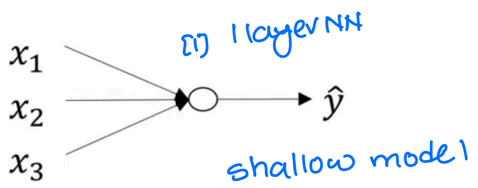
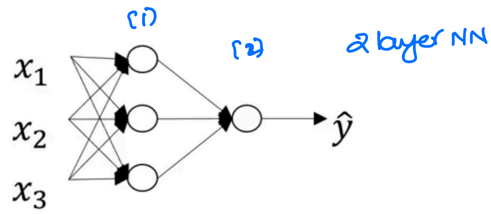
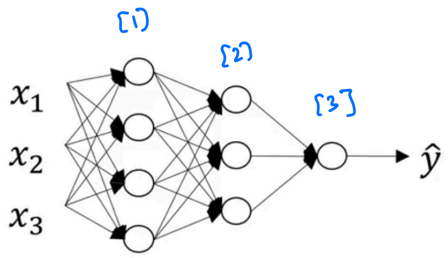


logistic regression

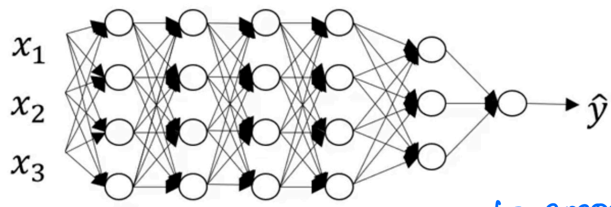


1 hidden layer



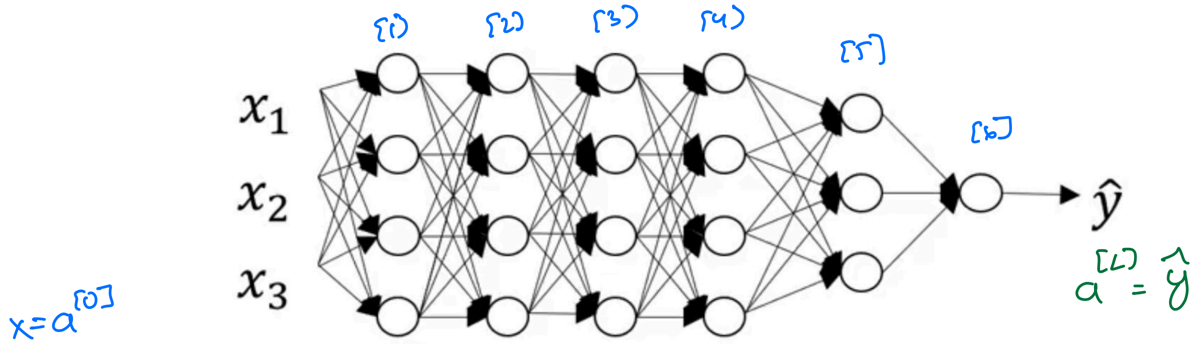
2 hidden layers

3 layer NN



5 hidden layers

deep model



5 hidden layers

L = No. of layer = 5

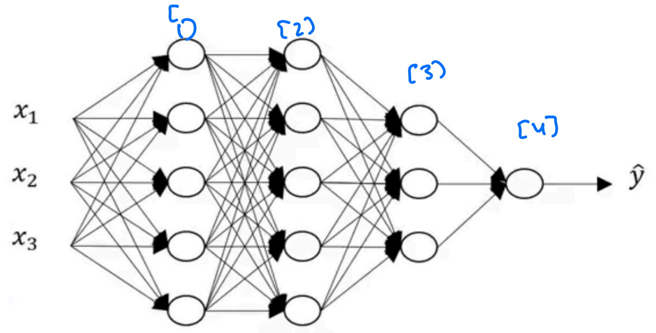
$n^{[l]}$ = no. of units in layers l $\{n^{[1]} = 4, n^{[2]} = 4, n^{[3]} = 4, n^{[4]} = 4, n^{[5]} = 3, n^{[6]} = 1\}$

$n^{[0]} = 3$

$a^{[l]}$ = activations in layer l $\quad a^{[l]} = g^{[l]}(z^{[l]})$

$W^{[l]}$ = weights for $z^{[l]}$ $\quad b^{[l]}$ = weights to compute $z^{[l]}$

Lec 2: Forward Propagation in a Deep Network:-



single training example : x

$$a^{[0]} = x$$
$$z^{[1]} = W^{[1]} a^{[0]} b^{[1]}$$
$$a^{[1]} = g^{[1]}(z^{[1]})$$
$$z^{[2]} = W^{[2]} a^{[1]} + b^{[2]}$$
$$a^{[2]} = g^{[2]}(z^{[2]})$$
$$z^{[3]} = W^{[3]} a^{[2]} + b^{[3]}$$
$$a^{[3]} = g^{[3]}(z^{[3]})$$
$$z^{[4]} = W^{[4]} a^{[3]} + b^{[4]}$$
$$a^{[4]} = g^{[4]}(z^{[4]})$$
$$\hat{y} = a^{[4]}$$

General equation

$$z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]}$$
$$a^{[l]} = g^{[l]}(z^{[l]})$$

vectorized for m example $A^{[0]} = X$

$$Z^{[1]} = W^{[1]} A^{[0]} + b^{[1]}$$
$$A^{[1]} = g^{[1]}(Z^{[1]})$$
$$Z^{[2]} = W^{[2]} A^{[1]} + b^{[2]}$$
$$A^{[2]} = g^{[2]}(Z^{[2]})$$

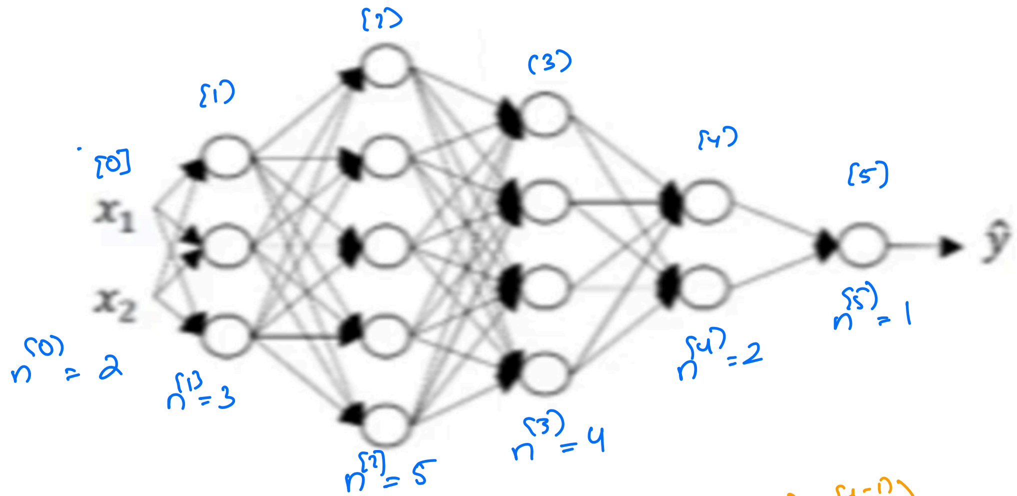
for loop
$l = 1, 2, 3$

$\vdots$

$$\hat{Y} = A^{[4]} = [a^{[4](1)} \cdots a^{[4](m)}]$$

$(1, m)$

Lec 3: Matrix Dimensions :-



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

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

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

$(3,1)$ $(3,2)$ $(2,1)$ $(3,1)$

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

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

$(5,1)$ $(5,3)$ $(3,1)$ $(5,1)$

$W^{[1]} = (3,2)$ $b^{[1]} = (3,1)$

$W^{[2]} = (5,3)$ $b^{[2]} = (5,1)$

$W^{[3]} = (4,5)$ $b^{[3]} = (4,1)$

$W^{[4]} = (2,4)$ $b^{[4]} = (2,1)$

$W^{[5]} = (1,2)$ $b^{[5]} = (1,1)$

Vectorized:-

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

$(3,1)$ $(3,2)$ $(2,1)$ $(3,1)$

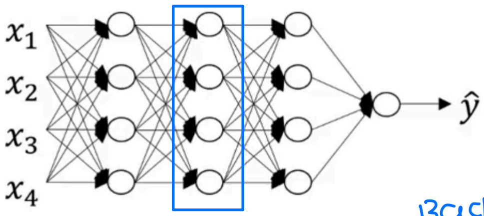
$$Z^{[1]} = W^{[1]} X + b^{[1]}$$

$(3,m)$ $(3,2)$ $(2,m)$ $(3,1)$

$Z^{[l]}$ $(n^{[l]}, m)$ $A^{[l]}$ $(n^{[l]}, m)$

$dZ^{[l]}$ $(n^{[l]}, m)$ $dA^{[l]}$ $(n^{[l]}, m)$

Lec5: Forward and Backward Functions:-



Forward

layer l: $W^{[l]}, b^{[l]}$

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

output: $a^{[l]}$

$$z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]}$$

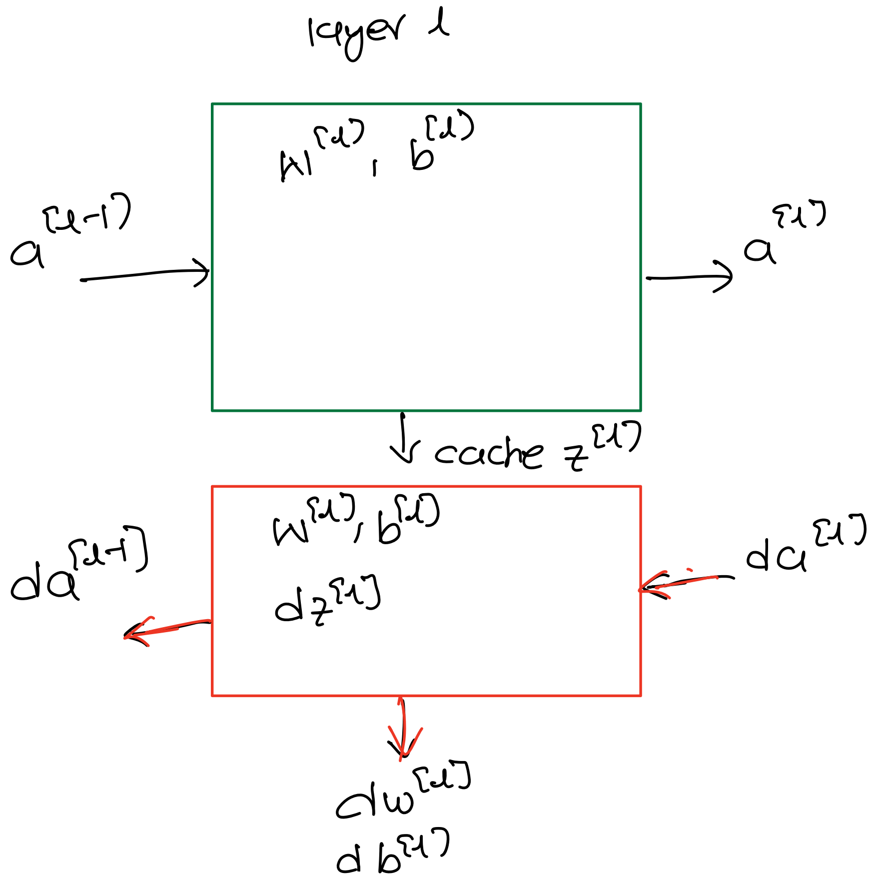
$$a^{[l]} = g^{[l]}(z^{[l]})$$

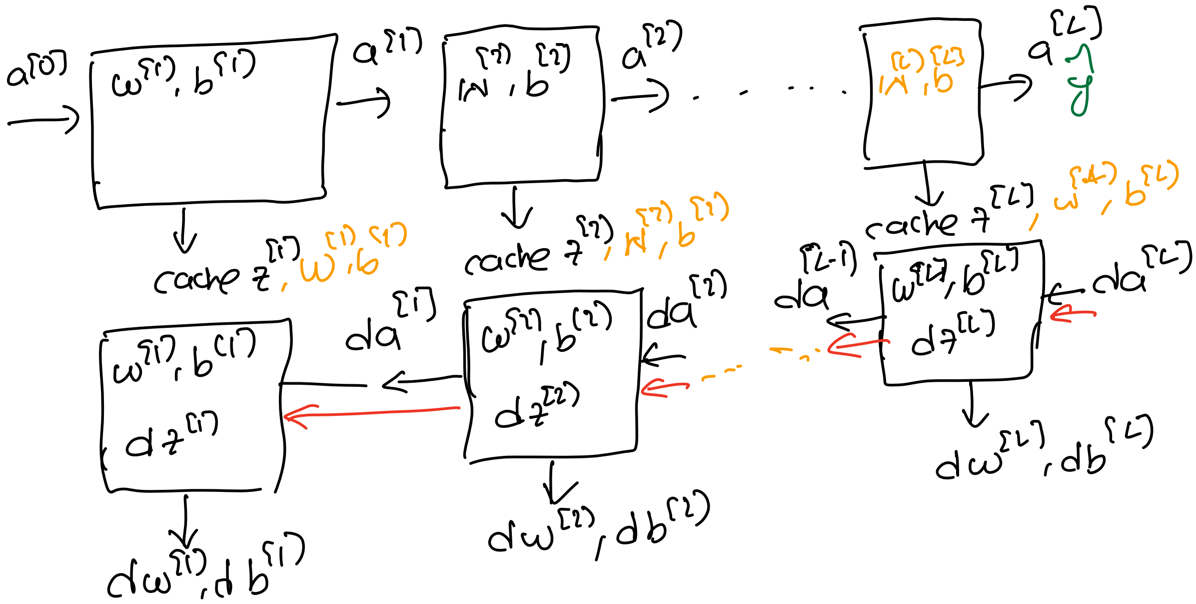
cache $z^{[l]}$

Backward:-

input: $da^{[l]}$, cache($z^{[l]}$)

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





$$w^{[l]} = w^{[l]} - \alpha\, dw^{[l]}$$
$$b^{[l]} = b^{[l]} - \alpha\, db^{[l]}$$

Forward Function:

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

Backward propagation:

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

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

$$dz^{[l]} = da^{[l]} * g^{[l]'}(z^{[l]})$$
$$dW^{[l]} = dz^{[l]} \cdot a^{[l-1]T}$$
$$db^{[l]} = dz^{[l]}$$
$$da^{[l-1]} = W^{[l]T} \cdot dz^{[l]}$$
$$dz^{[l]} = W^{[l]T} \cdot dz^{[l]} * g^{[l]'}(z^{[l]})$$
$$da^{[L]} = -\frac{y}{a} + \frac{(1-y)}{(1-a)}$$

Vectorized:

$$dz^{[l]} = dA^{[l]} * g^{[l]'}(z^{[l]})$$
$$dW^{[l]} = \frac{1}{m} dz^{[l]} \cdot A^{[l-1]T}$$
$$db^{[l]} = \frac{1}{m} np.sum(dz^{[l]}, axis=1, keepdims=True)$$
$$dA^{[l-1]} = W^{[l]T} \cdot dz^{[l]}$$

$\alpha(y, a^{[L]})$

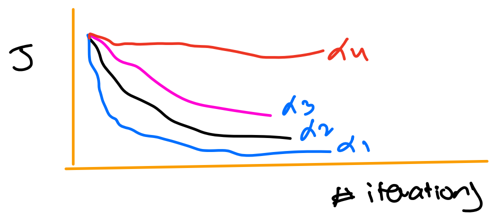
$$dA^{[L]} = \left(-\frac{y^{(1)}}{a^{(1)}} + \frac{(1-y^{(1)})}{(1-a^{(1)})} + \cdots \cdots - \frac{y^{(m)}}{a^{(m)}} + \frac{(1-y^{(m)})}{(1-a^{(m)})}\right)$$

Parameters vs Hyperparameters

Parameters: $W^{[1]}, b^{[1]}, W^{[2]}, b^{[2]} \ldots$

Hyper parameters: learning rate (α)
iterations
hidden layers L
hidden units $n^{[1]}, n^{[2]}, \ldots$
activation function ReLU, tanh

- momentum, minibatch, regularization.



Final Equations

$$
\begin{aligned}
Z^{[1]} &= W^{[1]}X + b^{[1]} \\
A^{[1]} &= g^{[1]}(Z^{[1]}) \\
Z^{[2]} &= W^{[2]}A^{[1]} + b^{[2]} \\
A^{[2]} &= g^{[2]}(Z^{[2]}) \\
&\vdots \\
A^{[L]} &= g^{[L]}\left(Z^{[L]}\right) = \hat{Y}
\end{aligned}
$$

Forward Function

Backward Function

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

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

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

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

Note that * denotes element-wise multiplication)

$$\vdots$$

$$dZ^{[1]} = W^{[2]^T} dZ^{[2]} * g'^{[1]}(Z^{[1]})$$

$$dW^{[1]} = \frac{1}{m} dZ^{[1]} A^{[0]^T}$$

Note that $A^{[0]^T}$ is another way to denote the input features, which is also written as X^T

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