EE412 Foundation of Big Data Analytics, Fall 2019
HW4

Name: Cao Viet Hai Nam

Student ID: 20200817

Discussion Group (People with whom you discussed ideas used in your answers):

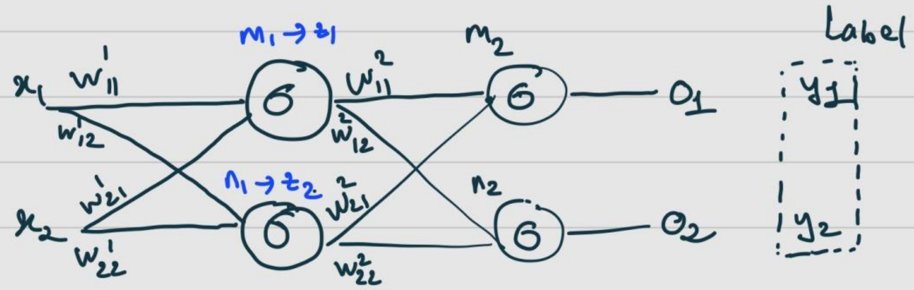
On-line or hardcopy documents used as part of your answers:

Answer to Problem 1

20200817 Cao Viet Hai Nam

1.

a)



let $\begin{cases} m_1 = x_1 w^1_{11} + x_2 w^1_{21} \\ n_1 = x_1 w^1_{12} + x_2 w^1_{22}. \end{cases}$

and $\begin{cases} z_1 = \sigma(m_1) \\ z_2 = \sigma(n_1) \end{cases}$

Note that: z_1, z_2, m_1, m_2, n_1, n_2 are all expressed by given parameters

$\begin{cases} m_2 = z_1 w^2_{11} + z_2 w^2_{21} \\ n_2 = z_1 w^2_{12} + z_2 w^2_{22}. \end{cases}$

and $\begin{cases} o_1 = \sigma(m_2) \\ o_2 = \sigma(n_2) \end{cases}$

$$L = \frac{1}{2}\left[(o_1 - y_1)^2 + (o_2 - y_2)^2\right]$$

$$\Rightarrow \frac{\partial L}{\partial o_1} = o_1 - y_1 \quad ; \quad \frac{\partial L}{\partial o_2} = o_2 - y_2$$

$$\frac{\partial L}{\partial w_{11}^2} = \frac{\partial L}{\partial o_1} \cdot \frac{\partial o_1}{\partial m_2} \cdot \frac{\partial m_2}{\partial w_{11}^2}$$

$$= (o_1 - y_1) \cdot o_1(1-o_1) \cdot z_1$$

$$\left(\text{Since } \frac{\partial o_1}{\partial m_2} = \frac{\partial\left(\frac{e^{m_2}}{e^{m_2}+1}\right)}{\partial m_2} = \frac{e^{m_2}}{(e^{m_2}+1)^2} = o_1(1-o_1) \right)$$

$$\frac{\partial L}{\partial w_{21}^2} = \frac{\partial L}{\partial o_1} \cdot \frac{\partial o_1}{\partial m_2} \cdot \frac{\partial m_2}{\partial w_{21}^2}$$

$$= (o_1 - y_1)\, o_1 (1-o_1)\, z_2$$

$$\frac{\partial L}{\partial w_{12}^2} = \frac{\partial L}{\partial o_2} \cdot \frac{\partial o_2}{\partial n_2} \cdot \frac{\partial n_2}{w_{12}^2}$$

$$= (o_2 - y_2)\, o_2 (1-o_2)\, z_1$$

$$\frac{\partial L}{\partial w_{22}^2} = \frac{\partial L}{\partial o_2} \cdot \frac{\partial o_2}{\partial n_2} \cdot \frac{\partial n_2}{\partial w_{22}^2}$$

$$= (o_2 - y_2)\, o_2 (1-o_2)\, z_2 \,.$$

$$\frac{\partial L}{\partial w_{11}^1} = \frac{\partial L}{\partial o_1} \cdot \frac{\partial o_1}{\partial m_2} \cdot \frac{\partial m_2}{\partial w_{11}^1} + \frac{\partial L}{\partial o_2} \cdot \frac{\partial o_2}{\partial n_2} \cdot \frac{\partial n_2}{\partial w_{11}^1} \quad (*)$$

we have: $\frac{\partial L}{\partial o_1} = o_1 - y_1$; $\frac{\partial L}{\partial o_2} = o_2 - y_2$

$$\frac{\partial o_1}{\partial m_2} = o_1(1-o_1) \; ; \; \frac{\partial o_2}{\partial n_2} = o_2(1-o_2)$$

$$\frac{\partial m_2}{\partial w_{11}^1} = \frac{\partial m_2}{\partial z_1} \cdot \frac{\partial z_1}{\partial m_1} \cdot \frac{\partial m_1}{\partial w_{11}^1}$$

$$= w_{11}^2 \cdot z_1(1-z_1) \cdot x_1$$

$$\frac{\partial n_2}{\partial w_{11}^1} = \frac{\partial n_2}{\partial z_1} \cdot \frac{\partial z_1}{\partial m_1} \cdot \frac{\partial m_1}{\partial w_{11}^1}$$

$$= w_{12}^2 \; z_1(1-z_1) x_1$$

Plug in (*) $\Rightarrow$

$$\frac{\partial L}{\partial w_{11}^1} = (o_1 - y_1) o_1(1-o_1) \cdot w_{11}^2 \cdot z_1(1-z_1) \cdot x_1$$
$$+ (o_2 - y_2) o_2(1-o_2) \cdot w_{12}^2 \; z_1(1-z_1) x_1$$

Similarly;

$$\frac{\partial L}{\partial w_{21}^1} = (o_1 - y_1) o_1(1-o_1) \; w_{11}^2 \; z_1(1-z_1) x_2$$
$$+ (o_2 - y_2) o_2(1-o_2) \; w_{12}^2 \; z_1(1-z_1) x_2$$

$$\frac{\partial L}{\partial w_{12}^{1}} = \frac{\partial L}{o_1} \cdot \frac{\partial o_1}{m_2} \cdot \frac{\partial m_2}{\partial w_{12}^{1}} + \frac{\partial L}{o_2} \cdot \frac{\partial o_2}{\partial n_2} \cdot \frac{\partial n_2}{\partial w_{12}^{1}}$$

$$= (o_1 - y_1)\, o_1 (1 - o_1)\, w_{21}^{2}\, z_2 (1 - z_2)\, x_1$$
$$+ (o_2 - y_2)\, o_2 (1 - o_2)\, w_{22}^{2}\, z_2 (1 - z_2)\, x_1$$

$$\frac{\partial L}{\partial w_{22}^{1}} = (o_1 - y_1)\, o_1 (1 - o_1)\, w_{21}^{2}\, z_2 (1 - z_2)\, x_2$$
$$+ (o_2 - y_2)\, o_2 (1 - o_2)\, w_{22}^{2}\, z_2 (1 - z_2)\, x_2$$

Answer to Problem 2

② Mining Data Stream

4.4.1 let $h_1(x) = (2x+1) \bmod 32$, $h_2(x) = (3x+7) \bmod 32$

$h_3(x) = 4x \pmod{32}$

Stream	$h_1(x)$	$h_2(x)$	$h_3(x)$
3	00111	10000	01100
1	00011	01010	00100
4	01001	10011	10000
1	00011	01010	00100
5	01011	10110	10100
9	10011	00010	00100
2	00101	01101	01000
6	01101	11001	11000
5	01011	10110	10100
Max tail length (R)	0	4	4

⇒ Estimated number of distinct elements for each hash function is :

$h_1 :\quad 2^{R_1} = 2^0 = 1$

$h_2 :\quad 2^{R_2} = 2^4 = 16$

$h_3 :\quad 2^{R_3} = 2^4 = 16$

For hash function $(ax+b) \bmod 2^k$, when a is even and b is odd, the hash function always results in odd number → binary value ends by 1's making the tail length always equal to 0

when a and b are both even, it always result in even number → higher chance of getting 2 different elements having a same representation following the hash function (hash function is supposed to generate different values)

Thus it is not advisable to choose a as an even number

4.5.3

i	X_i. element	X_i. value
1	3	2
2	1	3
3	4	2
4	1	2

5	3	1
6	4	1
7	2	2
8	1	1
9	2	1

Estimated surprise number

$$= \frac{1}{9} \sum_{i=1}^{9} 9(2X_i.\text{value} - 1)$$

$$= \big[(2.2-1) + (2.3-1) + (2.2-1) + (2.2-1) + (2.1-1) + (2.1-1) + (2.2-1) + (2.1-1) + (2.1-1) \big]$$

$$= 21$$