I/P layer    hidden layer    6/P layer



:Img

28 [grid]

28

28 × 28

784 pixels

→ Consign of 3 parts.
↳ forward propagation
↳ Backward Propagation
↳ Update Parameters.

784    10    10

[1]    [2]

---

① forward propagation.

$A^{[0]} = X$ (784 × m)

$Z^{[1]} = W^{[1]} \cdot A^{[1]} + b^{[1]}$

   10×784   784×m   10×1

10×m ← 10×m

$A^{[1]} = g(Z^{[1]}) = ReLU(Z_1)$

weight matrix: Initially any appreissible value

b : bias.

$\alpha$ - learning rate

→ $ReLU = \begin{cases} x & ; x > 0 \\ 0 & ; x \leq 0. \end{cases}$

Activation function:

$Z^{[2]} = W^{[2]} A^{[1]} + b^{[2]}$

   10×10   10×m   10×1

10×m ←

$(e^{1.3}) / (e^{1.3} + e^{1.5} + e^{2.2} + e^{0.7} + e^{1.1})$

→ 0.02

$A^{[2]} = Softmax(Z^{[2]})$

$\begin{bmatrix} 1.3 \\ 1.5 \\ 2.2 \\ 0.7 \\ 1.1 \end{bmatrix} \rightarrow \dfrac{e^{Z_i}}{\sum\limits_{j=1}^{x} e^{Z_j}} \rightarrow \begin{bmatrix} 0.02 \\ 0.96 \\ 0.05 \\ 0.01 \\ 0.02 \end{bmatrix}$

f/p : — take an image and get prediction out of it.

---

② Backward Propagation How much the predicted values deviated from the actual label. ⟹ Gives an Error and adjust the weight matrix and bias accordingly.

$dZ^{[2]}$ : Error of second layer. = $A^{[2]} - y$

                          10×m   10×m

10×m

$\dfrac{dW^{[2]}}{10×10} = \dfrac{1}{m} dZ^{[2]} A^{[1]T}$

               10×m   m×10

$db^{[1]} = \dfrac{1}{m} \sum dz^{[2]}$

$dz^{[1]}$ : Error of first layer.

   $= W^{[2]T} dz^{[2]} * g'(Z^{[1]})$

$\dfrac{dW^{[1]}}{10×784} = \dfrac{1}{m} dz^{[1]} X^T$

            10×m   m×784

$db^{[1]} = \dfrac{1}{m} \sum dz^{[1]}$

③ Update. Parameter

$$\omega^{[1]} = \omega^{[1]} - \alpha \, d\omega^{[1]}$$

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

$$\omega^{[2]} = b^{[2]} - \alpha \, d\omega^{[2]}$$

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

forward propogation

Backward propogation

till 'w' & 'b' get adjusted.

Update Parameter