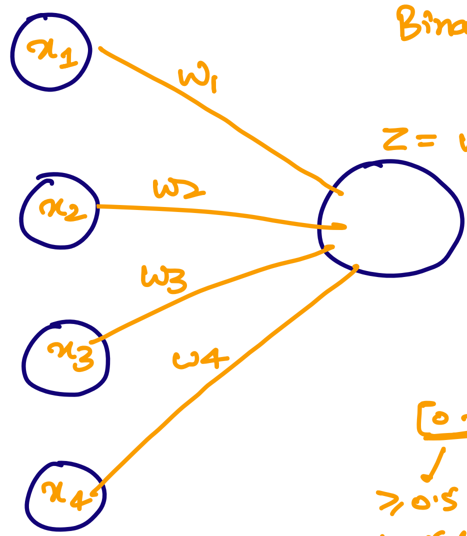
Neural Architecture

1 layer network

Logistic Regression



$\frac{1}{1+e^{x}}$

Binary Classification

$$z = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b$$

$\sigma(z)$ $\hat{y}$

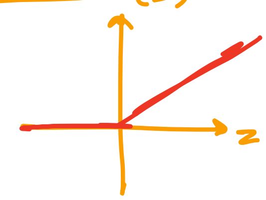
Activation fxn (sigmoid)

[0 - 1]

≥0.5 class 1

<0.5 class 0

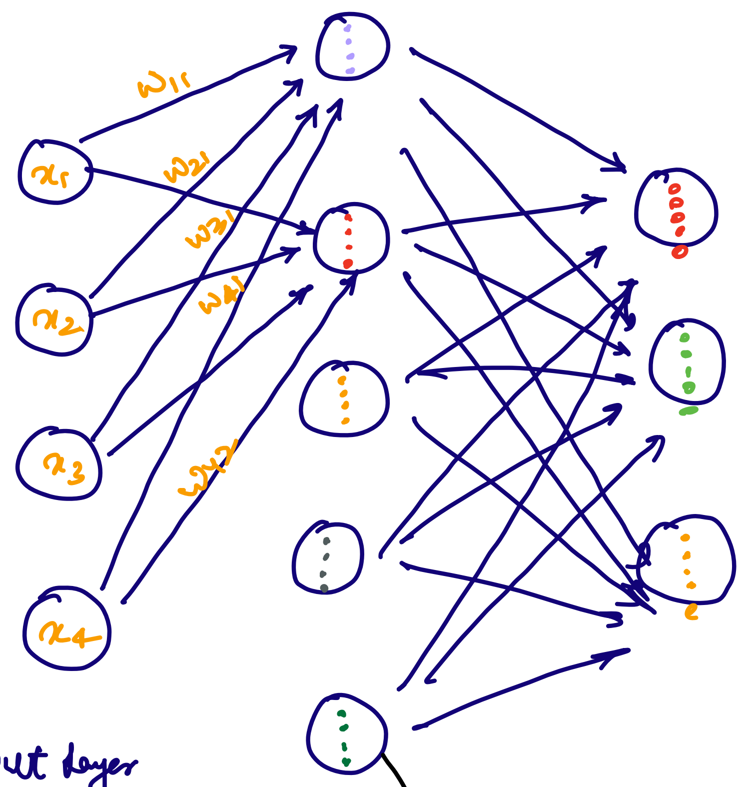
Relu:



$\sigma(z) = z \quad z > 0$

$\sigma(z) = 0 \quad z \leq 0$

2 layer network



$$W^{[1]} = \begin{bmatrix} w_{11} & w_{12} & \cdot & \cdot & \cdot \\ w_{21} & w_{22} & \cdot & \cdot & \cdot \end{bmatrix}$$

w_{21} w_{32}

w_{41} w_{42}

4×5

no of neurons in i/p layer

no of neurons in hidden layer

$$W^{[1]} = \begin{bmatrix} | & | & | & | & | \\ w_1^{[1]} & w_2^{[1]} & w_3^{[1]} & w_4^{[1]} & w_5^{[1]} \\ | & | & | & | & | \end{bmatrix}$$

weights associated with 5^{th} neuron of 1^{st} layer

$$W^{[2]} = \begin{bmatrix} \circ & \circ & \circ \\ \circ & \circ & \circ \\ \circ & \circ & \circ \\ \circ & \circ & \circ \\ \circ & \circ & \circ \end{bmatrix}_{5 \times 3}$$

Output Layer:

Dog $\bigcirc \rightarrow 0.6$

Cat $\bigcirc \rightarrow 0.1$

Horse $\bigcirc \rightarrow 0.3$

Dog class

10

20

30

Softmax

$$\frac{e^{10}}{e^{10}+e^{20}+e^{30}} = \bigcirc$$

$$\frac{e^{20}}{e^{10}+e \;\_\_\_\_} = \bigcirc$$

$$\frac{e^{10}}{e^{10}+e^{20}+e^{30}} = 0$$

No. of parameters:

$$\underbrace{4\times5}_{w^{[1]}} + \underbrace{5\times3}_{w^{[2]}} + \underbrace{5}_{b^{[1]}} + \underbrace{3}_{b^{[2]}}$$

$= 20 + 15 + 5 + 3$

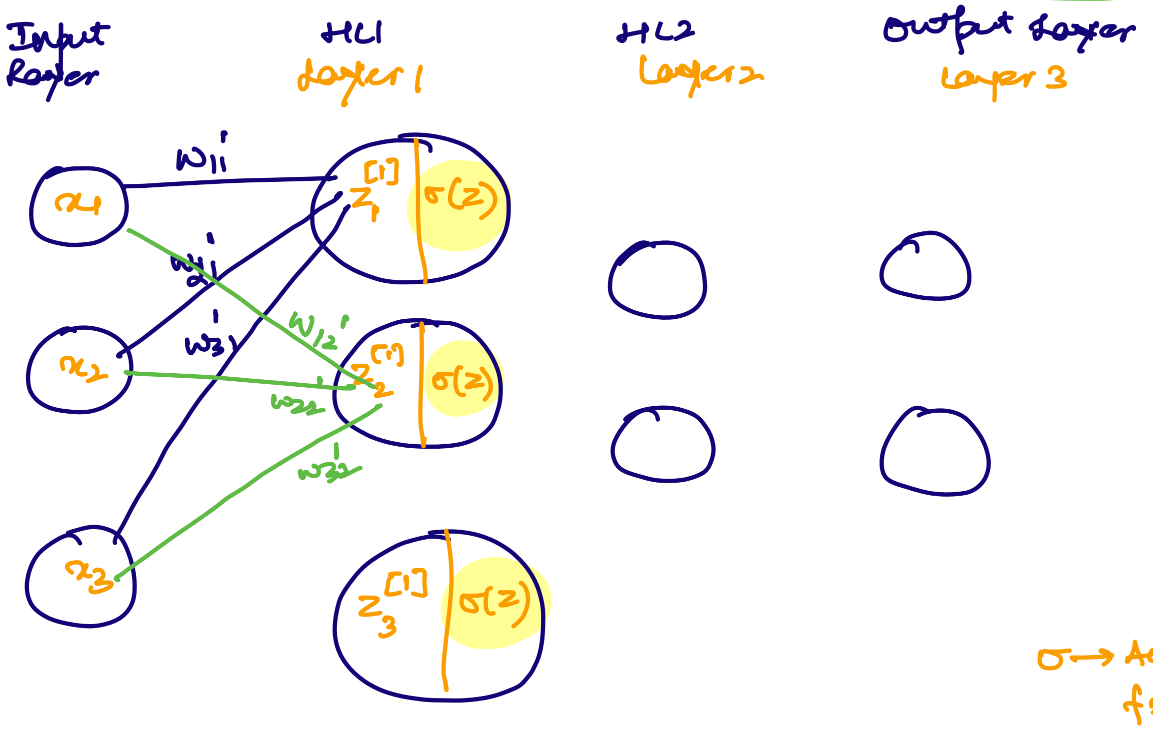
$= \underline{\underline{43}}$

LR:

$w_0x_0 + w_1x_1 + w_2x_2 \dots w_nx_n$

$x_0 = 1 \rightarrow (n+1)$

3 layer:



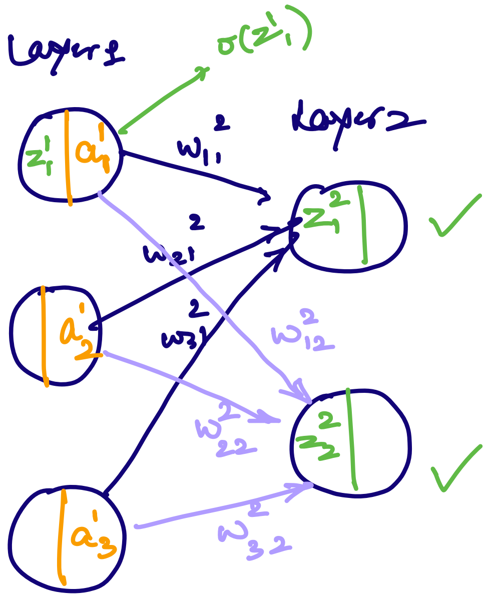
$\sigma \rightarrow$ Activation fn

$$z_1^{[1]} = x_1 * w_{11}^1 + x_2 * w_{21}^1 + x_3 * w_{31}^1 + b_1^{[1]} = \underline{\text{Real no}}$$

$$z_2^{[1]} = x_1 * w_{12}^1 + x_2 * w_{22}^1 + x_3 * w_{32}^1 + b_2^{[1]}$$

$$\boxed{z_j^{[l]} = \sum_{i=1}^{n} w_{ij}^{[l]} x_i + b_j^{[l]}}$$

n = no of features



$$z_1^2 = w_{11}^2 * a_1^1 + w_{21}^2 * a_2^1 + w_{31}^2 * a_3^1 + b_1^2$$

$$z_2^2 = w_{12}^2 * a_1^1 + w_{22}^2 * a_2^1 + w_{32}^2 * a_3^1 + b_2^2$$

$$\boxed{z_j^{[l]} = \sum_i w_{ij}^{[l]} * a_i^{[l-1]} + b_j^{[l]}}$$

$i \in$ all neurons from $l-1$ layer

$$\begin{bmatrix} - \\ - \end{bmatrix}$$

Vector:

$3 \times 2 \qquad 3 \times 1 \quad \Rightarrow \quad a_i = \begin{bmatrix} - \\ - \\ - \end{bmatrix} \; 3 \times 1$

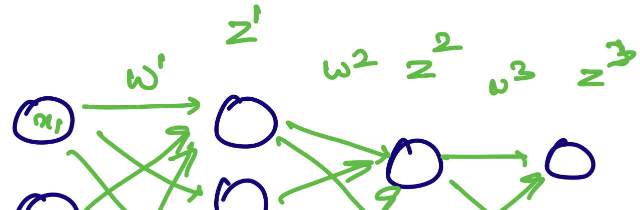
$w^T . a$

$2 \times 3 \quad 3 \times 1 = \underline{\underline{2 \times 1}}$

$(2 \times 1) + (2 \times 1)$

$$z^1 = (w^1)^T . x + b^1$$

$$a^1 = \sigma(z^1)$$

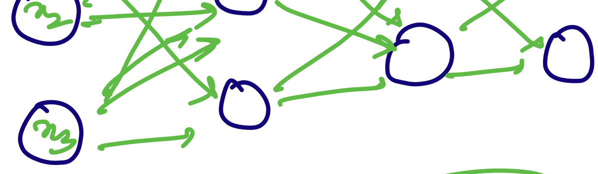


$$z^2 = (w^2)^T \cdot a^1 + b^2$$

$$a^2 = \sigma(z^2)$$

$$z^3 = (w^3)^T \cdot a^2 + b^3$$

$$\hat{y} = \text{softmax}(z^3)$$



σ:

- Relu
- Sigmoid
- Softmax

[1,2]

$$\frac{e^1}{e^1+e^2}, \quad \frac{e^2}{e^1+e^2}$$

3 neurons o/p layer: [1,2,3]

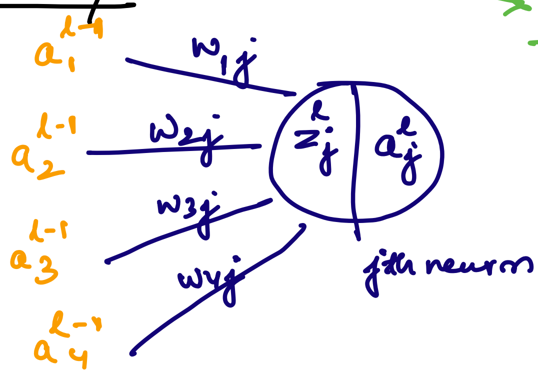
$$\frac{e^1}{e^1+e^2+e^3}, \quad \frac{e^2}{e^1+e^2+e^3}, \quad \frac{e^3}{e^1+e^2+e^3}$$

Backpropagation:

Loss function:

$$L = (y_{actual} - y_{pred})^2$$

Output layer:



L = last layer / output layer

Binary Classification task:

Binary Cross Entropy

$y = (x-5)^2$

$$z_j^l = \sum_i w_{ij}^l a_i^{l-1} + b_j^l$$

$$\frac{\partial L}{\partial w} = ? \qquad \frac{\partial L}{\partial b} = ?$$

$$w_{ij} \longrightarrow z_j \longrightarrow a_j \longrightarrow \text{Loss}.$$

$$\frac{\partial L}{\partial w_{ij}} = \frac{\partial L}{\partial a_j} \cdot \frac{\partial a_j}{\partial z_j} \cdot \frac{\partial z_j}{\partial w_{ij}}$$

$\frac{\partial L}{\partial a_j}$?

$$L = \frac{1}{2} \sum_i (y_i^l - a_i^l)^2$$

i ranges over the no. of neurons in o/p layer

$$\frac{\partial L}{\partial a_j} = \frac{1}{2} \cdot 2(y_j - a_j)(-1)$$

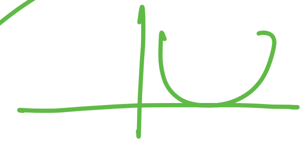
$$\frac{\partial L}{\partial a_j} = -(y_j - a_j)$$

$\frac{\partial a_j}{\partial z_j}$?

$$a_j = \sigma(z_j)$$

$$\frac{\partial a_j}{\partial z_j} = \sigma'(z_j)$$

$$\frac{\partial a_j}{\partial z_j} = \sigma(z_j)(1 - \sigma(z_j))$$



x? y is min?

$$\frac{\partial y}{\partial x}$$

w? L is min?

$$\frac{\partial L}{\partial w}$$

o/p layer

$a = 0.3$ Cat — TD: 1

$a = 0.7$ Dog — 0

$$\text{loss fn} = (1-0.3)^2 + (0-0.7)^2$$

$$=$$

$$\text{loss} = (y-a)^2$$

Sigmoid fn

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

$$\frac{\partial \sigma(z)}{\partial z} = \frac{\partial (1+e^{-z})^{-1}}{\partial z}$$

$$= \frac{-1(e^{-z})(-1)}{(1+e^{-z})^2}$$

$$= \frac{e^{-z}}{}$$

$$\frac{\partial z_j}{\partial w_{ij}} \; ?$$

$$= \left(\frac{1}{1+e^{-z}}\right)\left(1 - \frac{1}{1+e^{-z}}\right)$$

$$= \sigma(z)\left(1-\sigma(z)\right)$$

$$z_j^\ell = \sum_i w_{ij}^\ell a_i^{\ell-1} + b_j^\ell$$

$$\frac{\partial z_j}{\partial w_{ij}} = a_i^{\ell-1}$$

Combine:

$$\frac{\partial L}{\partial w_{ij}} = \frac{\partial L}{\partial a_j} \cdot \frac{\partial a_j}{\partial z_j} \cdot \frac{\partial z_j}{\partial w_{ij}}$$

$$\frac{\partial L}{\partial w_{ij}} = \underbrace{\boxed{-(y_j - a_j)\,\sigma(z_j)(1-\sigma(z_j))}}_{\delta_j^\ell} \; a_i^{\ell-1}$$

$$\frac{\partial L}{\partial w_{ij}} = \delta_j^\ell \, a_i^{\ell-1}$$

$$\delta_j^\ell = -(y_j - a_j)\,\sigma(z_j)(1-\sigma(z_j))$$

Bias:

$$\frac{\partial L}{\partial b_j} = \underbrace{\frac{\partial L}{\partial a_j} \cdot \frac{\partial a_j}{\partial z_j}}_{} \cdot \frac{\partial z_j}{\partial b_j}$$

$$= \delta_j^\ell$$

$$z_j^\ell = \sum_i w_{ij} a_i^{\ell-1} + b_j^\ell$$

$$\frac{\partial z_j}{\partial b_j} = 1$$

$$\frac{\partial L}{\partial b_j} = \delta_j^\ell \cdot 1$$

Output layer:

$$\frac{\partial L}{\partial w_{ij}} = \delta_j^{l} a_i^{l-1}$$

$$\delta_j^{l} = -(y_j - a_j)\,\sigma(z_j)(1-\sigma(z_j))$$

$$\frac{\partial L}{\partial b_j} = \delta_j^{l} \cdot 1$$