NEURAL NETWORK AND MNIST DATABASE

1- DATA

IMAGES USED TO TRAIN MODEL : 784 PX (28 X 28 PIXELES)

PIXEL VALUE : 0 -> COMPLETE BLACK ... 255 -> COMPLETE WHITE

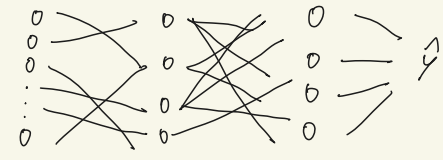
CLASSIFICATIONS : 10 -> FROM 0 to 9

TRANSPOSE

$$X = \begin{bmatrix} - x^{(1)} - \\ - x^{(2)} - \\ \vdots \\ - x^{(m)} - \end{bmatrix}^T = \begin{bmatrix} | & | & & | \\ x^{(1)} & x^{(2)} & \dots & x^{(m)} \\ | & | & & | \end{bmatrix}$$

$$X = \begin{bmatrix} a_{0,0} & a_{0,1} & a_{0,2} & \dots & a_{0,28} \\ a_{1,0} & a_{1,1} & a_{1,2} & \dots & a_{1,28} \\ \vdots & & & & \\ a_{28,0} & a_{28,1} & a_{28,2} & \dots & a_{28,28} \end{bmatrix}^T = \begin{bmatrix} a_{0,0} & a_{1,0} & a_{2,0} & \dots & a_{28,0} \\ a_{0,1} & a_{1,1} & a_{2,1} & \dots & a_{28,1} \\ \vdots & & & & \\ a_{0,28} & a_{1,28} & a_{2,28} & \dots & a_{28,28} \end{bmatrix}$$

2- NEURAL NETWORK PLAN



784 → INPUT LAYER

10 → 1ST HIDDEN LAYER

10 → OUTPUT LAYER

3- FORWARD PROPAGATION

1ST HIDDEN LAYER

$$A^{[0]} = X \quad (784 \cdot m) \quad \text{[INPUT]}$$

$$Z^{[1]}_{10} = W^{[1]}_{10 \cdot 784} A^{[1]}_{784} + b^{[1]}_{10} \quad \text{[LINEAR OPERATION]}$$

$$A^{[1]} = g\left(Z^{[1]}\right) = RELU\left[Z^{[1]}\right] \quad \text{[ACTIVATION]}$$

ACTIVATION

RELU

$$RELU(x) = \begin{cases} x \text{ if } x > 0 \\ 0 \text{ if } x < 0 \end{cases}$$

OUTPUT LAYER

$$Z^{[2]}_{10} = W^{[2]}_{10 \cdot 10} A^{[1]}_{10} + b^{[2]}_{10} \quad \text{[LINEAR OPERATION]}$$

$$A^{[2]} = SOFTMAX\left(Z^{[2]}\right) \quad \text{[ACTIVATION]}$$

OUTPUT LAYER → SOFTMAX ACTIVATION → PROBABILITY

$$\begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \end{bmatrix} \rightarrow \left[\frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \right] \rightarrow \begin{bmatrix} 0.02 \\ 0.90 \\ 0.05 \\ 0.01 \\ 0.02 \end{bmatrix}$$

$$\frac{e^{z_1}}{e^{z_1} + e^{z_2} + e^{z_3} + e^{z_4} + e^{z_5}} \quad \frac{e^{z_2}}{e^{z_1} + e^{z_2} + e^{z_3} + e^{z_4} + e^{z_5}}$$

$$\frac{e^{z_3}}{e^{z_1} + e^{z_2} + e^{z_3} + e^{z_4} + e^{z_5}} \quad \frac{e^{z_4}}{e^{z_1} + e^{z_2} + e^{z_3} + e^{z_4} + e^{z_5}}$$

$$\frac{e^{z_5}}{e^{z_1} + e^{z_2} + e^{z_3} + e^{z_4} + e^{z_5}}$$

4- BACK PROPAGATION

- $dz^{[2]} = Y - \hat{Y}$ of output LAYER
- $dz^{[1]} = Y - \hat{Y}$ of HIDDEN LAYERS
- dw = TOTAL WEIGHTS ERROR
- db = TOTAL BIASES ERROR
- $[2]$ = OUTPUT LAYER
- $[1]$ = 1ST HIDDEN LAYER

OUTPUT LAYER

ONE HOT ENCODER

$$\begin{bmatrix} 0.1 \\ 0.2 \\ 0.05 \\ 0.03 \\ 0.92 \end{bmatrix} \quad Y = 4 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix}$$

$$dZ^{[2]} = A^{[2]} - Y$$

[ERROR OF SECOND LAYER]

$$dW^{[2]} = \frac{1}{m} dZ^{[2]} A^{[1]T}$$

$$db^{[2]} = \frac{1}{m} \sum dZ^{[2]}$$

HIDDEN LAYER

$$dZ^{[1]} = W^{[2]T} dZ^{[2]} \cdot g'(Z)$$

[ERROR FROM SECOND LAYER, MULTIPLIED BY WEIGHTS OF LAST LAYER IN REVERSE, MULTIPLIED BY $G'(Z)$ TO UNDO ACTIVATION]

$$dW^{[1]} = \frac{1}{m} dZ^{[1]} X^T$$

$$db^{[1]} = \frac{1}{m} \sum dZ^{[1]}$$

5- LEARNING : GRADIENT DESCENT

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

$$b^{[1]} = b^{[1]} - \alpha \, dW^{[1]}$$

$$W^{[2]} = W^{[2]} - \alpha \, db^{[2]}$$

$$b^{[2]} = b^{2} - \alpha \, db^{[1]}$$