Tian Xiaoyang
26001904581

Q1

This is a binary classification problem:

- A 2-layer neural network as shown in Figure 1, is used for the binary classification.
- The data in the table is used for the training of the neural network. y has two values (0 or 1), and x (feature) has three dimensions.
- The initial value of the parameters of the neural network are as follows:

$$\mathbf{W}^{[1]} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & -2 \end{bmatrix}, \mathbf{b}^{[1]} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{W}^{[2]} = [1, \quad 1, \quad 1, \quad 1], \mathbf{b}^{[2]} = 1,$$

- The neurons in the hidden layer use the following **ReLU** function as the activation function.

$$g(x) = \begin{cases} 0, for\ x < 0 \\ x, for\ x \geq 0 \end{cases}$$

- The neurons in the output layer use the sigmoid function as the activation function.
- Learning rate is 1 in the training.

After the first iteration of the backpropagation in the training, what is the value of $w_{1,1}^{[1]}$ ($w_{1,1}^{[1]}$ is the first element in the first row of $\mathbf{W}^{[1]}$)?

Calculation process and formulars must be included in your answer!

x			y
x_1	x_2	x_3	
1	-1	1	1
-1	1	0	0

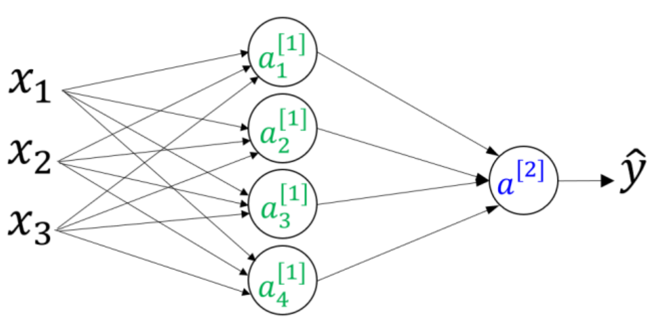

Figure 1

$$z_1 = \ \ w_1^T x + b_1$$
$$a_1 = \ ReLu(z_1) = \max(0, z_1)$$
$$z_2 = \ w_2^T a_2 + b_2$$
$$\hat{y} = a_2 = (z_2) = \frac{1}{1+e^{z2}}$$
$$z^1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & -2 \end{bmatrix} \times \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 2 \\ 1 & 0 \\ -1 & 1 \end{bmatrix}$$

$$a^1 = \max(0, \begin{bmatrix} 1 & -1 \\ 1 & 2 \\ 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix})$$

$$z^1 = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} + 1 = \begin{bmatrix} 4 & 4 \end{bmatrix}$$

$$a^2 = \sigma(z^2) - \frac{1}{1+e^{-[4 \quad 4]}} = [0.982 \quad 0.982]$$

Backpropagation:

$$\frac{dL}{dw_{1,1}^{[T]}} = \frac{dL}{da^{[2]}} \cdot \frac{da^{[2]}}{dz^{[2]}} \cdot \frac{dz^{[2]}}{da_1^{[z]}} \cdot \frac{da_1^{[z]}}{dz_1^{[z]}} \cdot \frac{dz^{[2]}}{dw_1^{[z]}}$$
$$= \frac{1}{m} \cdot (\frac{y}{\hat{y}} \cdot \frac{y-1}{1-\hat{y}}) \sigma(z^2)(1-\sigma(z^2)) w^2 \cdot 0 \wedge 1 \cdot x$$
$$= \frac{1}{2} \cdot ([-0.17 \ \ 0.98] \cdot 1[1 \ \ 0]) \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$
$$= \frac{1}{2} \cdot [-0.017\ , 0] \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$
$$= -0.0089$$

$$w_{1,1}^{[1]} = w_{1,1}^{[1]} - \alpha \frac{1}{m} \sum_{i=1}^{m} \frac{dj}{dw[1]}$$
$$= 1 - 1(-.0089)$$
$$= 1.0089$$

Q2.

For following sample data, what is the fitting straight line?

- Please use **Gradient Descent** for linear regression in this question.
- The initialization is: w = 0, b = 0, Learning rate is: α= 0.04.
- The regression should stop after two iterations.

Calculation process and formulars must be included in your answer!

X	Y
1	5
2	12
3	18

$$\frac{\partial L(w,b)}{\partial w} = \frac{\partial \sum_{i=1}^{m}(y[i]-(wx[i]+b))^2}{\partial w} = \frac{-2}{m} \sum_{i=1}^{m} x[i](y[i] - (wx[i[+b))$$

$$\frac{\partial L(w,b)}{\partial b} = \frac{\partial \sum_{i=1}^{m}(y[i]-(wx[i]+b))^2}{\partial b} = \frac{-2}{m} \sum_{i=1}^{m}(y[i] - (wx[i[+b))$$

$$w = w - \alpha \frac{\partial L(w,b)}{\partial b} \text{ , } b = b - \alpha \frac{\partial L(w,b)}{\partial b}$$

$$w[1] = \ w - \alpha \frac{-2}{3} \sum_{i=1}^{3} x[i](y[i] - (wx[i] + b))$$
$$= \ 0 - 0.04 \times -\frac{166}{3}$$
$$= \ \frac{166}{75}$$
$$= \ 2.2133$$

$$b[1] = \; b - \alpha \frac{-2}{m} \sum_{i=1}^{m} (y[i] - (wx[i] \; + b))$$
$$= \; 0 - 0.04 \times -\frac{70}{3}$$
$$= \; \frac{14}{15}$$
$$= \; 0.9333$$

$$w[2] = \; w[1] - \alpha \frac{-2}{3} \sum_{i=1}^{3} x[i] (y[i] - (w[1] x[i] + b[1]))$$
$$= \; \frac{166}{75} - 0.04 \times -\frac{6962}{225}$$
$$= \; 3.4510$$

$$b[2] = \; b - \alpha \frac{-2}{m} \sum_{i=1}^{m} (y[i] - (w[1] x[i] \; + b[1]))$$
$$= \; \frac{14}{15} - 0.04 \times -\frac{473}{25}$$
$$= \; \frac{2696}{1875}$$
$$= \; 1.4379$$

$$\widehat{y[i]} = inter + slope * x[i]$$

$$y[1] = 5$$
$$b[1] + w[1] * x[1] = 3.1466$$
$$b[2] + w[2] * x[1] = 4.8889$$

$$y[2] = 12$$
$$b[1] + w[1] * x[2] = 5.3599$$
$$b[2] + w[2] * x[2] = 8.3399$$

$$y[3] = 18$$
$$b[1] + w[1] * x[3] = 7.5732$$

$b[2] + w[2] * x[3] = 11.7909$

$y[i] = 1.4379 + 3.4510 * x[i]$ is the best fitting line