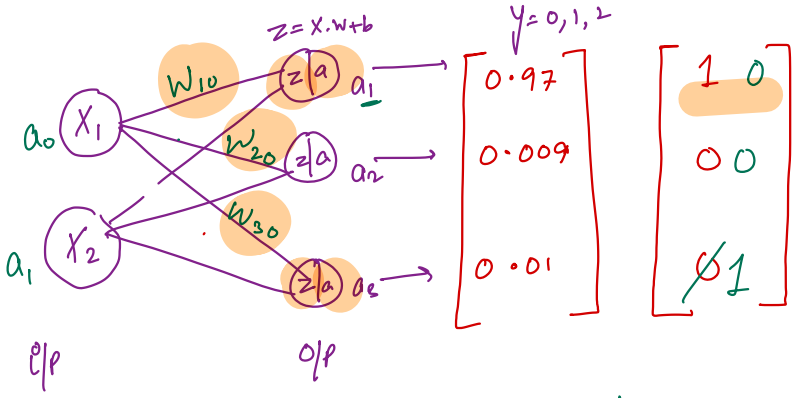
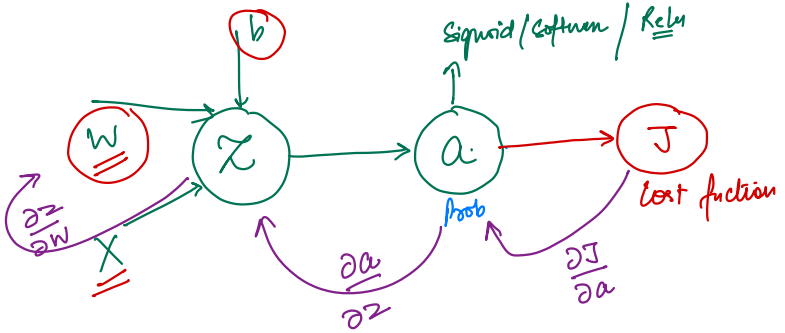




$z = X \cdot W + b$

$\frac{\partial J}{\partial W}$, $\frac{\partial J}{\partial b}$

$$\frac{\partial J}{\partial w} = \frac{\partial J}{\partial a} \times \frac{\partial a}{\partial z} \times \frac{\partial z}{\partial w}$$

Chain Rule (Backbone of NN)

Forward prop $z = \vec{w} \cdot \vec{x} + b$

$Act(z) =$ Sigmoid (z) or Softmax (z)

Multiclass $J \rightarrow$ Categorical cross entropy loss

Backprop

Log loss

$$J = -\frac{1}{I} \sum y_i \log(a_i) + (1-y_i) \log(1-a_i)$$

$$\frac{\partial J}{\partial a} = \frac{\partial}{\partial a} - \left(y \log(a) + (1-y) \log(1-a) \right)$$

$$= -\left(y \frac{\partial (\log a)}{\partial a} + (1-y) \frac{\partial \log(1-a)}{\partial a} \right)$$

$\log a$: $\frac{1}{a}(-1) = -\left(y \left(\frac{1}{a}\right) + (1-y) \left(\frac{-1}{1-a}\right) \right)$

$$= \frac{-y}{a} + \frac{(1-y)}{(1-a)} = \frac{(1-a)(-y) + (1-y)a}{a(1-a)}$$

$$= \frac{-y + ya + a - ya}{a(1-a)}$$

$\frac{1}{x} \Rightarrow x^{-1} \rightarrow \boxed{\frac{-1}{x^2}}$

$$\boxed{\frac{\partial J}{\partial a} \rightarrow \frac{a-y}{a(1-a)}} \quad (1)$$

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

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

$$\Rightarrow \frac{1 \times e^{-z}(+1)(-1)}{(1+e^{-z})(1+e^{-z})}$$

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

$$\Rightarrow \left(\frac{1}{1+e^{-z}}\right) - \left(\frac{1}{1+e^{-z}}\right)^2$$

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

$$\Rightarrow \boxed{a(1-a)}$$

$$\boxed{\frac{\partial a}{\partial z} = a(1-a)} \quad (2)$$

$$\frac{\partial z}{\partial w} = \frac{\partial (xw+b)}{\partial w} = x \quad \text{— (3)}$$

$$\frac{\partial J}{\partial w} = \frac{\partial J}{\partial a} \times \frac{\partial a}{\partial z} \times \frac{\partial z}{\partial w} = \frac{(a-y)}{a(1-a)} \times a(1-a) \times x$$

(1) (2) (3)

$$\frac{\partial J}{\partial w} = (a-y)\, x$$

$\frac{\partial J}{\partial z}$

$$\frac{\partial J}{\partial w} = (\text{error}) \times x$$

$(a-y)$

error = pred - actual

$$\boxed{\frac{\partial J}{\partial w} = \text{error} \times \frac{\partial z}{\partial w}}$$

$z = xw + b$

$\frac{\partial z}{\partial w} = x$

$$\frac{\partial J}{\partial b} = \text{error} \times \frac{\partial z}{\partial b}$$

$$\frac{\partial z}{\partial b} = (xw+b) \Rightarrow 1$$

$$\boxed{\frac{\partial J}{\partial b} = \text{error} \cdot 1}$$

Neural N/w from Scratch

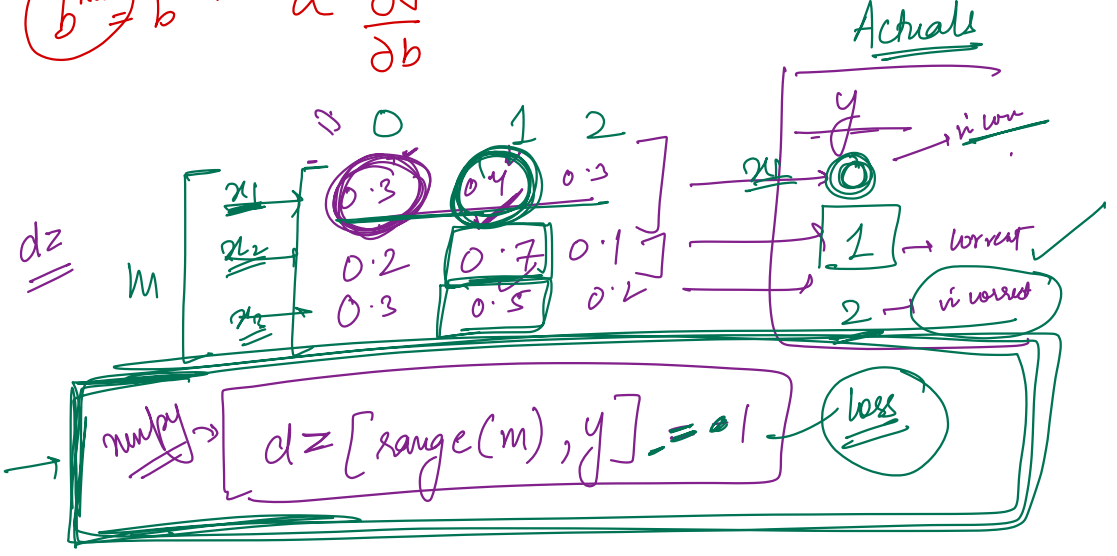
$$w^{new} = w^{old} - \alpha \sum^{N} \frac{\partial J}{\partial w}$$

α is learning rate

$$b^{new} = b^{old} - \alpha \frac{\partial J}{\partial b}$$

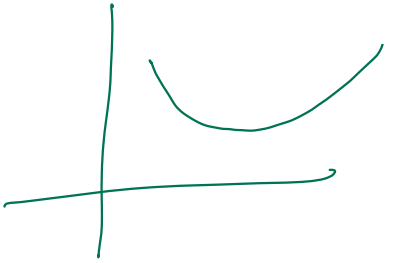
Stocastic GD

Batch GD



$z = 2x + 1$ → linear

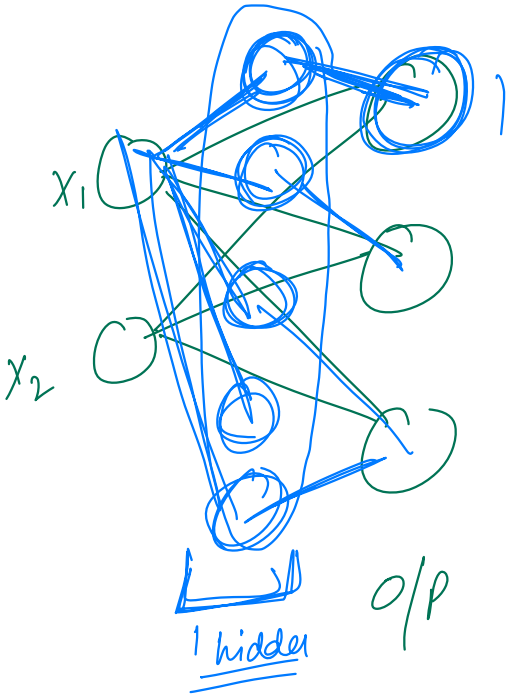
$\frac{1}{1+e^{(2x+1)}}$ ⇒ linear function



x^1 → x^2, x^3, x^4

Sq

$(2x+1)^2 \Rightarrow 4x^2 + 4x + 2$


1
x_1
x_2
o/p
1 hidden