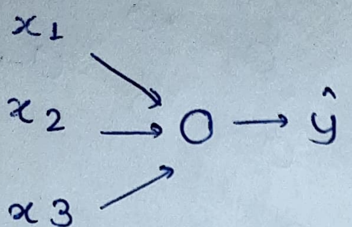
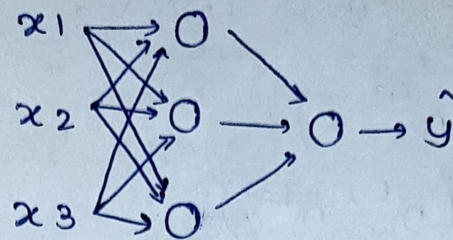
WEEK - 4

Deep Neural Network (L layered)



Logistic Regression → 1 Layer NN [SHALLOW]



1 hidden layer (NN) → 2 Layer [DEEP]

* 4 layer NN = 3 hidden layers

No. of nodes: Layer 1 = 5
Layer 2 = 5
Layer 3 = 3
Layer 4 = 1

Layers = L (4)
No. of nodes in layer l = $n^{[l]}$

$\therefore n^{[1]} = 5$; $n^{[2]} = 5$; $n^{[3]} = 3$; $n^{[4]} = 1$ & $n^{[0]} = 3$

$n^{[L]} = n^{[4]} = 1$

$a^{[l]}$ = activations of layer l.	$\Rightarrow a^{[l]} = g^{[l]}(z^{[l]})$
$w^{[l]}$ = weights for $z^{[l]}$	
$b^{[l]}$ = biases.	

- FORWARD PROPAGATION

$z^{[1]} = w^{[1]}x + b^{[1]}$

$a^{[1]} = g^{[1]}(z^{[1]})$

$z^{[2]} = w^{[2]}a^{[1]} + b^{[2]}$

$a^{[2]} = g^{[2]}(z^{[2]})$

$\vdots$

$z^{[4]} = w^{[4]}a^{[3]} + b^{[4]}$

$a^{[4]} = \hat{y} = g^{[4]}(z^{[4]})$

$\boxed{\text{vectorized}}$

$Z^{[1]} = W^{[1]}X + b^{[1]}$

$A^{[1]} = g^{[1]}(Z^{[1]})$

$Z^{[2]} = W^{[2]}A^{[1]} + b^{[2]}$

$A^{[2]} = g^{[2]}(Z^{[2]})$

$\vdots$

$\hat{y} = A^{[4]} = g^{[4]}(Z^{[4]})$

for $l = 1$ to L

Here $L = 4$

• MATRIX DIMENSIONS

$L = 5$

$n^{[1]} = 3$

$n^{[2]} = 5$

$n^{[3]} = 4$

$n^{[4]} = 2$

$n^{[5]} = 1$

$z^{[1]} = w^{[1]} x + b^{[1]}$

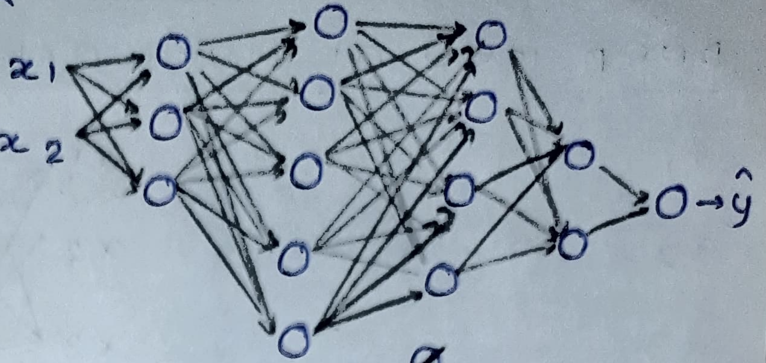
$(3,1) = (3,2)(2,1) + (3,1)$

$z^{[1]} = (n^{[1]}, 1)$

$w^{[1]} = (n^{[1]}, n^{[0]})$

$x = (n^{[0]}, 1)$

$b = (n^{[1]}, 1)$



$w^{[l]} = (n^{[l]}, n^{[l-1]}) = dw^{[l]}$

$b^{[l]} = (n^{[l]}, 1) = db^{[l]}$

* vectorized

$Z^{[1]} = W^{[1]} X + b^{[1]}$

$Z^{[1]}$: $(n^{[1]}, m)$

$W^{[1]}$: $(n^{[1]}, n^{[0]})$

X: $(n^{[0]}, m)$

$b^{[1]}$: $(n^{[1]}, 1) \xrightarrow{\text{broad cast}} (n^{[1]}, m)$

$Z^{[l]}, A^{[l]} : (n^{[l]}, m) = dZ^{[l]}, dA^{[l]}$
when $l = 0$ $A^{[0]} = X = (n^{[0]}, m)$

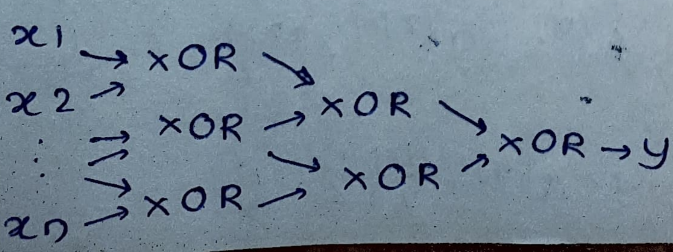
• Deep Network Importance

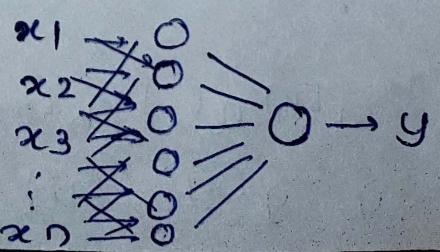
As we progress through layers of networks the features detected go from $\boxed{\text{simple}}$ to $\boxed{\text{complex.}}$

eg: $y = x_1$ XOR x_2 XOR .... XOR x_n (LOGICAL)

deep network $[\log_n]$ layers

one hidden layer $[2^{n-1}]$ nodes





• FORWARD & BACKWARD PROPAGATION

Layer l : $w^{[l]}, b^{[l]}$

→ **Forward** : Input $a^{[l-1]}$, output $a^{[l]}$

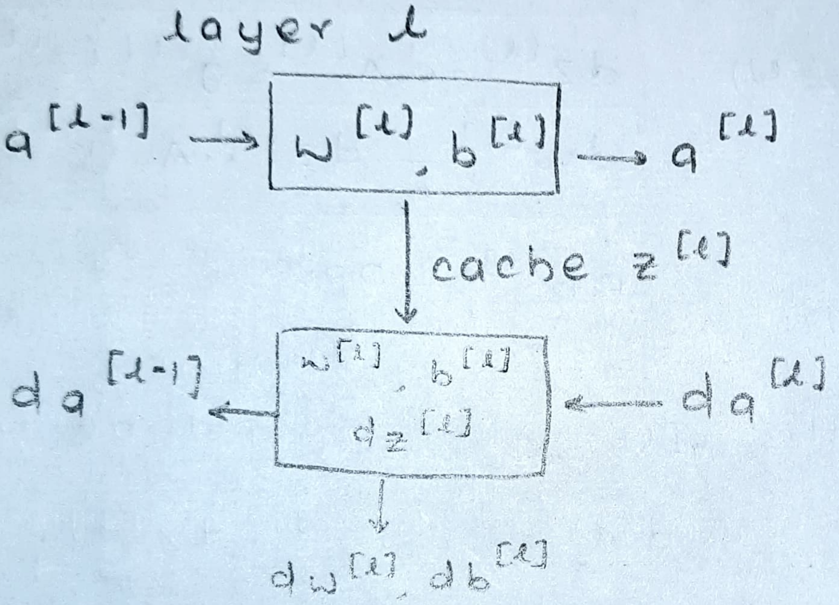
$z^{[l]} = w^{[l]} a^{[l-1]} + b^{[l]}$. cache $z^{[l]}$

$a^{[l]} = g^{[l]}(z^{[l]})$

→ **Backward** : Input $da^{[l]}$, output $da^{[l-1]}$

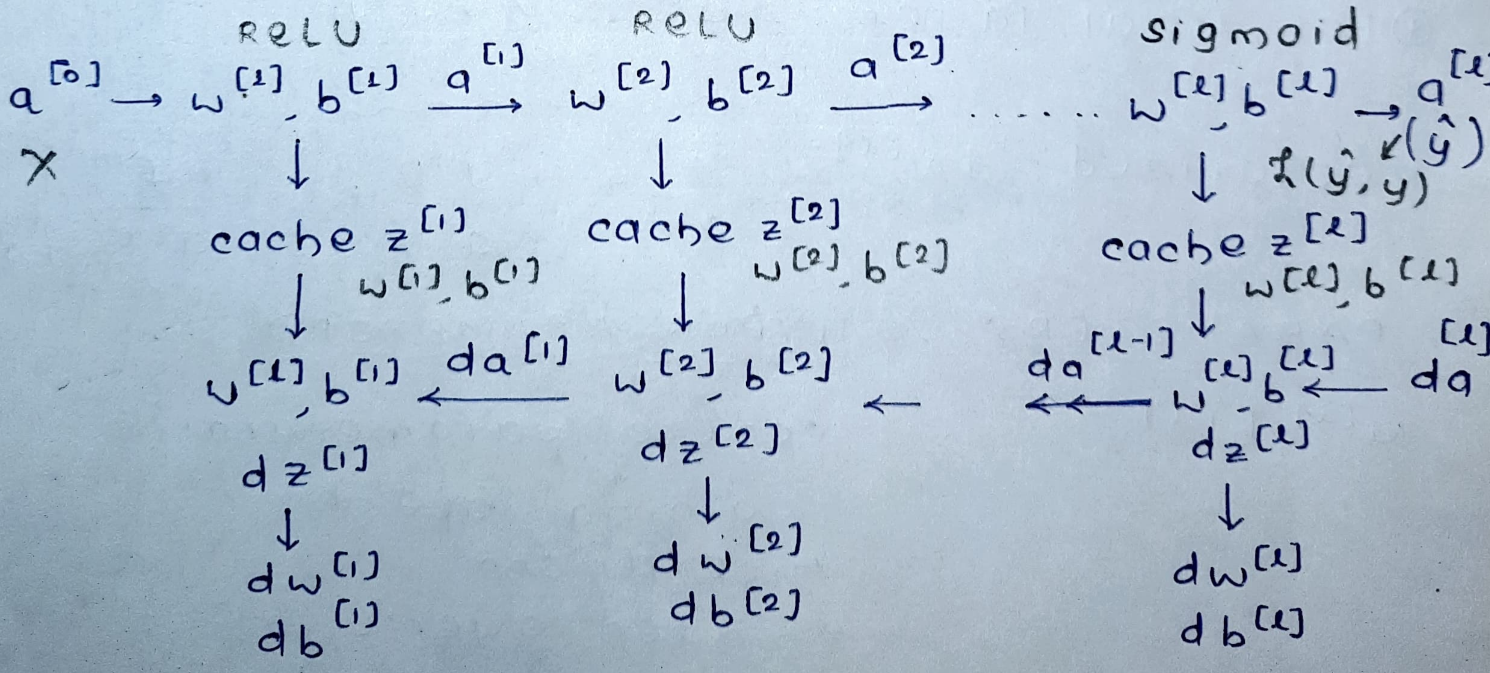
~~cache~~ cache $(z^{[l]})$ $dw^{[l]}$

$db^{[l]}$



cache

passes info from forward to back propagation step.



→ $w^{[l]} := w^{[l]} - \alpha dw^{[l]}$

→ $b^{[l]} := b^{[l]} - \alpha db^{[l]}$

- **FORWARD PROPAGATION FOR LAYER l**

Input: $a^{[l-1]}$

output: $a^{[l]}$, cache $(z^{[l]})$

VECTORIZED $Z^{[l]} = W^{[l]} \cdot A^{[l-1]} + b^{[l]}$

$A^{[l]} = g^{[l]}(Z^{[l]})$

- **BACKWARD PROPAGATION FOR LAYER l**

Input: $da^{[l]}$

output: $da^{[l-1]}$, $dw^{[l]}$, $db^{[l]}$

$dz^{[l]} = da^{[l]} * g^{[l]\prime}(z^{[l]})$

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

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

$da^{[l-1]} = w^{[l]T} \cdot dz^{[l]}$

$dz^{[l]} = w^{[l+1]T} dz^{[l+1]} * g^{[l]\prime}(z^{[l]})$

VECTORIZED

$dZ^{[l]} = dA^{[l]} * g^{[l]\prime}(Z^{[l]})$

$dW^{[l]} = \frac{1}{m} dZ^{[l]} \cdot A^{[l-1]T}$

$db^{[l]} = \frac{1}{m}$ np.sum$(dZ^{[l]}$, axis=1, keepdims=True)

$dA^{[l-1]} = W^{[l]T} \cdot dZ^{[l]}$

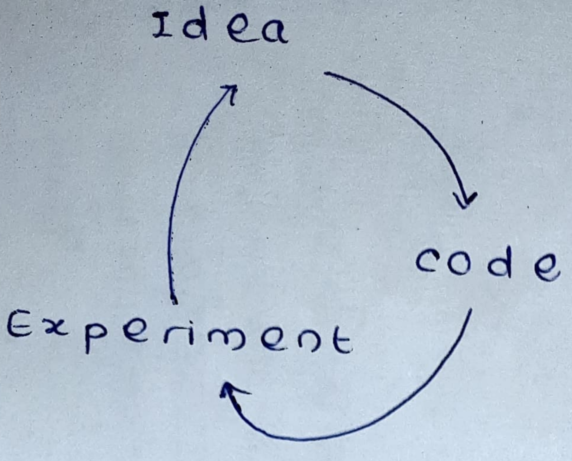
⊛ For final layer : $da^{[l]} = -\frac{y}{a} + \frac{(1-y)}{(1-a)}$

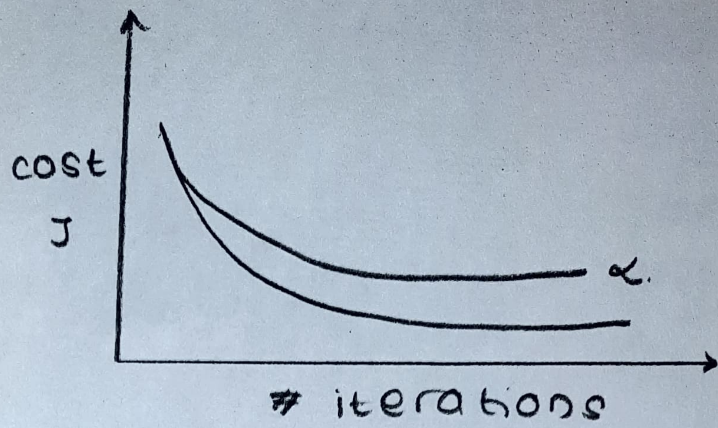
vectorized : $dA^{[l]} = \frac{-y^{(1)}}{a^{(1)}} + \frac{(1-y^{(1)})}{(1-a^{(1)})} + \dots + \frac{-y^{(m)}}{a^{(m)}} + \frac{(1-y^{(m)})}{(1-a^{(m)})}$

- **PARAMETERS & HYPERPARAMETERS**

$w^{[l]}, b^{[l]}$

↓
control parameters
↓
Learning rate (α)
#iterations
#hidden layer L
hidden units $n^{[1]}, n^{[2]}, \dots$
choice of activation function





{IMP}

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

$$dw^{[L]} = \frac{1}{m} dz^{[L]} A^{[L-1]T}$$

$$db^{[L]} = \frac{1}{m} \text{np.sum}(dz^{[L]}, \text{axis}=1, \text{keepdims}=\text{True})$$

$$dz^{[L-1]} = W^{[L]T} dz^{[L]} * g'^{[L-1]}(z^{[L-1]})$$

element wise multiplication

$$\therefore dz^{[1]} = A^{[1]} - Y$$

$$dw^{[1]} = \frac{1}{m} dz^{[1]} A^{[0]T} \qquad A^{[0]T} = X^{T}$$

$$db^{[1]} = \frac{1}{m} \text{np.sum}(dz^{[1]}, \text{axis}=1, \text{keepdims}=\text{True})$$