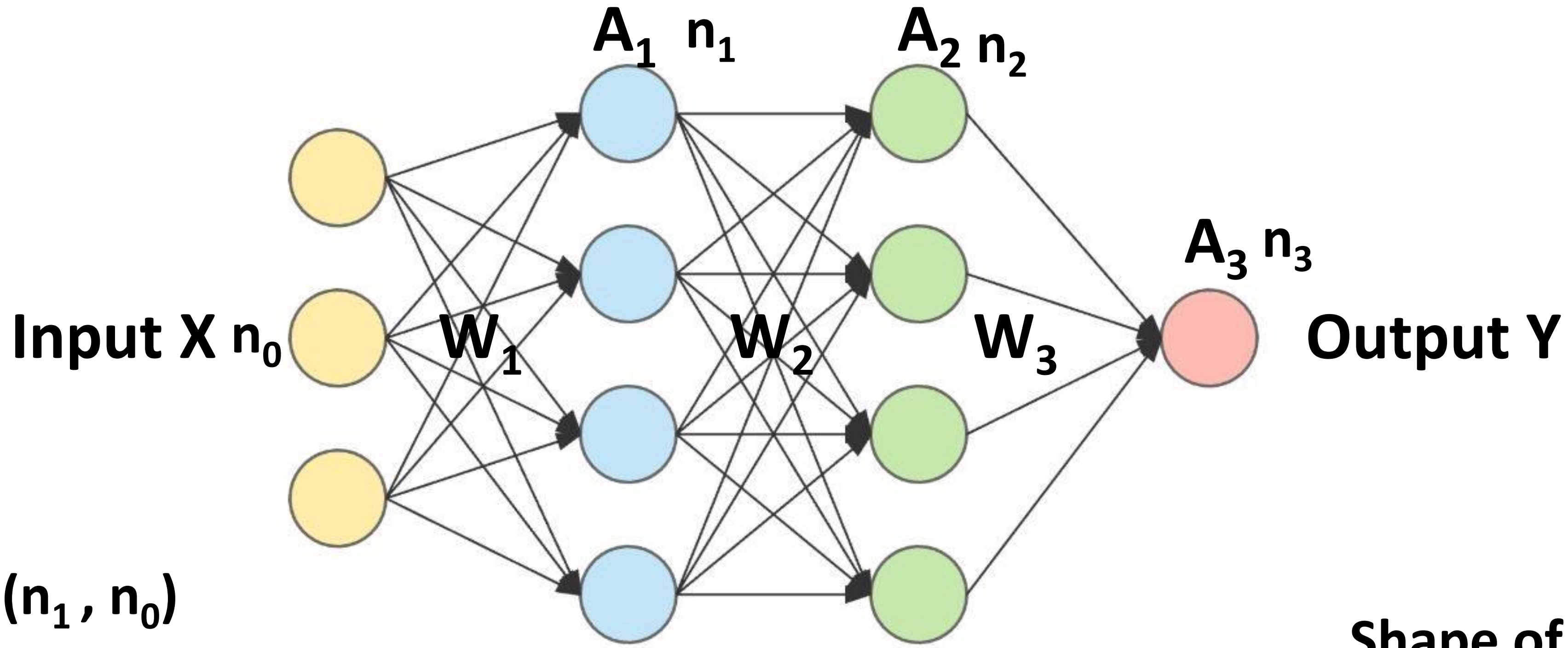
CODING
<LANE>

NOTES

m = total number of observations



Shape of W_1 = (n_1, n_0)

Shape of W_2 = (n_2, n_1)

Shape of W_3 = (n_3, n_2)

Shape of B_1 = $(n_1, 1)$

Shape of B_2 = $(n_2, 1)$

Shape of B_3 = $(n_3, 1)$

$A_1 = f_1(Z_1)$ $\quad$ $A_2 = f_2(Z_2)$ $\quad$ $A_3 = \text{Sigmoid}(Z_3)$

$$Z_k = W_k A_{k-1} + B_K$$

$$\text{Cost} = -\frac{1}{m}\sum_{i=1}^{m}[y_i * log(a_i) + (1 - y_i) * log(1 - a_i)]$$

$$\text{Loss}_i = -[y_i * log(a_i) + (1 - y_i) * log(1 - a_i)]$$

Shape of X = (n_0, m)

Shape of A_1 = (n_1, m)

Shape of A_2 = (n_2, m)

Shape of A_3 = (n_3, m)

Shape of Y = (n_3, m)

$Loss_i = -[y_i * log(a_i) + (1 - y_i) * log(1 - a_i)]$ $a_3 = Sigmoid(z_3)$ $Z_3 = W_3 A_2 + B_3$

Sigmoid

$$f(x) = \frac{1}{1+e^{-x}}$$

$$\frac{\partial L}{\partial w_3} = \frac{\partial L}{\partial a_3} \times \frac{\partial a_3}{\partial z_3} \times \frac{\partial z_3}{\partial w_3}$$

$$\frac{\partial L}{\partial a_3} = \frac{\partial}{\partial a_3}\left[-y \cdot log(a_3) - (1-y) \cdot log(1-a_3)\right]$$

$$= \frac{-y}{a_3} + \frac{(1-y)}{(1-a_3)}$$

$$= \frac{-y + \cancel{ya_3} + a_3 - \cancel{ya_3}}{a_3 \cdot (1-a_3)}$$

$$\frac{\partial L}{\partial a_3} = \frac{a_3 - y}{a_3 \cdot (1-a_3)}$$

$$\frac{\partial a_3}{\partial z_3} = \frac{\partial}{\partial z_3}\left(\frac{1}{1+e^{-z_3}}\right) = \frac{\partial}{\partial z_3}\left[(1+e^{-z_3})^{-1}\right]$$

$$= (-1) \cdot (1+e^{-z_3})^{-2} \cdot \frac{\partial}{\partial z_3}(1+e^{-z_3})$$

$$= \frac{e^{-z_3}}{(1+e^{-z_3})^2}$$

$$= a_3^2 \cdot \frac{(1-a_3)}{a_3}$$

$$\left[\because a_3 = \frac{1}{(1+e^{-z_3})} \text{ \& } e^{-z_3} = \frac{(1-a_3)}{a_3}\right]$$

$$\frac{\partial a_3}{\partial z_3} = a_3 \cdot (1-a_3)$$

$$\frac{\partial z_3}{\partial w_3} = \frac{\partial}{\partial w_3}\left[w_3 a_2 + B_3\right]$$

$$= a_2$$

$$\frac{\partial L}{\partial W_3} = \frac{\partial L}{\partial a_3} \times \frac{\partial a_3}{\partial z_3} \times \frac{\partial z_3}{\partial W_3}$$

$$= \frac{(a_3 - y)}{\cancel{a_3 \cdot (1-a_3)}} \cdot \cancel{a_3 \cdot (1-a_3)} \cdot a_2$$

$$= (a_3 - y) \cdot a_2$$ ← for 1 observation

$$\frac{\partial cost}{\partial W_3} = \frac{1}{m} (A_3 - Y) \cdot A_2^T$$ ← For all m observations

$$= \frac{1}{m} dZ_3 \cdot A_2^T$$

where $\frac{\partial cost}{\partial Z_3} = dZ_3 = A_3 - Y$

Similarly, $\frac{\partial L}{\partial B_3} = \frac{\partial L}{\partial a_3} \times \frac{\partial a_3}{\partial Z_3} \times \frac{\partial Z_3}{\partial B_3} = (a_3 - Y) \cdot (1)$

$$\frac{\partial cost}{\partial B_3} = \frac{1}{m} \mathrm{sum}(dZ_3, 1)$$

↑ Represent columnwise summation

Now, For $\partial cost / dW_2$, $d\,cost / dB_2$

$$\frac{\partial L}{\partial W_2} = \frac{\partial L}{\partial a_3} \times \frac{\partial a_3}{\partial z_3} \times \frac{\partial z_3}{\partial a_2} \times \frac{\partial a_2}{\partial z_2} \times \frac{\partial z_2}{\partial W_2}$$

$$= dz_3 \cdot w_3 \cdot f_2'(z_2) \cdot a_1$$

$$\frac{\partial cost}{\partial W_2} = \frac{1}{m}\left[W_3^T \cdot dZ_3 * f_2'(Z_2)\right] \cdot A_1^T$$

$$= \frac{1}{M}\, dZ_2 \cdot A_1^T$$

$$\left[\because \frac{dL}{da_3} \times \frac{da_3}{dz_3} = \frac{dL}{dz_3} = dz_3 ,\right.$$

$$\frac{\partial z_3}{\partial a_2} = \frac{\partial (W_3 \cdot a_2 + B_3)}{\partial a_2} = a_2 , \quad \frac{\partial a_2}{\partial z_2} = f_2'(z_2)$$

$$\left.\frac{\partial z_2}{\partial W_2} = \frac{\partial (W_2 \cdot a_1 + B_2)}{\partial W_2} = a_1 \right]$$

$$\left[\because dZ_2 = \frac{d\,cost}{dZ_2} = W_3^T \cdot dZ_3 * f_2'(Z_2)\right]$$

Similarly, $\frac{\partial cost}{\partial B_2} = \frac{1}{m} Sum(dZ_2, 1)$ [∵ column wise sum of dZ_2]

Similarly, for $\partial cost / dw_1$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial a_3} \times \frac{\partial a_3}{\partial z_3} \times \frac{\partial z_3}{\partial a_2} \times \frac{\partial a_2}{\partial z_2} \times \frac{\partial z_2}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial w_1}$$

$$= dz_2 \cdot w_2 \cdot f_1'(z_1) \cdot x$$

$$\frac{\partial cost}{\partial w_1} = \frac{1}{m}\left[w_2^T \cdot dz_2 * f_1'(z_1)\right] \cdot x^T$$

$$= \frac{1}{m} dz_1 \cdot x^T$$

Similarly,

$$\frac{\partial cost}{\partial B} = \frac{1}{m} \mathrm{Sum}(dz_1, 1)$$

Complete Back Propagation

$$dZ_3 = (A_3 - Y)$$

$$dW_3 = \frac{1}{m} . dZ_3 . A_2^T$$

$$dB_3 = \frac{1}{m} . sum(dZ_3, 1)$$

$$dZ_2 = W_3^T . dZ_3 * f_2^{'}(Z_2)$$

$$dW_2 = \frac{1}{m} . dZ_2 . A_2^T$$

$$dB_2 = \frac{1}{m} . sum(dZ_2, 1)$$

$$dZ_1 = W_2^T . dZ_2 * f_1^{'}(Z_1)$$

$$dW_1 = \frac{1}{m} . dZ_1 . A_1^T$$

$$dB_1 = \frac{1}{m} . sum(dZ_1, 1)$$

$W_3 = W_3 -$ alpha * dW_3

$B_3 = B_3 -$ alpha * dB_3

$W_2 = W_2 -$ alpha * dW_2

$B_2 = B_2 -$ alpha * dB_2

$W_1 = W_1 -$ alpha * dW_1

$B_1 = B_1 -$ alpha * dB_1