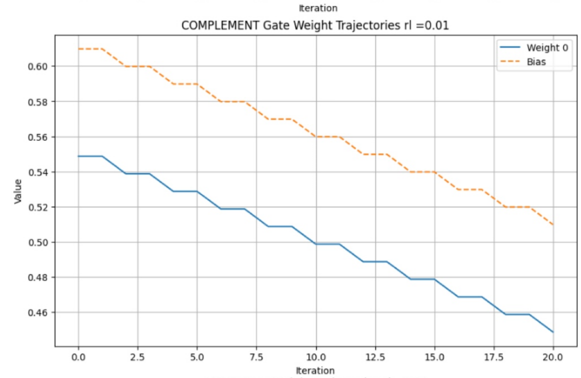


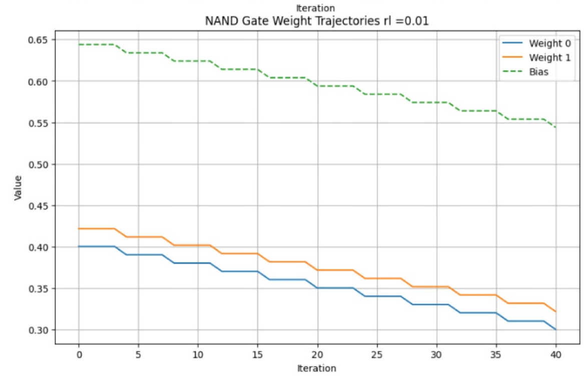


When learning rate = 0.01



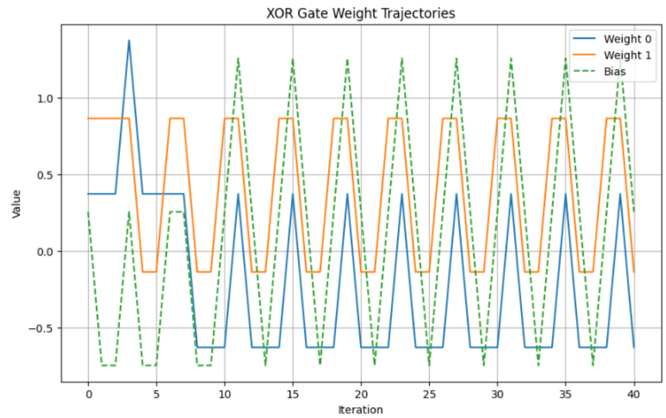






Through comparison, we can intuitively find that a smaller learning rate will slow down the parameter update, so we need more epochs to iterate the parameters, while a larger learning rate can improve the training efficiency. However, if the learning rate is too large, it may lead to missing the optimal solution.

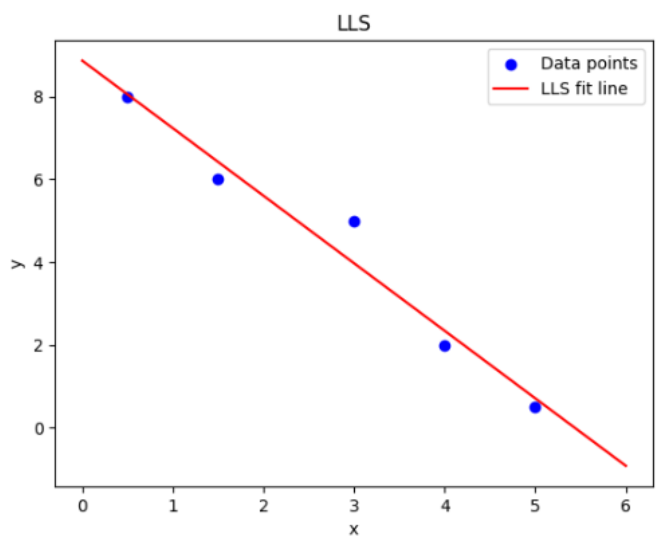
c)



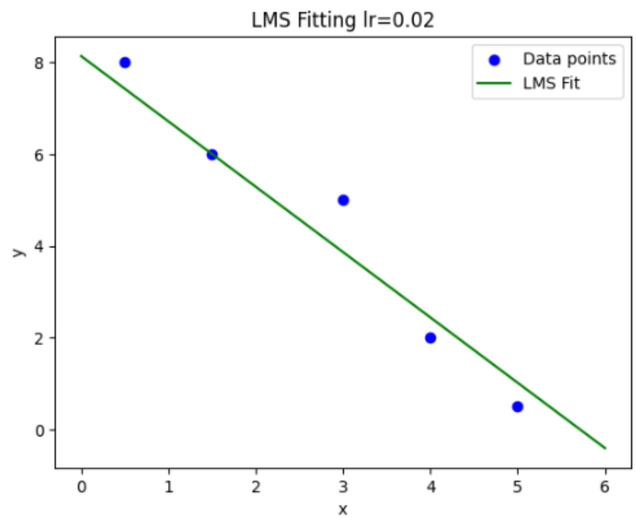
We find that the perceptron cannot solve the XOR problem, which makes the parameters always oscillate and will not converge.

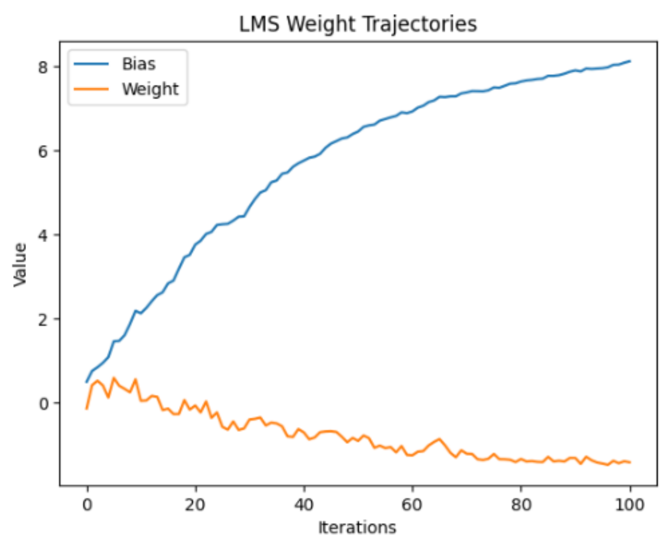
Q4

a)

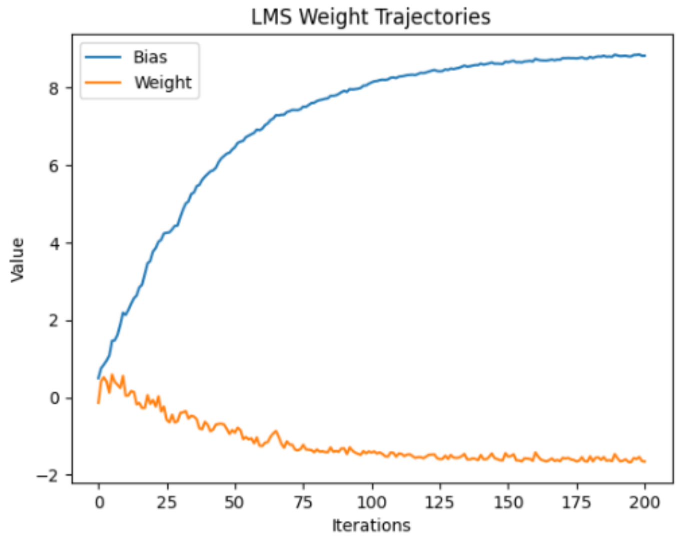


b)





We found that weights failed to converge in the condition of 100 epochs, so we conducted further experiments and started to converge at 150 epochs.

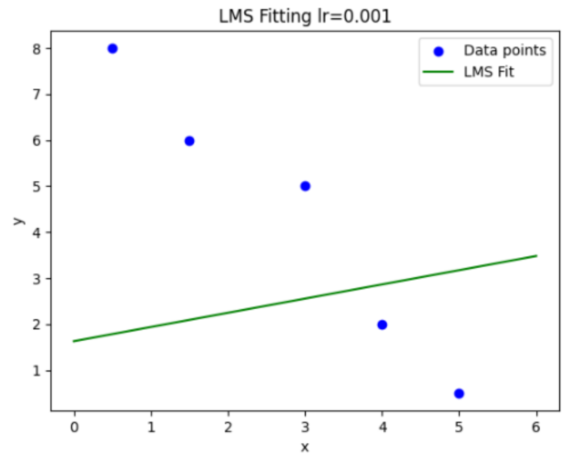


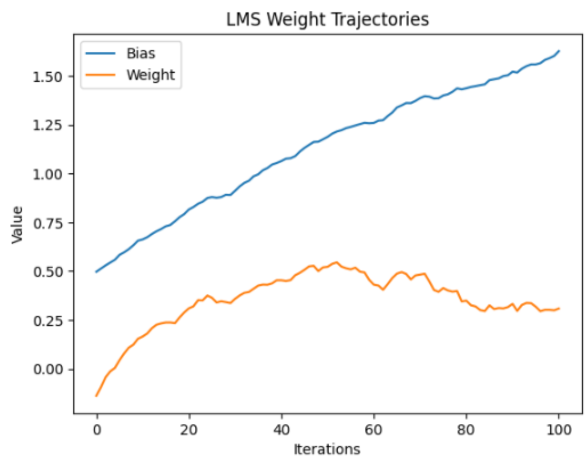
c)

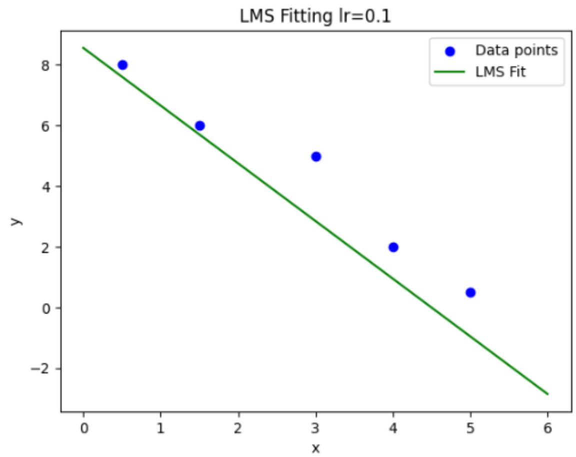
```
LLS Solution (Bias, Weight): [ 8.86842105 -1.63157895]
LMS Solution (Bias, Weight): [ 8.13020857 -1.42219745]
```

In this case, LLS and LMS performance similarly. LLS is solved directly, the result is stable, and there is no need to worry about convergence. LMS cannot guarantee to reach the global optimal solution in all cases, especially if the initial weights are poorly chosen or the learning rate is inappropriate.

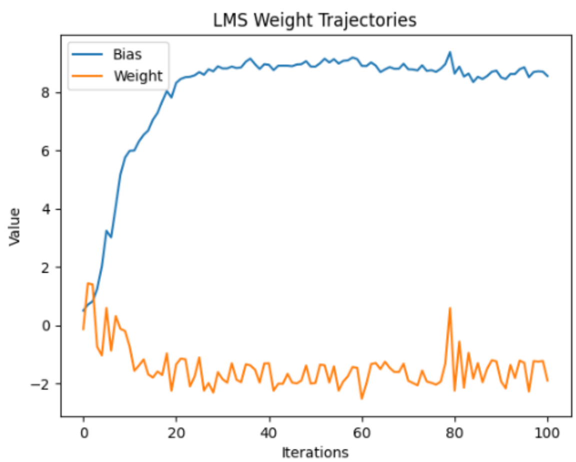
d)


LMS Fitting lr=0.001
Data points
LMS Fit
y
x
0
1
2
3
4
5
6
7
8

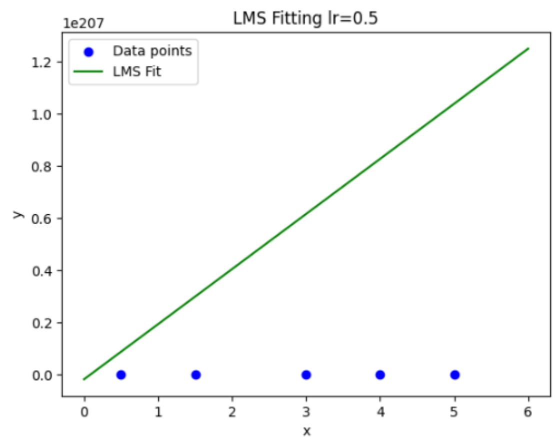

LMS Weight Trajectories
Bias
Weight
Value
Iterations
0.00
0.25
0.50
0.75
1.00
1.25
1.50
0
20
40
60
80
100

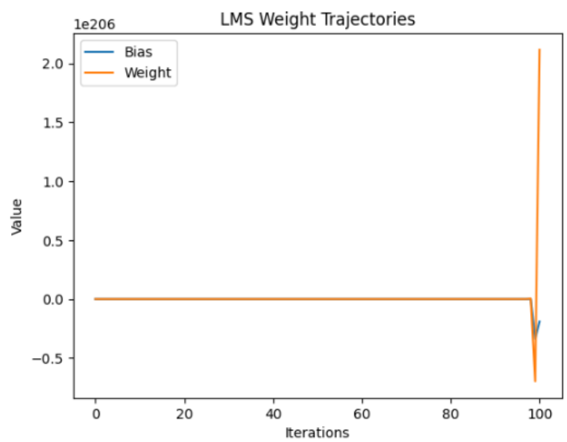


LMS Fitting lr=0.1
Data points
LMS Fit
y
x
8
6
4
2
0
−2
0
1
2
3
4
5
6



LMS Weight Trajectories
Bias
Weight
Value
Iterations
8
6
4
2
0
−2
0
20
40
60
80
100





Both too large and too small learning rates lead to inaccurate results. If the learning rate is too small, the optimal solution is not obtained under the condition of 100 epochs. However, too high a learning rate will lead to failure to converge.

Q5

$$\frac{\partial E(n)}{\partial w(n)} = \frac{\partial}{\partial (n)}(\frac{1}{2}e^2 n) + \frac{\partial}{\partial w(n)}(\frac{\lambda}{2}||w(n)||^2)$$
$$= -e(n)x(n) + \lambda w(n)$$

$$w(n+1) = w(n) - \eta(-e(n)x(n) + \lambda w(n))$$

$$w(n+1) = (1 - \eta\lambda)w(n) + \eta e(n)x(n)$$