Assignment-2

(Q-1) Linear regression:

Loss function, $L = \frac{1}{2}(y-\hat{y})^2$

Our system:

Input (2 features) $\rightarrow$ 1 hidden layer (sigmoid) $\rightarrow$ 1 output.

$y \rightarrow$ Target

$\hat{y} \rightarrow$ Prediction

Backward propagation:

Output Layer:

Output: $a_2 = \hat{y}$

Chain rule:

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial a_2}\frac{\partial a_2}{\partial z_2}\frac{\partial z_2}{\partial w_2}$$

$$\frac{\partial L}{\partial a_2} = (y-a_2)(-1)$$

$$\frac{\partial a_2}{\partial z_2} = 1$$

$$\frac{\partial z_2}{\partial w_2} = a_1$$

$$\frac{\partial L}{\partial w_2} = \boxed{(a_2-y)}\, a_1^T$$

$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial a_2}\frac{\partial a_2}{\partial z_2}\frac{\partial z_2}{\partial b_2}$$

$$\frac{\partial L}{\partial a_2} = (y-a_2)(-1)$$

$$\frac{\partial a_2}{\partial z_2} = 1$$

$$\frac{\partial z_2}{\partial b_2} = 1.$$

$$\frac{\partial L}{\partial b_2} = \boxed{(a_2-y)}$$

hidden Layer:

Similarly,

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial a_2}\frac{\partial a_2}{\partial z_2}\frac{\partial z_2}{\partial a_1}\frac{\partial a_1}{\partial z_1}\frac{\partial z_1}{\partial w_1}$$

$$= \boxed{(a_2-y)} \cdot w_1 \cdot a_1 \cdot (1-a_1) X$$

$$\frac{\partial L}{\partial b_1} = \boxed{(a_2-y)} \cdot w_1 \cdot a_1 (1-a_1) \cdot 1$$

Forward passes:

$a_2 = g(z_2) = \hat{y}$

$z_2 = w_2 a_1 + b_2$

$a_1 = g(z_1) = \frac{1}{1+e^{-z_1}}$

$z_1 = w_1 X + b_1$

$g'(z_1) = g(z_1)(1-g(z_1))$

output sigmoid

$a_2 = \hat{y} = \sigma(z_2)$

Binary classifaction:

Loss function, $L = -[y \times \log(\hat{y}) + (1-y) \times \log(1-\hat{y})]$

$$\frac{\partial L}{\partial \omega_2} = \frac{\partial L}{\partial a_2} \frac{\partial a_2}{\partial z_2} \frac{\partial z_2}{\partial \omega_2}$$

F.P.

$a_2 = g(z_2)$

$= \frac{1}{1+e^{-z_2}}$

$$\frac{\partial L}{\partial a_2} = -\frac{y}{a_2} - \frac{(1-y)(-1)}{1-a_2} = \frac{a_2 - y}{a_2(1-a_2)}$$

$z_2 = \omega_2 a_1 + b_2$

$a_1 = g(z_1)$

$= \frac{1}{1+e^{-z_2}}$

$$\frac{\partial a_2}{\partial z_2} = a_2(1-a_2)$$

$$\frac{\partial L}{\partial \omega_2} = \frac{a_2 - y}{\cancel{a_2(1-a_2)}} \cdot \cancel{a_2} \cancel{a_2 \cdot (1-a_2)} \cdot a_1$$

$z_1 = \omega_1 x + b_1$

$g'(z_2) = g(z)(1-g(z))$

$$= (a_2 - y) \cdot a_1^T$$

Similarly, $\frac{\partial L}{\partial b_2} = (a_2 - y)$

$$\frac{\partial L}{\partial \omega_1} = (a_2 - y)\omega_1 a_1 (1-a_1) x$$

$$\frac{\partial L}{\partial b_1} = (a_2 - y) \cdot \omega_1 \cdot a_1(1-a_1) \cdot 1$$

The main difference for the update rule for regression & binary classification is the loss function (Mean Square Error & log loss) and the activation function in output layer (Linear & sigmoid). Each pair provides the same update for both cases.

So we can update rule. for this given system we have same