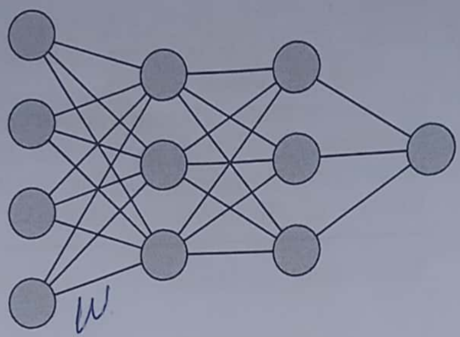
Each question is 2.5 marks each.

[10 points]

Question 1:

Consider the multi-layer perceptron with 2 hidden layers below.



The MLP with 2 - hidden layers is defined as below:

$$\mathbf{z}_1 = \mathbf{W}_1\mathbf{x}^{(i)} + \mathbf{b}_1$$

$$\mathbf{a}_1 = ReLU(\mathbf{z}_1)$$

$$\mathbf{z}_2 = \mathbf{W}_2\mathbf{a}_1 + \mathbf{b}_2$$

$$\mathbf{a}_2 = ReLU(\mathbf{z}_2)$$

$$\mathbf{z}_3 = \mathbf{W}_3\mathbf{a}_2 + \mathbf{b}_3$$

$$\hat{y}^{(i)} = \sigma(\mathbf{z}_3)$$

$$L^{(i)} = y^{(i)}\log(\hat{y}^{(i)}) + (1 - y^{(i)})\log(1 - \hat{y}^{(i)})$$

$$J = -\frac{1}{m}\sum_{i=1}^{m} L^{(i)}$$

Note that $x^{(i)}$ represents a pair of single input examples and each is of shape $n \times 1$. Further $y^{(i)}$ is a single output label and is a scalar. There are m examples in our dataset. We use h nodes in our hidden layers; that is z_1's and z_2's shape is $h \times 1$.

(a) What are the shapes of W_1, b_1, W_2, b_2 ? If we were vectorizing across multiple examples, what would be the shapes of X and Y be instead?

marking scheme

$b_1, b_2 \rightarrow 0.5$
$X \rightarrow 0.5$
$Y \rightarrow 0.5$
$W_1 \rightarrow 0.5$
$W_2 \rightarrow 0.5$

$x^{(i)} = n \times 1$. $y^{(i)} = 1*1$

$X = m*n$.

$Y = m*1$.

$W_1 = \cancel{m}\, n*h$.

$b_1 = \cancel{1*h}\ h*1$.

$W_2 = h*h$.

$b_2 = h*1$.

$\underset{m*h}{Z_1} = X W_1 + b_1$
$m*n(n*h) + (h*1)$.

$\underset{m*h}{Z_2} = \underset{m*h}{Z_1}\, \underset{h*h}{W_2} + \underset{h*1}{b_2}$.

$\underset{m*1}{Z_3} = \underset{m*h}{Z_2}\, \underset{h*1}{W_3} + \underset{1*1}{b_3}$.

(b) What is $\frac{\partial J}{\partial z_3}$? Hint: You can simplify the equation in terms of $\hat{y}^{(i)}$.

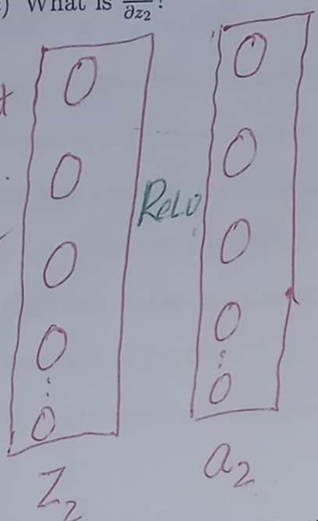
$$\frac{\partial J}{\partial L^{(i)}} \cdot \frac{\partial L^{(i)}}{\partial y^{(i)}} \cdot \frac{\partial y'}{\partial z_3}$$

$$-\frac{1}{m}\left(y^{(i)}/\hat{y}^{i} - (1-y^{(i)})/(1-\hat{y}^{i})\right) * a(z_3)\cdot(1-a(z_3)).$$

$$\frac{1}{m}\left(\hat{y}^{(i)} - y^{(i)}\right).$$

(c) What is $\frac{\partial a_2}{\partial z_2}$?

marking scheme:
2.5 for correct answer.
Format may vary.

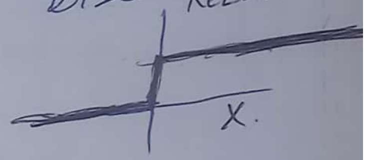


Since we are applying ReLU.

$$\mathrm{ReLU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases}$$

$$\mathrm{ReLU}'(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases}$$

ReLU'(x)



$$\frac{\partial a_2^i}{\partial z_2^j} = \begin{cases} 0 & i \neq j \\ 1 & i = j \;\&\; z_2^j \geq 0 \\ 0 & i = j \;\&\; z_2^j < 0 \end{cases}$$

(d) Leaky ReLU is an activation function that is a variant of ReLU. The leaky ReLU is defined as $f(x) = max(0.1x, x)$. Compute $f'(x)$ and sketch the graphs of $f(x)$ and $f'(x)$

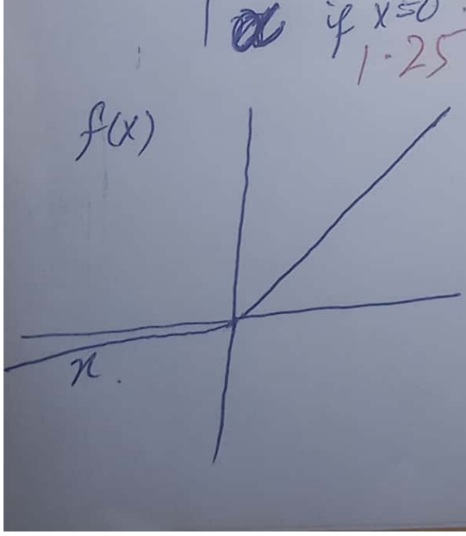
$$f(x) = \begin{cases} x & x > 0 \\ 0.1x & x < 0 \\ \alpha & \text{if } x = 0 \end{cases}$$

1.25 marks for Leaky ReLU plot

1.25 marks for Leaky Relu plot derivative.

$$f'(x) = \begin{cases} 1 & x > 0 \\ 0.1 & x < 0 \end{cases}$$

f(x)



f'(x).