# 2. Computation Graphs and Deep Learning.

## 2.1

$$\because h(w^Tx_i+b) = \frac{1}{1+e^{-(w^Tx+b)}}$$

$$\therefore \log(h(w^Tx_i+b)) = -\log(1+e^{-(w^Tx+b)})$$

$$\therefore \log(1-h(w^Tx_i+b)) = \log(e^{-(w^Tx+b)}) - \log(1+e^{-(w^Tx+b)})$$

$$= -(w^Tx+b) - \log(1+e^{-(w^Tx+b)})$$

$$\because L(w,b) = -\frac{1}{m}\sum_{i=1}^{M}[y_i\log(h(w^Tx_i+b)) + (1-y_i)\log(1-h(w^Tx_i+b))]$$

$$= -\frac{1}{m}\sum_{i=1}^{M}[-y_i\log(1+e^{-(w^Tx+b)}) + (1-y_i)(-(w^Tx+b)-\log(1+e^{-(w^Tx+b)}))]$$

$$\therefore L(w,b) = -\frac{1}{m}\sum_{i=1}^{M}[-y_i\log(1+e^{-(w^Tx+b)}) -(w^Tx+b) + y_i(w^Tx+b) - \log(1+e^{-(w^Tx+b)}) + y_i\cdot\log(1+e^{-(w^Tx+b)})]$$

$$= -\frac{1}{m}\sum_{i=1}^{M}[-(w^Tx+b)+y_i(w^Tx+b) - \log(1+e^{-(w^Tx+b)})]$$

$$\because -(w^Tx+b) - \log(1+e^{-(w^Tx+b)}) = -\log e^{w^Tx+b} - \log(1+e^{-(w^Tx+b)})$$

$$= -\log(1+e^{w^Tx+b})$$

$$\therefore L(w,b) = -\frac{1}{m}\sum_{i=1}^{M}[y_i(w^Tx+b) - \log(1+e^{w^Tx+b})]$$

$$\cancel{\therefore \frac{\partial L}{\partial w_1} = -\frac{1}{m}\sum_{i=1}^{M}[y_i x_i]}$$

$$\because \frac{\partial}{\partial w} \log(1+e^{w^Tx+b}) = \frac{x \cdot e^{w^Tx+b}}{1+e^{w^Tx+b}} = \frac{e^{-(w^Tx+b)} \cdot x \cdot e^{w^Tx+b}}{e^{-(w^Tx+b)}(1+e^{w^Tx+b})}$$

$$= \frac{x}{1+e^{-(w^Tx+b)}} = x_i \cdot h(w^Tx_i+b)$$

$$\boxed{\therefore \frac{\partial L}{\partial w_1} = -\frac{1}{m}\sum_{i=1}^{m}[y_i x_1^i - x_1^i \cdot h(w^Tx^i+b)]}$$

$$\boxed{\therefore \frac{\partial L}{\partial w_2} = -\frac{1}{m}\sum_{i=1}^{m}[y_i x_2^i - x_2^i \cdot h(w^Tx^i+b)]}$$

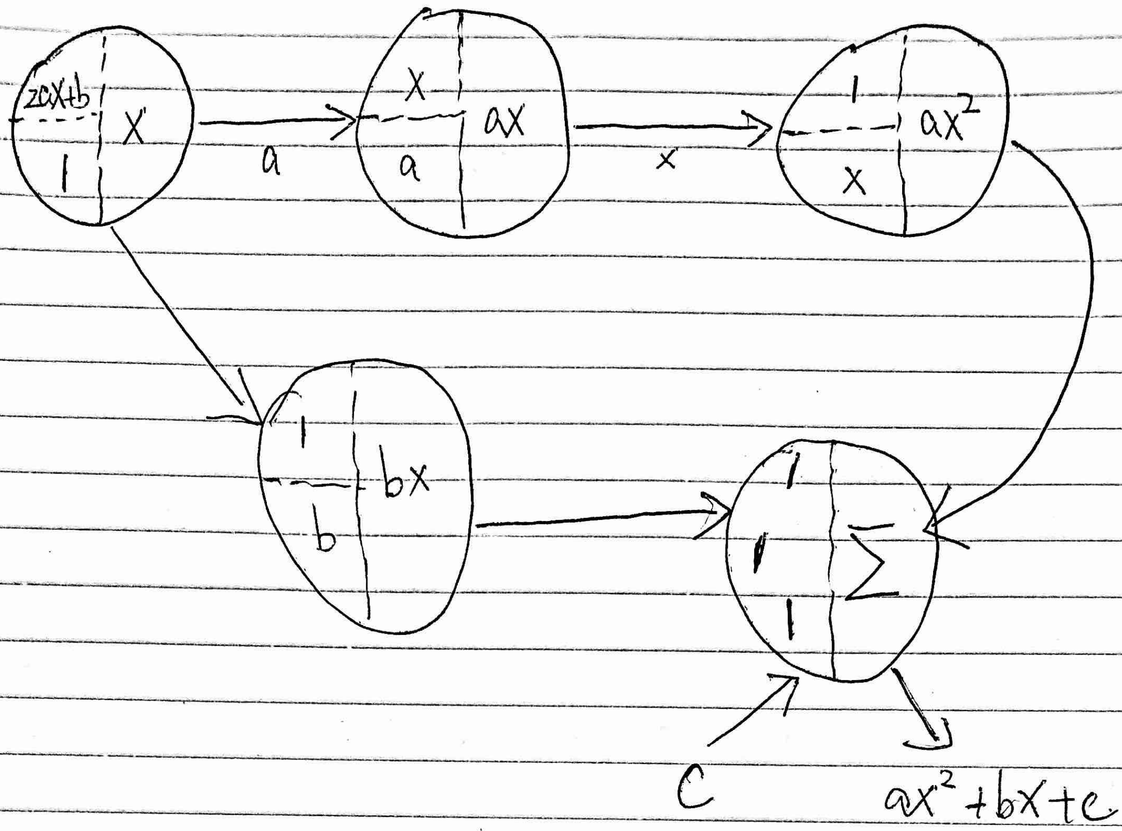
$$\because \frac{\partial}{\partial b} \log(1+e^{w^Tx+b}) = \frac{e^{w^Tx+b}}{1+e^{w^Tx+b}}$$

$$= \frac{1}{1+e^{-(w^Tx+b)}}$$

$$= h(w^Tx+b).$$

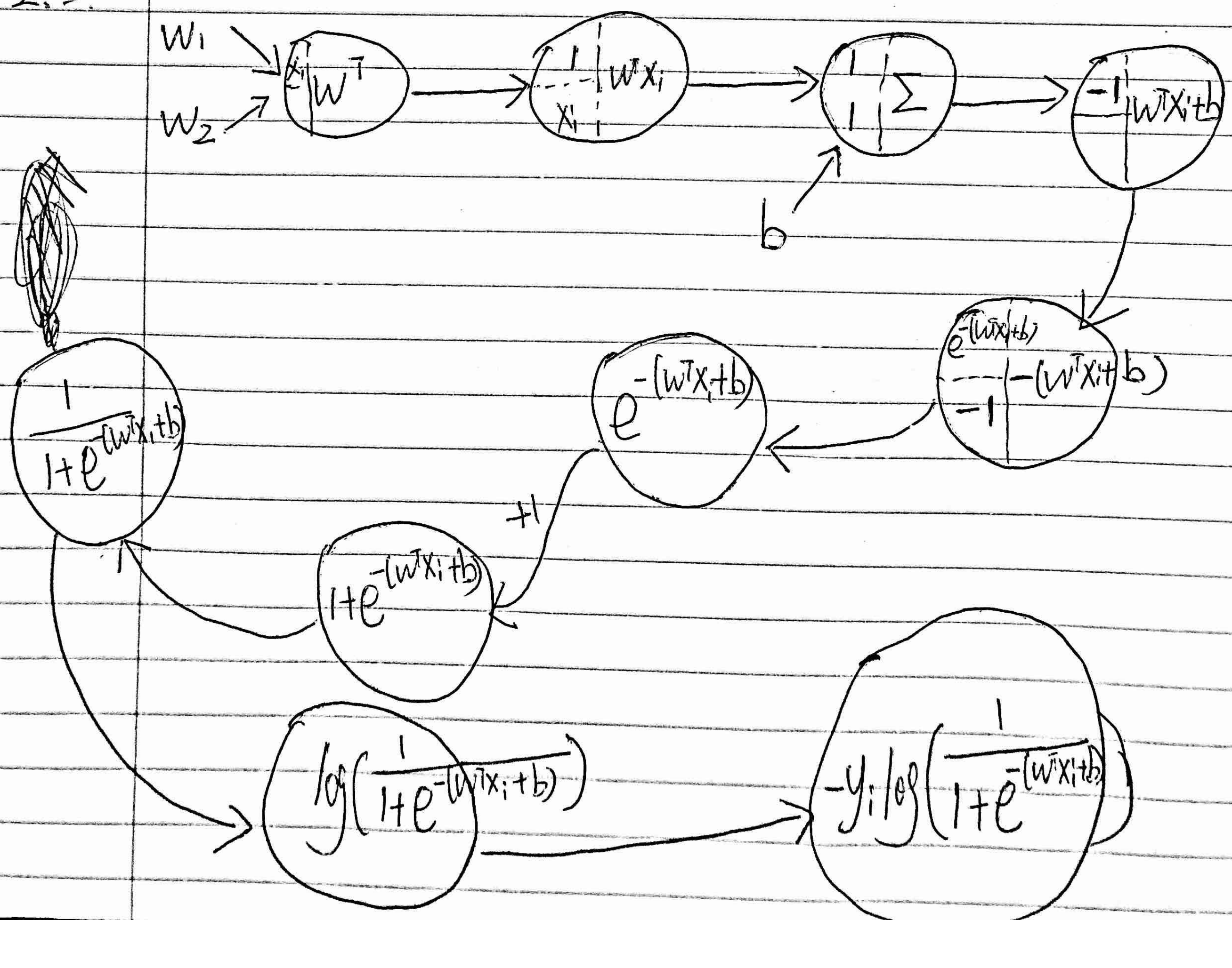
$$\boxed{\therefore \frac{\partial L}{\partial b} = -\frac{1}{m}\sum_{i=1}^{m}[y_i - h(w^Tx_i+b)]}$$

2.2



X'
1
a
X
a
aX
x
1
X
$aX^2$
1
b
bX
$\Sigma$
c
$aX^2+bX+c$


2.3



$w_1$
$w_2$
$x_i$
$W^T$
1
$x_i$
$W^Tx_i$
1
1
$\Sigma$
b
-1
$W^Tx_i+b$
$e^{-(W^Tx_i+b)}$
-1
$-(W^Tx_i+b)$
$e^{-(W^Tx_i+b)}$
+1
$1+e^{-(W^Tx_i+b)}$
1
$\frac{1}{1+e^{-(W^Tx_i+b)}}$
$\log\left(\frac{1}{1+e^{-(W^Tx_i+b)}}\right)$
$-y_i\log\left(\frac{1}{1+e^{-(W^Tx_i+b)}}\right)$

3. Extra Credit:

3.1

$$\frac{\partial L}{\partial w_1} = 1 \cdot 5 - 5 \cdot h\left(\cancel{w^T \cdot} \begin{bmatrix} 1.1 \\ -6.0 \end{bmatrix} \begin{bmatrix} 5 & 10 \end{bmatrix} + 2\right)$$

$$= 5 - 5 \cdot h(-52.5)$$

$$= 5 - 5 \cdot \frac{1}{1 + e^{52.5}}$$

$$\approx 5$$

$$\frac{\partial L}{\partial w_2} = 1 \cdot 10 - 10 \cdot h\left(\begin{bmatrix} 1.1 \\ -6.0 \end{bmatrix} \begin{bmatrix} 5 & 10 \end{bmatrix} + 2\right)$$

$$= \cancel{5}\, 10 - 10 \cdot h(-52.5)$$

$$= 10 - 10 \cdot \frac{1}{1 + e^{52.5}}$$

$$\approx 10.$$

$$\frac{\partial L}{\partial b} = 1 - h\left(\begin{bmatrix} 1.1 \\ -6.0 \end{bmatrix} \begin{bmatrix} 5 & 10 \end{bmatrix} + 2\right)$$

$$= 1 - \frac{1}{1 + e^{52.5}}$$

$$= 1$$