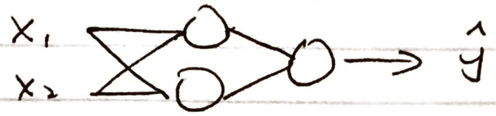
1 NN By Hand

BEIMING LIN



$b_1^{[1]}$ -30

$W_{11}^{[1]}$ 20

$W_{12}^{[1]}$ 20

$b^{[2]} = -10$

$b_2^{[1]}$ 10

$W_{21}^{[1]}$ -20

$W_{22}^{[1]}$ -20

$w_1^{[2]} = 20$

$w_2^{[2]} = 20$

$\sigma(a) = h(a) = \begin{cases} 1 & \text{if } a > 0 \\ 0 & \text{o/w} \end{cases}$

$x = (-1, -1)$.

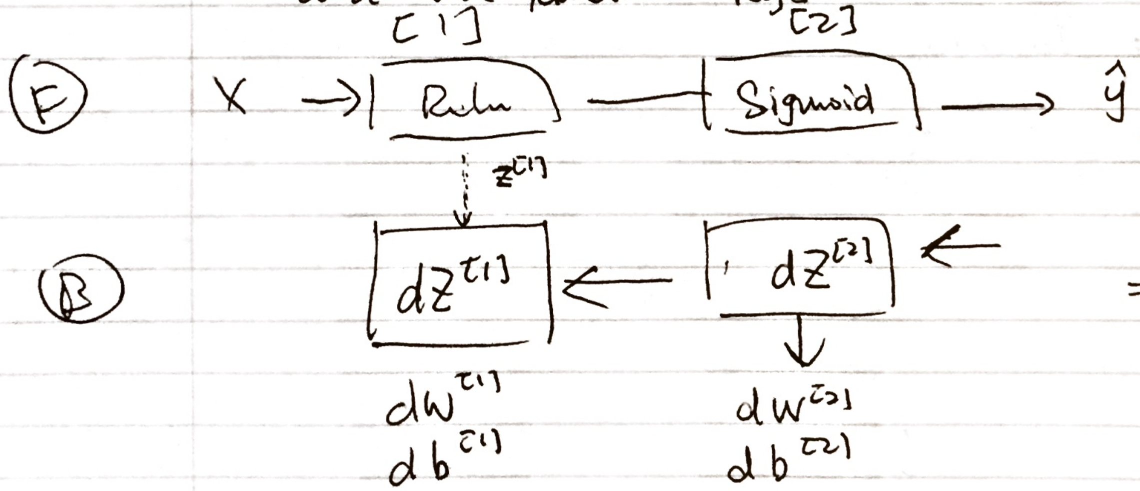
$Z_1^{[1]} = -30 - 20 - 20$ $\quad a_1^{[1]} = 0$

$Z_2^{[1]} = 10 + 20 + 20$ $\quad a_2^{[1]} = 1$

$Z^{[2]} = -10 + 0 + 20$ $\quad \underline{a^{[2]} = 1 = \hat{y}}$

2. Assume W, X, Z, A are all vectorized and use Andrew Ng's notation.

m data in total



$\mathcal{L}(\hat{y}, y) = \mathcal{L}(a, y)$

$= -(y \log(a) + (1-y)\log(1-a))$

$W^{[l]} := W^{[l]} - \alpha dW^{[l]}$

$b^{[l]} := b^{[l]} - \alpha db^{[l]}$

$$\mathcal{L}(a, y) = -\left(y \log(a) + (1+y)\log(1+a)\right)$$

FWD:

$$A^{[0]} := X$$
$$Z^{[1]} = W^{[1]} A^{[0]} + b^{[1]}$$
$$A^{[1]} = g(Z^{[1]})$$
$$Z^{[2]} = W^{[2]} A^{[1]} + b^{[2]}$$
$$A^{[2]} = \sigma(Z^{[2]})$$

$$g(z) = \begin{cases} z & \text{if } z > 0 \\ 0, & \text{o/w} \end{cases}$$

$$\sigma(z) = a = \frac{1}{1+e^{-z}}$$

BWD.

$$dz^{[2]} = \frac{\partial \mathcal{L}}{\partial z} = \frac{\partial \mathcal{L}}{\partial a} \cdot \frac{\partial a}{\partial z} = \left(-\frac{y}{a} + \frac{1-y}{1-a}\right)\left(a(1-a)\right)$$
$$= a - y$$

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

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

where m is the # of data.

elementwise product

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

where

$$g'(z) = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{o/w} \end{cases}$$

$$dW^{[1]} = \frac{1}{m} dZ^{[1]} A^{[0]T}$$
$$db^{[1]} = \frac{\sum_i dZ_i^{[1]}}{m}$$