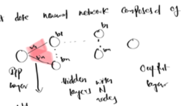
what dose neural network composed of



weights

→ They are weights that we assign to previous nodes to determine the Next nodes value

Now we also have

bias

→ why do we Need bias because we dont compute zeros to avoid that risk

weight	&	bias
for every connected link except I/P		for every node ~~PR~~

→ Now how to determin node value ⇒

$Z =$ weights & previous values + bias of that node
and ↓ node vall

→ After getting Z we pass the values to the Activation function to djust all of our parameters

⇒ Activated value = function(z)

⇒ we do this till we get our output layer
⇒ we get our output layer values which are Ypred

Now what
→ we find the loss of our first cycle

Ypred - Yoriginal

⇒ pass it to the function which we call a loss function → which is

binary classification

$$\Rightarrow eq^n = -(y_i \ln(\hat{y}_i) + (1-y_i)\ln(1-\hat{y}_i))$$

Then we find cost
this is nothing but sumation of the values

$$\Rightarrow \frac{1}{m}\sum_{i=1}^{m} L(\hat{y}_i, y_i)$$

$$\text{cost} \rightarrow \frac{1}{m}\sum_{i=1}^{m} -(y_i \ln(\hat{y}_i) + (1-y_i)\ln(1-\hat{y}_i))$$

Now after doing this
here comes the fun part

How to reduce cost
to reduce the cost we need to tune the parameter which are w, b

we use gradient to minimize our cost

$$\nabla C = \left[\frac{\partial C}{\partial w_1}, \frac{\partial C}{\partial w_2}, \frac{\partial C}{\partial w_n}, \frac{\partial C}{\partial b_1}, \frac{\partial C}{\partial b_2}, \frac{\partial C}{\partial b_n}\right]$$

get the loss function
→ use partial derivative to find a slope
→ because gradient gives us max we need to go opposite with the smaller learning rate to avoid over-fitting

→ Rinse and Repeat.

So yea the good old chain rule in partial derivaties

$$\frac{\partial C}{\partial w^L} = \frac{\partial C}{\partial A^L}\frac{\partial A^L}{\partial Z^L}\frac{\partial Z^L}{\partial w^L}$$

$$\frac{\partial C}{\partial b^L} = \frac{\partial C}{\partial A^L}\frac{\partial A^L}{\partial Z^L}\frac{\partial Z^L}{\partial b^L}$$

→ Now what to do after finding derivatives "simply put we go othe side"

$$w = w - \alpha\frac{\partial C}{\partial w} \qquad \alpha = \text{learning rate}$$

$$b = b - \alpha\frac{\partial C}{\partial b}$$

→ Some derivations for our code

$$\frac{\partial C}{\partial w^L} = \frac{\partial C}{\partial A^L}\frac{\partial A^L}{\partial Z^L}\frac{\partial Z^L}{\partial w^L}$$

$$\frac{\partial C}{\partial A^L} = \frac{\partial}{\partial A^L}\frac{1}{m}(L)$$

$$= \frac{1}{m}\frac{\partial}{\partial A^L} -(y \ln A^L + (1-y)\ln(1-A^L))$$

$$= \frac{1}{m} - \frac{\partial}{\partial A^L}\frac{y}{A^L} + \frac{(1-y)}{(1-A^L)} \cdot -1$$

$$\frac{\partial C}{\partial A^L} = -\frac{1}{m}\left(\frac{y}{A^L} - \frac{(1-y)}{(1-A^L)}\right)$$

$$\frac{\partial A^L}{\partial Z^L} = \left(\frac{1}{1+e^{-z}}\right)'$$

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

$$\frac{\partial A^L}{\partial Z^L} = \frac{e^{-z}}{1+e^{-z}}$$

$$\frac{\partial Z^L}{\partial w^L} = (w^L A^{(L-1)} + b^L)'$$

$$\frac{\partial Z^L}{\partial w^L} = A^{(L-1)}$$

$$\text{Now} \Rightarrow \frac{\partial C}{\partial w^L} = \frac{\partial C}{\partial A^L}\frac{\partial A^L}{\partial Z^L}\frac{\partial Z^L}{\partial w^L}$$

$$= -\frac{1}{m}\left(\frac{y}{A^L} - \frac{(1-y)}{A(1-A)}\right) * \frac{e^{-z}}{1+e^{-z}}$$

but we know
$$\frac{1}{1+e^{-z}} = A^L$$
we can use $\frac{e^{-z}}{1+e^{-z}} = A^L(1-A^L)$

$$= -\frac{1}{m}\left(\frac{y}{A^L} - \frac{(1-y)}{(1-A^L)}\right) * A^L(1-A^L) * A^{(L-1)}$$

$$= -\frac{1}{m}(y(1-A^L) - (1-y)A^L) * A^{(L-1)}$$

$$= -\frac{1}{m}(y - yA^L - (A^L - yA^L))$$

$$\frac{\partial C}{\partial w^L} = \frac{1}{m}\left(\underbrace{A^L - y}_{n^L \times m}\right)\underbrace{(A^{[L-1]})^T}_{m \times n^{L-1}} \rightarrow \text{transpose}$$

$$\frac{\partial C}{\partial b^L} = \frac{\partial C}{\partial A^L}\frac{\partial A^L}{\partial Z^L}\frac{\partial Z^L}{\partial b^L} \{\Rightarrow 1$$

$$\frac{\partial C}{\partial b^L} = \frac{1}{m}\sum_{i=1}^{m}(A^{Li} - y^i)$$

The chain

$$\begin{tikzcd}
 & \frac{\partial C}{\partial A^{[3]}} \arrow[r] \arrow[dl] \arrow[d] & \frac{\partial C}{\partial b^3} \\
\frac{\partial C}{\partial w^3} & \frac{\partial C}{\partial A^{[2]}} \arrow[r] \arrow[dl, dash] \arrow[d] & \frac{\partial C}{\partial b^2} \\
\frac{\partial C}{\partial w^2} & \frac{\partial C}{\partial A^{[1]}} \arrow[r] \arrow[dl] & \frac{\partial C}{\partial b^1} \\
\frac{\partial C}{\partial w^1} & &
\end{tikzcd}$$

$$\frac{\partial C}{\partial w^2} = \underline{\frac{\partial C}{\partial A^{[3]}}\frac{\partial A^3}{\partial Z^3}}\,\underline{\frac{\partial Z^3}{\partial A^2}}\frac{\partial Z^2}{\partial w^2}$$

$$= \frac{\partial C}{\partial A^{[2]}}\frac{\partial A^2}{\partial Z^2}\frac{\partial Z^2}{\partial w^2}$$

$$Z^{[3]} = w^{[3]}A^{[2]} + b^{[3]}$$

$$\frac{\partial C}{\partial A^{[3]}} * \frac{\partial A^3}{\partial Z^3} = \frac{1}{m}[A^3] - y$$

$$= (w^3)^T \frac{1}{m}(A^3 - y)$$