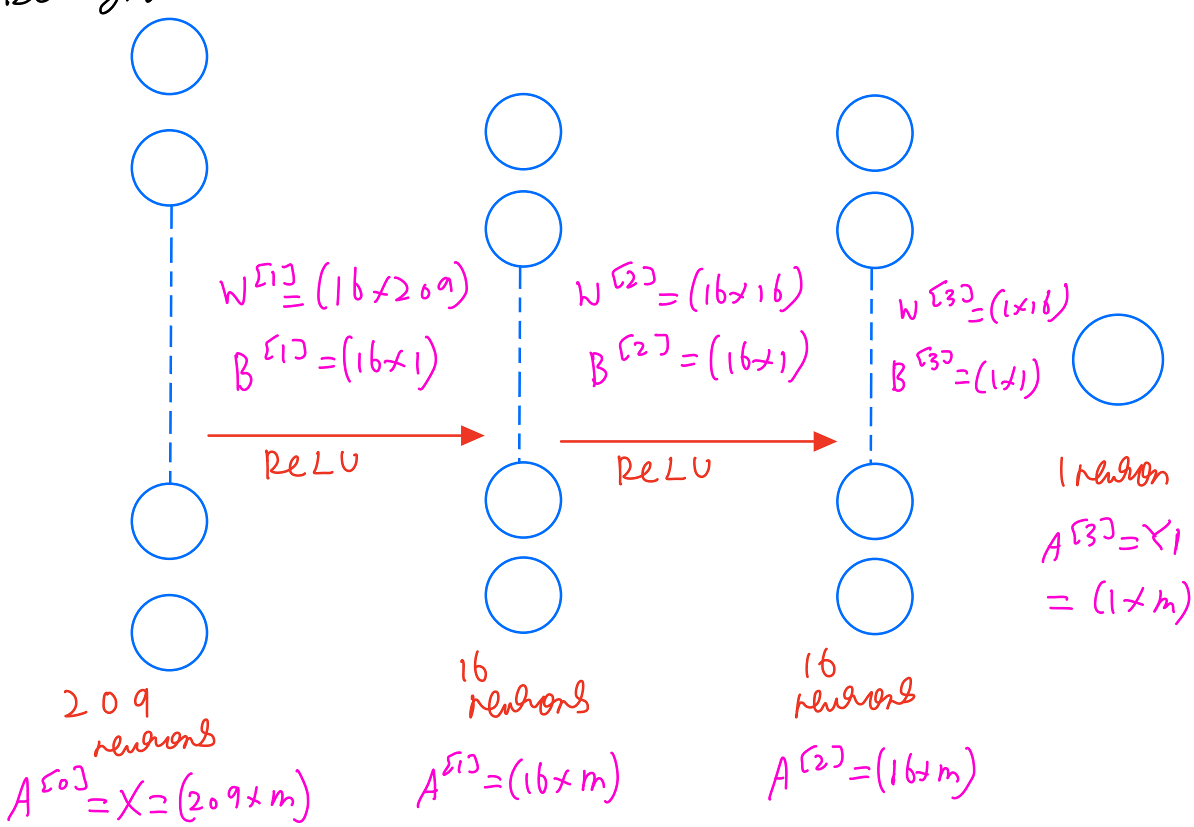
Initial data:

1.) $X = (m \times 209)$

2.) $Y = (m \times 1)$

→) Take transpose of X, $X = (209 \times m)$ and Y, $Y = (1 \times m)$

Design:



Forward Propagation:

1.) $Z^{[1]} = (W^{[1]} @ A^{[0]}) + B^{[1]}$

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

3.) $Z^{[2]} = (W^{[2]} @ A^{[1]}) + B^{[2]}$

4.) $A^{[2]} = ReLU(Z^{[2]})$

5.) $Z^{[3]} = (W^{[3]} @ A^{[2]}) + B^{[3]}$

6.) $A^{[3]} = Y1 = Z^{[3]}$

Backward Propogation:

1.) error $= Y1 - Y$ $(1 \times m)$

$16 \times m$

2.) $dW^{[l]} = \left(\frac{1}{m}\right) * \left(\text{error}^{[l]} @ A^{[l-1]}.T\right)$

3.) $dB^{[l]} = \left(\frac{1}{m}\right) * \sum \text{error}$

$(1 \times m)$

4.) $\text{error}^{[l-1]} = \left(W^{[l]}.T @ \text{error}\right) * g'\left(Z^{[l-1]}\right)$

(1×16)

5.) repeat from step 2

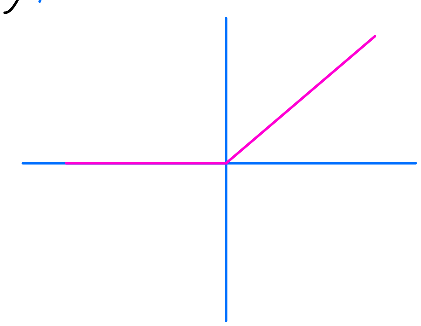
Update Parameters:

1.) $W^{[l]} = W^{[l]} - \alpha \, dW^{[l]}$

2.) $B^{[l]} = B^{[l]} - \alpha \, dB^{[l]}$

Activation Functions:

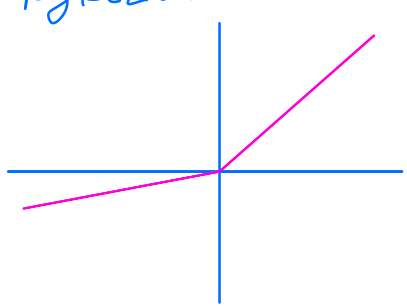
1.) ReLU: (Rectified Linear Unit)



$$f(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ x & \text{for } x > 0 \end{cases}$$

$$f'(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ 1 & \text{for } x > 0 \end{cases}$$

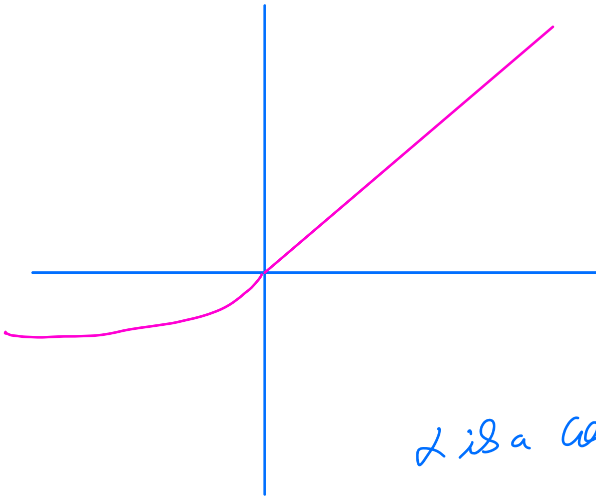
2.) leakyReLU:



$$f(x) = \begin{cases} 0.01x & \text{for } x \leq 0 \\ x & \text{for } x > 0 \end{cases}$$

$$f'(x) = \begin{cases} 0.01 & \text{for } x \leq 0 \\ 1 & \text{for } x > 0 \end{cases}$$

3.) ELU: (Exponential Linear Unit)

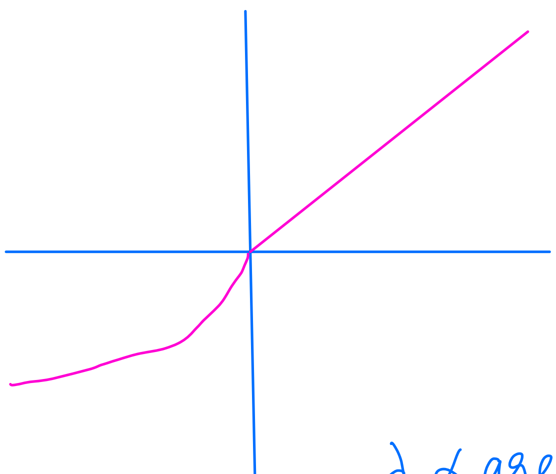


$$f(x) = \begin{cases} x & \text{for } x \geq 0 \\ \alpha(e^x - 1) & \text{for } x < 0 \end{cases}$$

$$f'(x) = \begin{cases} 1 & \text{for } x \geq 0 \\ \alpha e^x & \text{for } x < 0 \end{cases}$$

α is a constant

4.) SELU: (Scaled Exponential Linear Unit)



$$f(x) = \begin{cases} \underbrace{\lambda}_{\text{scale}} x & \text{for } x > 0 \\ \lambda\alpha(e^x - 1) & \text{for } x \leq 0 \end{cases}$$

$$f'(x) = \begin{cases} \lambda & x > 0 \\ \lambda\alpha e^x & x \leq 0 \end{cases}$$

λ, α are fixed already found values

→) All gave around the same accuracy but SELU and ELU were computationally expensive.

Data Preprocessing tried:

1.) Tried removing outliers but messed up the model

2.) Rather than one hot encoding 3 categorical features each one was labelled. But drastically reduced the accuracy.

Initial weights and Biases:

1.) For biases 3 things tried

i) initially all zero

ii) all set to 0.01

iii) all from random normal distribution

2nd one worked best

2.) For weigths 3 thing tried

i) Random values from normal distribution

ii) Glorot Intialization → mean = 0

$$\sigma^2 = \frac{1}{fan_{avg}} = \frac{1}{\frac{fan_{in} + fan_{out}}{2}}$$

fan_{in} = i/p neurons

fan_{out} = curr neurons

iii) He Initialization → mean = 0

$$\sigma^2 = \frac{1}{fan_{in}}$$

3rd one worked best