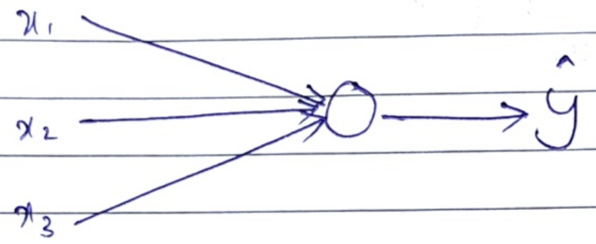
Week 4:

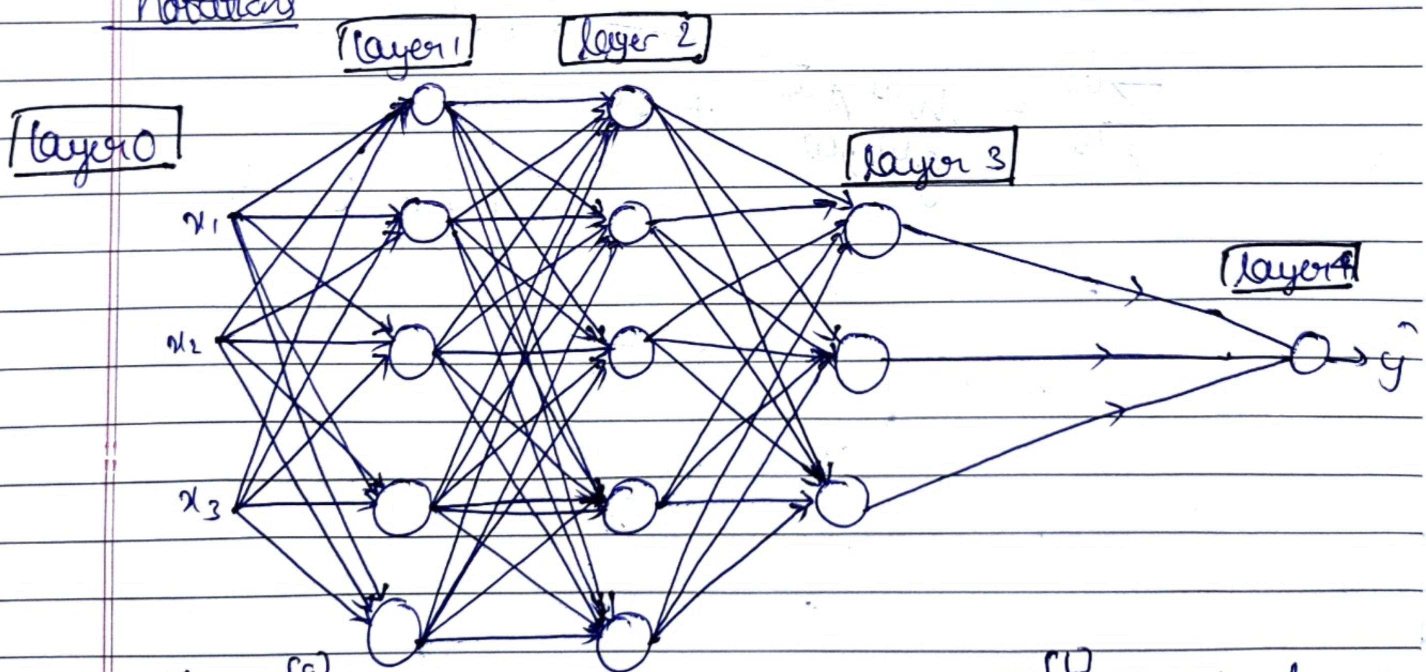
Video 1: Deep L layer neural network

logistic regression - shallow
1 layer NN



Deep Neural Network

notations



$X = a^{[0]}$
$l = 4$ (layers)
$n^{[l]}$ = units in layer l
$a^{[l]}$ = activations in layer l
$a^{[l]} = g^{[l]}(z^{[l]})$

$w^{[l]}$ - wts. for $z^{[l]}$
$b^{[l]}$ - wts

Forward Propagation in Deep ~~Layer~~ Network

$x = a^{[0]}$

$$X: \quad z^{[1]} = W^{[1]} x + b^{[1]}$$

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

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

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

$$\vdots$$

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

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

general equation:

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

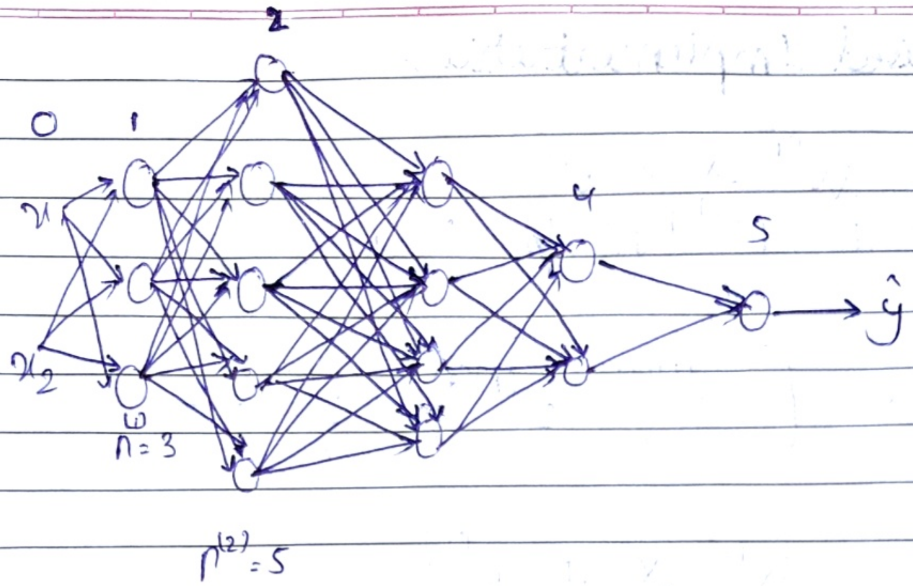
$$a^{[l]} = g^{[l]} z^{[l]}$$

$X = A^{[0]}$

$$Z^{[1]} = W^{[1]} A^{[0]} + b^{[1]}$$

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

for $l = 1-4$



$$z^{[1]} = w^{[1]} . x + b^{[1]}$$
$$(3,1) \quad (3,2)(2,1) + (3,1)$$
$$\boxed{(n^{[1]}, 1)} \quad \boxed{(n^{[1]}, n^{[0]})} (n^{[0]}, 1) \boxed{(n^{[1]}, 1)}$$

$$\boxed{w^{[1]} = (n^{[1]} \; n^{[0]})}$$
$$w^{[2]} = (5,3) \rightarrow (n^{[2]} \; n^{[1]})$$
$$z^{[2]} = w^{[2]} \cdot a^{[1]} + b^{[2]}$$
$$(5,1) \quad (5,3) . (3,1)$$
$$w^{[3]} = (4,5)$$
$$w^{[4]} = (2,4)$$
$$w^{[5]} = (1,2)$$

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

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

Vectorised Implementation

$$\underset{(n^{[1]},1)}{Z^{[1]}} = \underset{(n^{[1]},n^{[0]})}{W^{[1]}} \cdot \underset{(n^{[0]},1)}{x} + \underset{(n^{[1]},1)}{b^{[1]}}$$

$$Z = \begin{bmatrix} | & | & | & \\ z^{[1](1)} & z^{[1](2)} & z^{[1](3)} & \cdots \\ | & | & | & \end{bmatrix}$$

$$\underset{(n^{[1]},m)}{Z^{[1]}} = \underset{(n^{[1]},n^{[0]})}{W^{[1]}} \cdot \underset{(n^{[0]},m)}{X} + \underset{(n^{[1]},1)}{b^{[1]}}$$

$z^{[l]}, a^{[l]} : (n^{[l]}, 1)$

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

$l = 0$

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

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

Why Deep Representation?

* Deep neural networks make relations with data from simpler to complex. In each layer it tries to make a relation to the previous layer.

* Face recognition:

image → edges → face parts → face → desired face.

* Audio recognition:

audio → low level sound features → phonemes → words → sentences

Building Blocks of Deep Neural Networks

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

cache - storing value of z for backward propagation

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

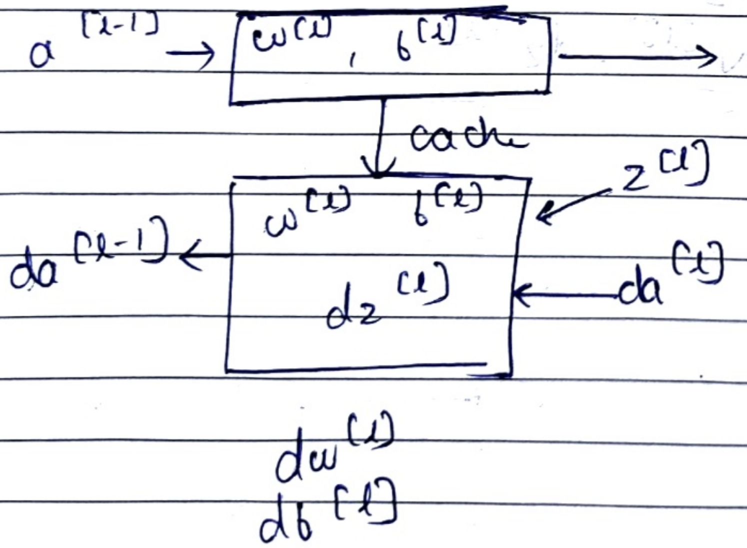
$z^{[l]} = w^{[l]} a^{[l-1]} + b^{[l]}$

cache $z^{[l]}$

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

Backward: input $da^{[l]}$
output $da^{[l-1]}$

layer l



Forward Propagation of layer l

Input : $a^{[l-1]}$ — $w^{[l]}$, $b^{[l]}$

Output : $a^{[l]}$ cache $(z^{[l]})$ $\quad z^{[l]} = w^{[l]} a^{[l-1]} + b^{[l]}$

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

Vectorised

$$Z^{[l]} = W^{[l]} . A^{[l-1]} + b^{[l]}$$
$$A^{[l]} = g^{[l]}(Z^{[l]})$$

Backward Propagation of layer l

Input : $da^{[l]}$

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

$$dz^{[l]} = da^{[l]} * g^{[l]'}(z^{[l]})$$
$$dW^{[l]} = dz^{[l]} . a^{[l-1]T}$$
$$db^{[l]} = dz^{[l]}$$
$$da^{[l-1]} = W^{[l]T} . dz^{[l]}$$

Vectorised

Hyperparameters and Parameters

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

* Hyperparameters: (i) learning rate α
(ii) iterations
(iii) hidden layers L
(iv) hidden units $n^{[1]}, n^{[2]} \dots$
(v) choice of activation function

* hyperparameters control parameters