my NN:

$X \rightarrow$
$(784 \times m)$

$\downarrow$

$A^{[0]}$

$W^{[1]} = 16 \times 784$
$B^{[1]} = 16 \times 1$

$W^{[2]} = 16 \times 16$
$B^{[2]} = 16 \times 1$

$W^{[3]} = 10 \times 16$
$B^{[3]} = 10 \times 1$

784 neurons

16 neurons
$A^{[1]} = 16 \times m$

16 neurons
$A^{[2]} = 16 \times m$

10 neurons
$A^{[3]} = 10 \times m$
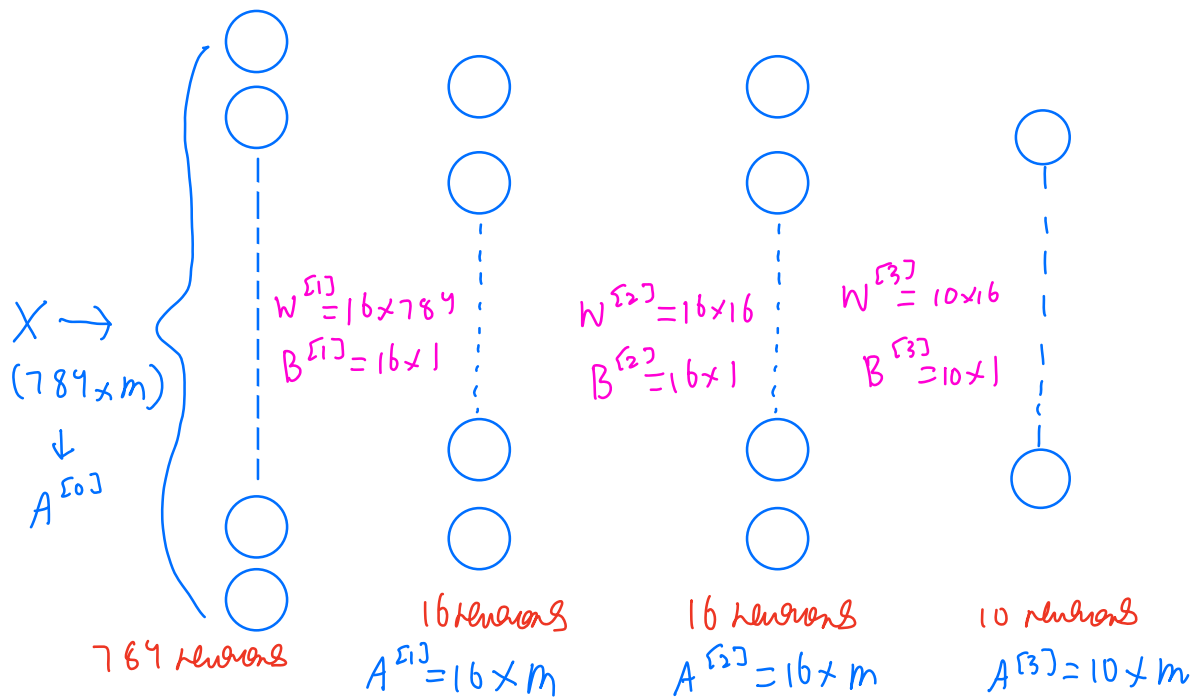
Forward Prop:

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

$Z = 10 \times m$

$(10 \times m)@$

One-hot-Encoding Y:

1.) The Y which are correct labels look like
$= [1, 0, 1, --- 7, ---]_{1 \times m}$

↓
means for 0th training label is 1 which we need
to convert to

$$
\begin{matrix}
0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9
\end{matrix}
\begin{bmatrix}
0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0
\end{bmatrix}
\longrightarrow \text{extended across m images}
$$

4.) So first one_hot_Y will have dimensions

$\underline{10 \times m}$

Nb. zeroes ((10, Y. Size))

5.) for i in range (0, m):

num = Y[i]   // tells which number is the label
             so this will tell which grow
             mein 1 set karna hain and
             i will tell which column

One_hot [num, i] = 1

Back Propogation:
Algo for back Prop:
error = Y1 - Y
$l = len (layers) - 1$
while True:
dW[l] = $\frac{1}{m}$ * (error @ A[l-1].T)

$$dB[l] = \frac{1}{m} * (\Sigma \text{ errors})$$

Then for next layer error becomes

$\Big($ error $= (w[l].T @ \text{ error}) * \text{ derivative} (z[l-1])$

$\quad$ $l = l-1$

if $l == 1$:
$\quad$ break