DATE: / /

1. Forward propagation

In this process we move in forward path from input to the output. In this process we have different hidden layers and one output layer. Each hidden layer consist of many hidden units which are independent to each other and are the outcome of the activation function applied to the some linear function of the previous hidden layer. By calculating these equations we find our predicted output from given input.

Backward propagation

This process is exactly opposite to the above process. Using this we find gradients (or more precisely some refinement term) from the loss function which is the net error in computing predicted output. So, we tune our all parameters to ensure less cost (less error). For this we use different optimizers. Once we tuned our parameters, then finally we use it to predict output for the test input data.

2. Input X {4x1}
Output Y {1x1}
W1 {5x4} b1 {5x1}
W2 {1x5} b2 {1x1}
} hidden unit parameters

g_1 activation function for layer 1
g_2 " " layer 2

PAGE NO.
DATE: / /

Forward prop. (Vectorised)

$z1 = w1 \cdot x + b1$

$A1 = g_1(z1)$

$z2 = w2 \cdot A1 + b2$

$A2 = g_2(z2)$

$cost = -Y \log A2 - (1-Y) \log(1-A2)$

Backward prop. (Vectorised)

$dz2 = A2 - Y$ { If g_2 is sigmoid }

$dw2 = dz2 \cdot A1$

$db2 = dz2$

$dA1 = w2^T \cdot dz2$

$dz1 = dA1 * g_1'(z1)$ (element-wise product)

$dw1 = dz1 \cdot X$

$db1 = dz1$

Generalization for MLP : Forward prop.

consider X as $A^{[0]}$.

For every layer l in [1, L]

$$z^{[l]} = w^{[l]} \cdot A^{[l-1]} + b^{[l]}$$
$$A^{[l]} = g^{[l]}(z^{[l]})$$

{ Store $A^{[l]}$ and $z^{[l]}$ for Back prop. }

Generalization for MLP : Backward prop.

$dz^{[L]} = A^{[L]} - Y$

$dw^{[L]} = dz^{[L]} \cdot A^{[L-1]}$

$db^{[L]} = dz^{[L]}$

For (L-1) to 1 :- ($l \in [1, \dots L-1]$)

$dz^{[l]} = dA^{[l]} * g^{[l]'}(z^{[l]})$

$dw^{[l]} = dz^{[l]} \cdot A^{[l-1]}$

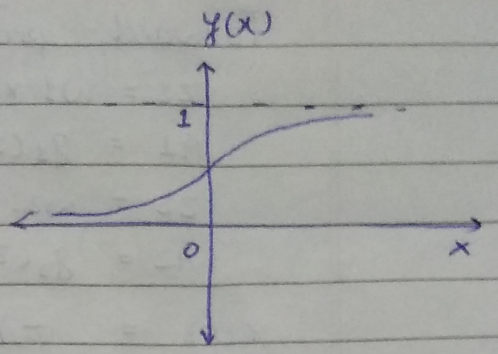
$db^{[l]} = dz^{[l]}$, $dA^{[l-1]} = w^{[l]T} \cdot dz^{[l]}$

PAGE NO.
DATE: / /

3. a) Sigmoid

$$y(x) = \frac{1}{1+e^{-x}}$$

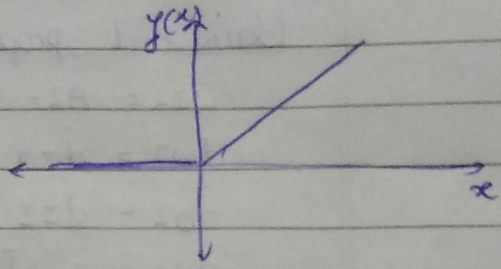
$$\frac{dy}{dx} = y(x)(1-y(x))$$



b) Relu

$$y(x) = \max(0, x)$$

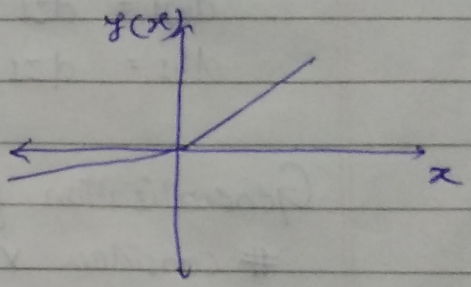
$$y'(x) = \begin{cases} 0 & x<0 \\ 1 & x \geq 0 \end{cases}$$



c) Leaky Relu

$$y(x) = \begin{cases} 0.01x & x<0 \\ x & x \geq 0 \end{cases}$$

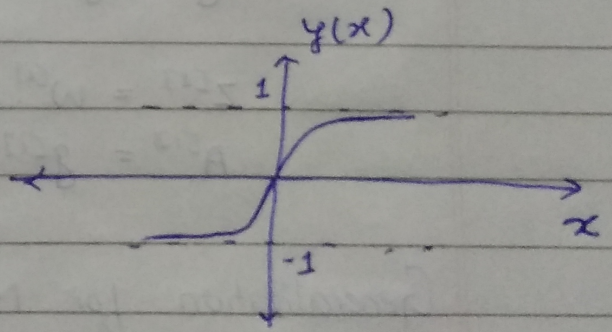
$$\frac{dy}{dx} = \begin{cases} 0.01 & x<0 \\ 1 & x \geq 0 \end{cases}$$



d) Tanh

$$y(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

$$\frac{dy}{dx} = 1 - y^2(x)$$



e) Softmax

Z ~ column vector / Vector

K ~ sized of Z vector

$$y(z[i]) = \frac{e^{z[i]}}{\sum_{i=1}^{K} e^{z[i]}}$$

$$\frac{dy}{dz[i]} = y(z[i])(1 - y(z[i]))$$