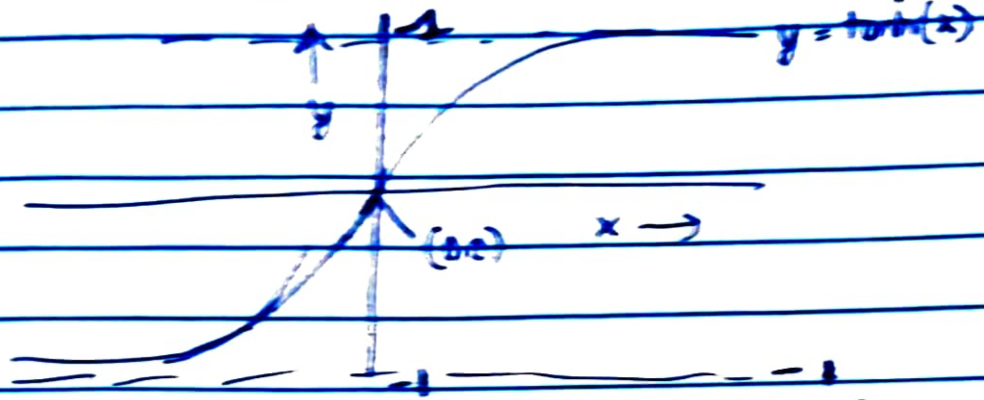
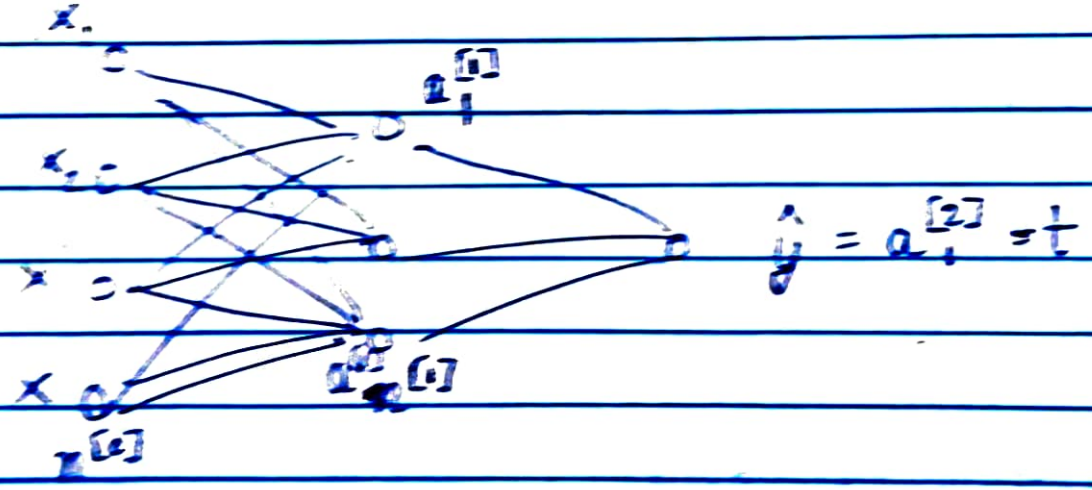
2. This is true, as the range of tanh is [-1,1], all values get centered around 0, which is the mid point of the range of tanh.



2

On the other hand sigmoid squishes values to a range [0,1] which is centered at 0.5 for the inputs on average, thereby having a positive ~~[illegible]~~ gradient on average.



$$\mathcal{L}(t, y) = -(y \log(t) + 1-y \log(1-t))$$

We have forward propogation :

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

$$a^{[1](i)} = ReLU(z^{[1](i)}) = \max(z^{[1](i)}, 0)$$

Now, derivative of ReLU = $\begin{cases} 1, & x \geq 0 \quad \text{(assumed at } x=0) \\ 0, & x < 0. \end{cases}$

$z = ReLU'(x)$

$= \text{signum}(x)$

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

$$t = a^{[2]} = \sigma(z^{[2](i)}) \quad \text{(Sigmoid)}$$

Vectorizing :

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

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

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

$$T = A^{[2]} = \sigma(Z^{[2]})$$

Now, For update, gradient descent —:

$$W^{[1]} = W^{[1]} - \alpha \frac{\partial \mathcal{L}}{\partial W^{[1]}}$$

$$W^{[2]} = W^{[2]} - \alpha \frac{\partial \mathcal{L}}{\partial W^{[2]}}$$

$$b^{[1]} = b^{[1]} - \alpha \frac{\partial \mathcal{L}}{\partial b^{[1]}}, \quad b^{[2]} = b^{[2]} - \alpha \frac{\partial \mathcal{L}}{\partial b^{[2]}}$$

Now, $\frac{\partial \mathcal{L}}{\partial W^{[2]}} = \frac{\partial \mathcal{L}}{\partial A^{[2]}} \frac{\partial A^{[2]}}{\partial Z^{[2]}} \frac{\partial Z^{[2]}}{\partial W^{[2]}}$ (Chain Rule)

$= \frac{\partial \mathcal{L}}{\partial A^{[2]}} = \frac{-Y}{T} + \frac{1-Y}{1-T} = \frac{T-Y}{T(1-T)}$

$\frac{\partial A^{[2]}}{\partial Z^{[2]}} = A^{[2]}(1-A^{[2]})$ $(\because \sigma'(z) = \sigma(z)(1-\sigma(z)))$

$= T(1-T)$

$\frac{\partial Z^{[2]}}{\partial W^{[2]}} = A^{[1]T}$

$\Rightarrow \frac{\partial \mathcal{L}}{\partial W^{[2]}} = (T-Y)\,A^{[1]T}$

8

Similarly, $\frac{\partial \mathcal{L}}{\partial b^{[2]}} = (T-Y)$. Vectorized $= \frac{1}{m}$ np.sum(dZ, axis=1, keepdims=True)

$= dZ$

$= \frac{\partial \mathcal{L}}{\partial W^{[1]}} = \frac{\partial \mathcal{L}}{\partial Z^{[2]}} \frac{\partial Z^{[2]}}{\partial A^{[1]}} \frac{\partial A^{[1]}}{\partial W^{[1]}}$

$= (T-Y)\,\mathrm{signum}(A^{[1]})\,A^{[0]T}$