MEMC-5173

Assignment #4
Reference solutions

Q.1
(Question 5-5 in Book 2)

To find the gradient of Equation (5.25), first we must expand it:

$$\begin{aligned} E_{\mathbf{W}}(\boldsymbol{\theta}) &= (\mathbf{y}-\mathbf{A}\boldsymbol{\theta})^T\mathbf{W}(\mathbf{y}-\mathbf{A}\boldsymbol{\theta}) \\ &= (\mathbf{y}^T-\boldsymbol{\theta}^T\mathbf{A}^T)(\mathbf{W}\mathbf{y}-\mathbf{W}\mathbf{A}\boldsymbol{\theta}) \\ &= \mathbf{y}^T\mathbf{W}\mathbf{y}-\mathbf{y}^T\mathbf{W}\mathbf{A}\boldsymbol{\theta}-\boldsymbol{\theta}^T\mathbf{A}^T\mathbf{W}\mathbf{y}+\boldsymbol{\theta}^T\mathbf{A}^T\mathbf{W}\mathbf{A}\boldsymbol{\theta} \\ &= \mathbf{y}^T\mathbf{W}\mathbf{y}-2\boldsymbol{\theta}^T\mathbf{A}^T\mathbf{W}\mathbf{y}+\boldsymbol{\theta}^T\mathbf{A}^T\mathbf{W}\mathbf{A}\boldsymbol{\theta}. \end{aligned}$$

By using the identities in Equations (5.2) and Equation (5.6), we can find the gradient of the preceding equation and set it to zero when $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}$:

$$\nabla E_{\mathbf{W}}(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -2\mathbf{A}^T\mathbf{W}\mathbf{y}+2\mathbf{A}^T\mathbf{W}\mathbf{A}\hat{\boldsymbol{\theta}}=0.$$

Therefore the weighted LSE is

$$\hat{\boldsymbol{\theta}} = (\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{y}.$$

Q. 2
Redo Example 4.1 in Book 1 but with $\boldsymbol{W}^{(1)} = [1, -1, 0, -1]^T$, $\eta = 0.5$

with $\boldsymbol{W}^{(1)} = [1, -1, 0, -1]^T$, $\eta = 0.5$

Suppose that learning rate $\eta = 0.5$.
In the example 1, we have $\theta^{(1)} = w_4^{(1)} = 0.5$, whereas here we have $\theta^{(1)} = w_4^{(1)} = -1$.

Epoch 1:
Introducing the first input vector $x^{(1)}$ to the network, we get:

$$o^{(1)} = \operatorname{sgn}(w^{(1)^T}x^{(1)}) = \operatorname{sgn}\left(\begin{bmatrix}1,-1,0,-1\end{bmatrix}\begin{bmatrix}1\\-2\\0\\-1\end{bmatrix}\right) = +1 \neq t^{(1)}$$

$$w^{(2)} = w^{(1)} + \eta\left[t^{(1)} - o^{(1)}\right]x^{(1)} = w^{(1)} + 0.5(-2)x^{(1)} = \left[0,1,0,0\right]^{(T)}$$

Introducing the second input vector $x^{(2)}$ to the network, we get:

$$o^{(2)} = \text{sgn}(w^{(2)^T} x^{(2)}) = \text{sgn}\left([0,1,0,0] \begin{bmatrix} 0 \\ 1.5 \\ -0.5 \\ -1 \end{bmatrix} \right) = 1 \neq t^{(2)}$$

$$w^{(3)} = w^{(2)} + \eta \left[t^{(2)} - o^{(2)} \right] x^{(2)} = w^{(2)} + 0.5(-2)x^{(2)} = [0,-0.5,0.5,1]^{(T)}$$

Introducing the third input vector $x^{(3)}$ to the network, we get:

$$o^{(3)} = \text{sgn}(w^{(3)^T} x^{(3)}) = \text{sgn}\left([0,-0.5,0.5,1] \begin{bmatrix} -1 \\ 1 \\ 0.5 \\ -1 \end{bmatrix} \right) = -1 \neq t^{(3)}$$

$$w^{(4)} = w^{(3)} + \eta \left[t^{(3)} - o^{(3)} \right] x^{(3)} = w^{(3)} + 0.5(2)x^{(3)} = [-1,+0.5,1,0]^{(T)}$$

Epoch 2:

We reuse here the training set $(x^{(1)}, t^{(1)}), (x^{(2)}, t^{(2)}), (x^{(3)}, t^{(3)})$ as $(x^{(4)}, t^{(4)}), (x^{(5)}, t^{(5)}), (x^{(6)}, t^{(6)})$, respectively.

Introducing the first input vector $x^{(4)}$ to the network, we get:

$$o^{(4)} = \text{sgn}(w^{(4)^T} x^{(4)}) = \text{sgn}\left([-1,0.5,1,0] \begin{bmatrix} 1 \\ -2 \\ 0 \\ -1 \end{bmatrix} \right) = -1 = t^{(4)}$$

$$w^{(5)} = w^{(4)}$$

Introducing the second input vector $x^{(5)}$ to the network, we get:

$$o^{(5)} = \text{sgn}(w^{(5)^T} x^{(5)}) = \text{sgn}\left([-1,0.5,1,0] \begin{bmatrix} 0 \\ 1.5 \\ -0.5 \\ -1 \end{bmatrix} \right) = +1 \neq t^{(5)}$$

$$w^{(6)} = w^{(5)} + \eta \left[t^{(5)} - o^{(5)} \right] x^{(5)} = w^{(5)} - 0.5(2)x^{(5)} = [-1,-1,1.5,1]^{(T)}$$

Introducing the third input vector $x^{(6)}$ to the network, we get:

$$o^{(6)} = \text{sgn}(w^{(6)^T} x^{(6)}) = \text{sgn}\left(\left[-1,-1,1.5,1\right] \begin{bmatrix} -1 \\ 1 \\ 0.5 \\ -1 \end{bmatrix} \right) = -1 \neq t^{(6)}$$

$$w^{(7)} = w^{(6)} + \eta\left[t^{(6)} - o^{(6)}\right]x^{(6)} = w^{(6)} + 0.5(2)x^{(6)} = \left[-2,0,2,0\right]^{(T)}$$

Epoch 3:
We reuse here the training set $(x^{(1)},t^{(1)}),(x^{(2)},t^{(2)}),(x^{(3)},t^{(3)})$ as $(x^{(7)},t^{(7)}),(x^{(8)},t^{(8)}),(x^{(9)},t^{(9)})$, Respectively.
Introducing the first input vector $x^{(7)}$ to the network, we get:

$$o^{(7)} = \text{sgn}(w^{(7)^T} x^{(7)}) = \text{sgn}\left(\left[-2,0,2,0\right] \begin{bmatrix} 1 \\ -2 \\ 0 \\ -1 \end{bmatrix} \right) = -1 = t^{(7)}$$

$$w^{(8)} = w^{(7)}$$

Introducing the second input vector $x^{(8)}$ to the network, we get:

$$o^{(8)} = \text{sgn}(w^{(8)^T} x^{(8)}) = \text{sgn}\left(\left[-2,0,2,0\right] \begin{bmatrix} 0 \\ 1.5 \\ -0.5 \\ -1 \end{bmatrix} \right) = -1 = t^{(8)}$$

$$w^{(9)} = w^{(8)}$$

Introducing the third input vector $x^{(9)}$ to the network, we get:

$$o^{(9)} = \text{sgn}(w^{(9)^T} x^{(9)}) = \text{sgn}\left(\left[-2,0,2,0\right] \begin{bmatrix} -1 \\ 1 \\ 0.5 \\ -1 \end{bmatrix} \right) = 1 = t^{(9)}$$

$$w^{(10)} = w^{(9)}$$

Introducing the input vectors for another epoch will result no change to the weights, which indicates $w^{(10)}$ is the solution of this problem; that is

$$w_1 = -2 \;,\; w_2 = 0 \;,\; w_3 = 2 \;\text{ And }\; w_4 = 0 \text{ or } \theta = 0$$

And boundary line is equal to:

$$-2x_1 + 2x_3 = 0 \quad \text{OR} \quad -x_1 + x_3 = 0$$

Q 3

Question 4.3(a) in Book 1.

Suppose that $w^{(1)} = [-1 \quad 1]^T$, $\theta^{(1)} = -1$

Epoch 1:

Introducing $(T, -1)$ as vector $(x^{(1)}, t^{(1)})$ to the network, we get:

$$o^{(1)} = \text{sgn}(w^{(1)^T} x^{(1)} - \theta^{(1)}) = \text{sgn}\left([-1, 1]\begin{bmatrix} 3 \\ 0 \end{bmatrix} - (-1) \right) = -1 = t^{(1)}$$

$$w^{(2)} = w^{(1)} \ , \ \theta^{(2)} = \theta^{(1)}$$

Introducing $(U, -1)$ as vector $(x^{(2)}, t^{(2)})$ to the network, we get:

$$o^{(2)} = \text{sgn}(w^{(2)^T} x^{(2)} - \theta^{(2)}) = \text{sgn}\left([-1, 1]\begin{bmatrix} 2 \\ 1.5 \end{bmatrix} - (-1) \right) = 1 \neq t^{(2)}$$

$$w^{(3)} = w^{(2)} + \eta\left[t^{(2)} - o^{(2)}\right]x^{(2)} = w^{(2)} + 0.5(-2)x^{(2)} = [-3, -0.5]^{(T)}$$

$$\theta^{(3)} = \theta^{(2)} - 2\eta t^{(2)} = -1 - 2(0.5)(-1) = 0$$

Introducing $(V, -1)$ as vector $(x^{(3)}, t^{(3)})$ to the network, we get:

$$o^{(3)} = \text{sgn}(w^{(3)^T} x^{(3)} - \theta^{(3)}) = \text{sgn}\left([-3, -0.5]\begin{bmatrix} 1 \\ 3 \end{bmatrix} - 0 \right) = -1 = t^{(3)}$$

$$w^{(4)} = w^{(3)}$$

$$\theta^{(4)} = \theta^{(3)}$$

Introducing $(X, 1)$ as vector $(x^{(4)}, t^{(4)})$ to the network, we get:

$$o^{(4)} = \text{sgn}(w^{(4)^T} x^{(4)} - \theta^{(4)}) = \text{sgn}\left([-3, -0.5]\begin{bmatrix} -1 \\ 0 \end{bmatrix} - 1 \right) = 1 = t^{(4)}$$

$$w^{(5)} = w^{(4)} \ , \ \theta^{(5)} = \theta^{(4)}$$

Introducing $(Y, 1)$ as vector $(x^{(5)}, t^{(5)})$ to the network, we get:

$$o^{(5)} = \text{sgn}(w^{(5)^T} x^{(5)} - \theta^{(5)}) = \text{sgn}\left([-3, -0.5]\begin{bmatrix} -2 \\ 0 \end{bmatrix} \right) = 1 = t^{(5)}$$

$$w^{(6)} = w^{(5)} \ , \ \theta^{(6)} = \theta^{(5)}$$

Introducing $(Z, 1)$ as vector $(x^{(6)}, t^{(6)})$ to the network, we get:

$$o^{(6)} = \text{sgn}(w^{(6)^T} x^{(6)} - \theta^{(6)}) = \text{sgn}\left([-3, -0.5]\begin{bmatrix} -1 \\ 2 \end{bmatrix} \right) = 1 = t^{(6)}$$

$$w^{(7)} = w^{(6)} \ , \ \theta^{(7)} = \theta^{(6)}$$

Epoch 2:

Introducing $(T,-1)$ as vector $(x^{(7)},t^{(7)})$ to the network, we get:

$$o^{(7)}=\text{sgn}(w^{(7)^T}x^{(7)}-\theta^{(7)})=\text{sgn}\left([-3,-0.5]\begin{bmatrix}3\\0\end{bmatrix}\right)=-1=t^{(7)}$$

$$w^{(8)}=w^{(7)}\ ,\ \theta^{(8)}=\theta^{(7)}$$

Introducing $(U,-1)$ as vector $(x^{(8)},t^{(8)})$ to the network, we get:

$$o^{(8)}=\text{sgn}(w^{(8)^T}x^{(8)}-\theta^{(8)})=\text{sgn}\left([-3,-0.5]\begin{bmatrix}2\\1.5\end{bmatrix}\right)=-1=t^{(8)}$$

$$w^{(9)}=w^{(8)}\ ,\ \theta^{(9)}=\theta^{(8)}$$

Introducing $(V,-1)$ as vector $(x^{(9)},t^{(9)})$ to the network, we get:

$$o^{(9)}=\text{sgn}(w^{(9)^T}x^{(9)}-\theta^{(9)})=\text{sgn}\left([-3,-0.5]\begin{bmatrix}1\\3\end{bmatrix}\right)=-1=t^{(4)}$$

$$w^{(10)}=w^{(9)},\ \theta^{(10)}=\theta^{(9)}$$

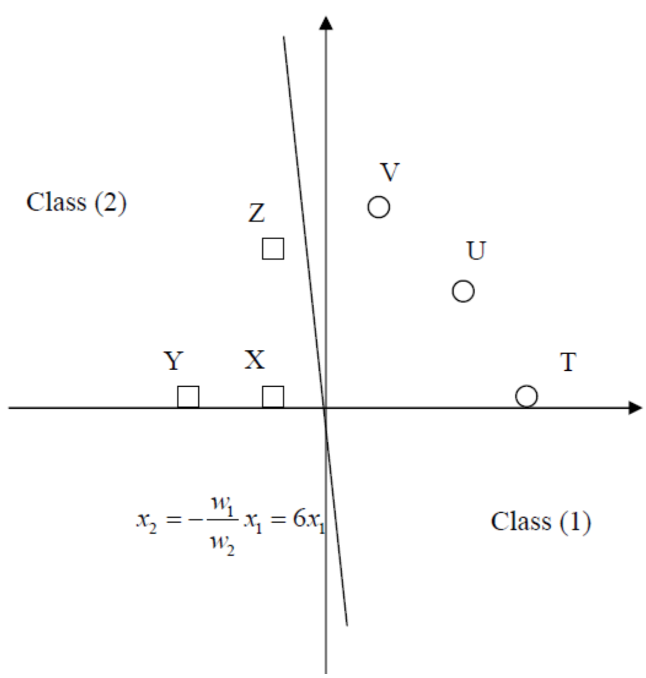
Introducing $(X,1)$ as vector $(x^{(10)},t^{(10)})$ to the network, we get:

$$o^{(10)}=\text{sgn}(w^{(10)^T}x^{(10)}-\theta^{(10)})=\text{sgn}\left([-3,-0.5]\begin{bmatrix}-1\\0\end{bmatrix}\right)=1=t^{(10)}$$

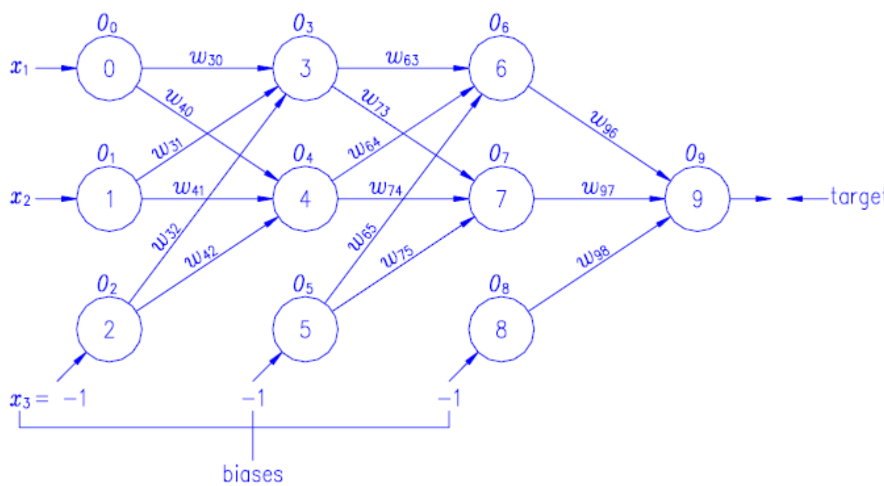
$$w^{(11)}=w^{(10)}\ ,\ \theta^{(10)}=\theta^{(10)}$$

As such, the boundary for classifying the patterns is as follows:

$$x_2=-6x_1$$



Q 4.
Book 1:
Use the same training patterns as those in Example 5.1. Assume all initial link weights are 0.5, learning rate $\eta = 0.2$, and maximum tolerable error $E_{max} = 0.01$. Training should be undertaken over at least 2 epochs.



Step 1: Initialize weights and thresholds to small random values

Step 2: Apply input pattern

$$o_0 = x_1$$
$$o_1 = x_2$$
$$o_2 = x_3 = -1$$

Step 3: Forward propagation

$$O_3 = f(w_{30}o_0 + w_{31}o_1 + w_{32}o_2) = f(w_{30}o_0 + w_{31}o_1 - w_{32})$$
$$O_4 = f(w_{40}o_0 + w_{41}o_1 + w_{42}o_2) = f(w_{40}o_0 + w_{41}o_1 - w_{42})$$
$$O_5 = -1$$
$$O_6 = f(w_{63}o_3 + w_{64}o_4 + w_{65}o_5) = f(w_{63}o_3 + w_{64}o_4 - w_{65})$$
$$O_7 = f(w_{73}o_3 + w_{74}o_4 + w_{75}o_5) = f(w_{73}o_3 + w_{74}o_4 - w_{75})$$
$$O_8 = -1$$
$$O_9 = f(w_{96}o_6 + w_{97}o_7 + w_{98}o_8) = f(w_{96}o_6 + w_{97}o_4 - w_{98})$$

Step 4: Output error measure

$$E = \tfrac{1}{2}(t - o_9)^2 + E$$
$$\delta_9 = f'(tot_9)(t - o_9)$$
$$= o_9(1 - o_9)(t - o_9)$$

Step 5: Error back-propagation

Fourth layer weight updates:

$$\Delta w_{96} = \eta \delta_9 o_6 \quad w_{96}^{new} = w_{96}^{old} + \Delta w_{96}$$
$$\Delta w_{97} = \eta \delta_9 o_7 \quad w_{97}^{new} = w_{97}^{old} + \Delta w_{97}$$
$$\Delta w_{98} = \eta \delta_9 o_8 \quad w_{98}^{new} = w_{98}^{old} + \Delta w_{98}$$

Third layer error signals:

$$\delta_6 = f'(tot_6)\sum_{i=9}^{9} w_{i6}\delta_i = o_6(1-o_6)w_{96}\delta_9$$
$$\delta_7 = f'(tot_7)\sum_{i=9}^{9} w_{i7}\delta_i = o_7(1-o_7)w_{97}\delta_9$$

Third layer weight updates:

$$\Delta w_{63} = \eta \delta_6 o_3 \quad w_{63}^{new} = w_{63}^{old} + \Delta w_{63}$$
$$\Delta w_{64} = \eta \delta_6 o_4 \quad w_{64}^{new} = w_{64}^{old} + \Delta w_{64}$$
$$\Delta w_{65} = \eta \delta_6 o_5 \quad w_{65}^{new} = w_{65}^{old} + \Delta w_{65}$$

$$\Delta w_{73} = \eta \delta_7 o_3 \quad w_{73}^{new} = w_{73}^{old} + \Delta w_{73}$$
$$\Delta w_{74} = \eta \delta_7 o_4 \quad w_{74}^{new} = w_{74}^{old} + \Delta w_{74}$$
$$\Delta w_{75} = \eta \delta_7 o_5 \quad w_{75}^{new} = w_{75}^{old} + \Delta w_{75}$$

Second layer error signals:

$$\delta_3 = f'(tot_3)\sum_{i=6}^{7} w_{i3}\delta_i = 0.5o_3(1-o_3)\left(w_{63}\delta_6 + w_{73}\delta_7\right)$$
$$\delta_4 = f'(tot_4)\sum_{i=6}^{7} w_{i4}\delta_i = 0.5o_4(1-o_4)\left(w_{64}\delta_6 + w_{74}\delta_7\right)$$

Second layer weight updates:

$$\Delta w_{30} = \eta \delta_3 o_0 \quad w_{30}^{new} = w_{30}^{old} + \Delta w_{30}$$
$$\Delta w_{31} = \eta \delta_3 o_1 \quad w_{31}^{new} = w_{31}^{old} + \Delta w_{31}$$
$$\Delta w_{32} = \eta \delta_3 o_2 \quad w_{32}^{new} = w_{32}^{old} + \Delta w_{32}$$

$$\Delta w_{40} = \eta \delta_4 o_0 \quad w_{40}^{new} = w_{40}^{old} + \Delta w_{40}$$
$$\Delta w_{41} = \eta \delta_4 o_1 \quad w_{41}^{new} = w_{41}^{old} + \Delta w_{41}$$
$$\Delta w_{42} = \eta \delta_4 o_2 \quad w_{42}^{new} = w_{42}^{old} + \Delta w_{42}$$

Step 6: Repeat the process starting from step 2 using another exemplar. Once all exemplars have been used, we finish one epoch.

Step 7: Check if the cumulative error in the output layer is acceptable. If so the network has been trained. If not, repeat the whole process until the desired cumulative error is achieved.