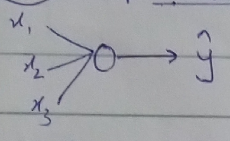
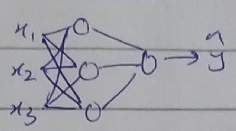
Week-4 Deep Neural Network

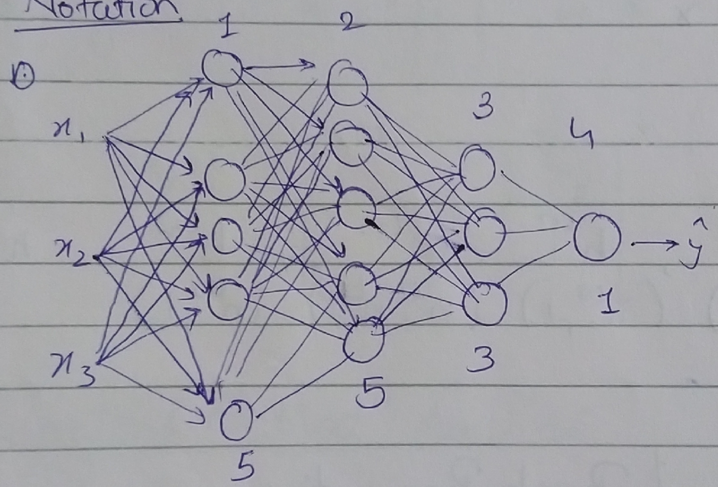


1 layerNN



2layerNN

Notation



$L = 4$ (# layers)

$n^{[l]}$ = # units in layer l $\quad n^{[1]}=5 \quad n^{[2]}=5 \quad n^{[3]}=3 \quad n^{[4]}=1$

$a^{[l]}$ = activat^{ns} in " " $\quad n^{[0]} = n_x = 3$

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

$w^{[l]}$ = weights for $z^{[l]}$

$b^{[l]}$ = bias for $z^{[l]}$

$a^{[0]} = x \qquad a^{[l]} = \hat{y}$

Forward Propagat^n

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

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

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

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

Layer 3 ⋮

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

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

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

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

Vectorized

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

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

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

Matrix Dimensions

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

$(3,1) \leftarrow (3,2)\ (2,1)$

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

$\begin{bmatrix} \circ \\ \circ \\ \circ \end{bmatrix} \quad \begin{bmatrix} \circ & \circ \\ \circ & \circ \\ \circ & \circ \end{bmatrix} \quad \begin{bmatrix} \circ \\ \circ \end{bmatrix}$

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

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

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

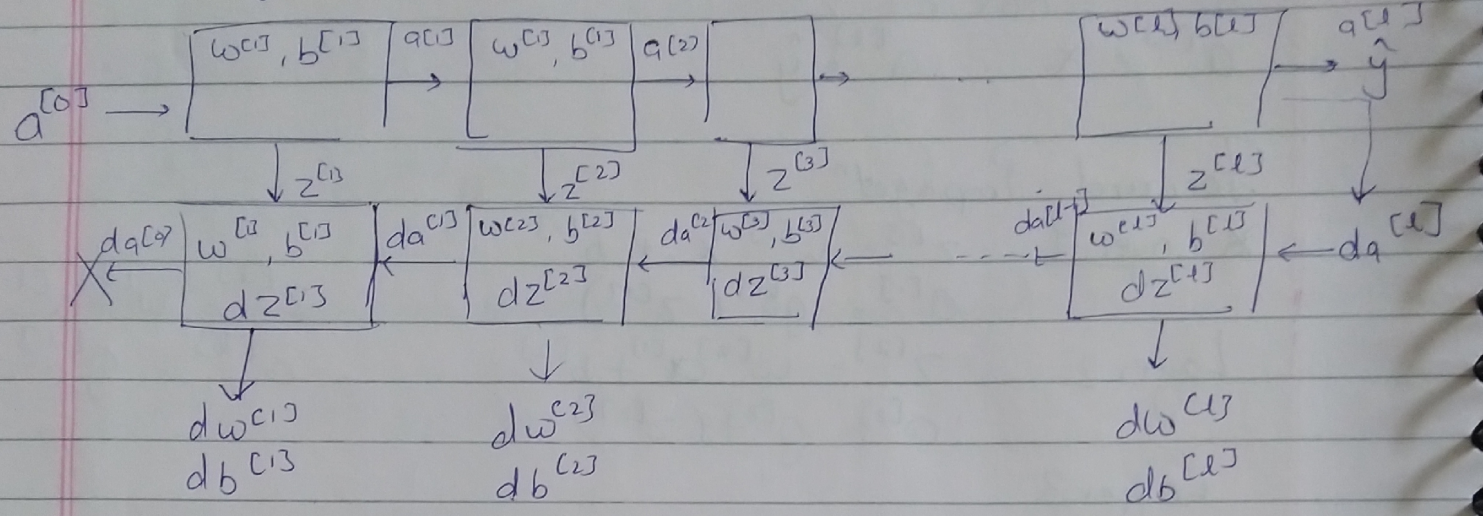
$Z^{[l]} = (n^{[l]}, m)$

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

$a^{[l]} = \cancel{g^{[l]}(z^{[l]})}\ (n^{[l]}, m) = dA^{[l]}$

There are functions you can compute with a "small" L-layer deep neural network that shallower networks require exponentially more hidden units to compute.

Building Blocks of NN



Backward Propagation

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

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

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

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

Vectorized

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

$dw^{[l]} = 1/m \; dZ^{[l]} \cdot A^{[l-1]T}$

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

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

For the last layer,

$$da^{[l]} = -\frac{y}{a} + \frac{(1-y)}{(1-a)}$$

y = x**2

$y = x^2$

$\frac{dy}{dx} = 2x$