EE412 Foundation of Big Data Analytics, Fall 2022
HW4

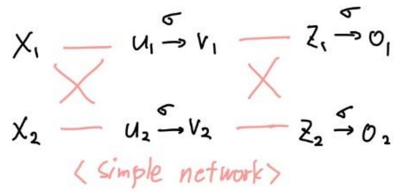
Name: TaewookHam
Student ID: 20180716

Answer to Problem 1

1-(a)

"Chainrule" $$O = \sigma(\overbrace{W^2 \underbrace{\sigma(\underbrace{W^1 x}_{u})}_{v}}^{z})$$

$$O = \sigma(z)$$



$$MSE = L(o) = \frac{1}{2}\sum_i^n (o_i - y_i)^2$$

$$o_1 = \sigma(z_1)$$
$$o_2 = \sigma(z_2)$$
$$z_1 = w_{11}^2 v_1 + w_{21}^2 v_2$$
$$z_2 = w_{12}^2 v_1 + w_{22}^2 v_2$$
$$Z = \begin{bmatrix} w_{11}^2 & w_{21}^2 \\ w_{12}^2 & w_{22}^2 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} w_1^T \\ w_2^T \end{bmatrix} v$$
$$= [w_1^2 \ w_2^2]^T$$

$$\frac{\partial L_1}{\partial o_1} = (o_1 - y_1) \qquad \frac{\partial L_2}{\partial o_2} = (o_2 - y_2)$$
$$\frac{\partial o_1}{\partial z_1} = o_1(1-o_1) \qquad \frac{\partial o_2}{\partial z_2} = o_2(1-o_2)$$
$$\frac{\partial z_1}{\partial w_{11}^2} = v_1 \qquad \frac{\partial z_1}{\partial w_{12}^2} = v_1$$
$$\frac{\partial z_1}{\partial w_{21}^2} = v_2 \qquad \frac{\partial z_1}{\partial w_{22}^2} = v_2$$

$$\frac{\partial L_1}{\partial w_{11}^2} = \frac{\partial z_1}{\partial w_{11}^2}\cdot\frac{\partial o_1}{\partial z_1}\cdot\frac{\partial L_1}{\partial o_1} = \sigma(w_{11}^1 x_1 + w_{21}^1 x_2)\cdot o_1(1-o_1)(o_1-y_1)$$

$$\frac{\partial L_1}{\partial w_{21}^2} = \frac{\partial z_1}{\partial w_{21}^2}\cdot\frac{\partial o_1}{\partial z_1}\cdot\frac{\partial L_1}{\partial o_1} = \sigma(w_{12}^1 x_1 + w_{22}^1 x_2)\cdot o_1(1-o_1)(o_1-y_1)$$

$$\frac{\partial L_2}{\partial w_{12}^2} = \frac{\partial z_2}{\partial w_{12}^2}\cdot\frac{\partial o_2}{\partial z_2}\cdot\frac{\partial L_2}{\partial o_2} = \sigma(w_{11}^1 x_1 + w_{21}^1 x_2)\cdot o_2(1-o_2)(o_2-y_2)$$

$$\frac{\partial L_2}{\partial w_{22}^2} = \frac{\partial z_2}{\partial w_{22}^2}\cdot\frac{\partial o_2}{\partial z_2}\cdot\frac{\partial L_2}{\partial o_2} = \sigma(w_{12}^1 x_1 + w_{22}^1 x_2)\cdot o_2(1-o_2)(o_2-y_2)$$

Ans

where
$$\left(\begin{aligned} v_1 &= \sigma(u_1) \\ v_2 &= \sigma(u_2) \\ u_1 &= w_{11}^1 x_1 + w_{21}^1 x_2 = w_1^T x \\ u_2 &= w_{12}^1 x_1 + w_{22}^1 x_2 = w_2^T x \end{aligned}\right)$$

$$\frac{\partial L_1}{\partial v_1} = \frac{\partial z_1}{\partial v_1} \cdot \frac{\partial o_1}{\partial z_1} \cdot \frac{\partial L_1}{\partial o_1} = w_{11}^2\, o_1(1-o_1)(o_1-y_1)$$

$$\frac{\partial L_2}{\partial v_1} = \frac{\partial z_2}{\partial v_1} \cdot \frac{\partial o_2}{\partial z_2} \cdot \frac{\partial L_2}{\partial o_2} = w_{12}^2 \cdot o_2(1-o_2)\cdot(o_2-y_2)$$

$$\frac{\partial L_1}{\partial v_2} = \frac{\partial z_1}{\partial v_2} \cdot \frac{\partial o_1}{\partial z} \cdot \frac{\partial L_1}{\partial o_1} = w_{21}^2 \cdot o_1(1-o_1)\cdot(o_1-y_1)$$

$$\frac{\partial L_2}{\partial v_2} = \frac{\partial z_2}{\partial v_2} \cdot \frac{\partial o_2}{\partial z_2} \cdot \frac{\partial L_2}{\partial o_2} = w_{22}^2\, o_2(1-o_2)(o_2-y_2)$$

$$V_1 = \sigma(u_1) \qquad \frac{\partial V_1}{\partial u_1} = \sigma(u_1)(1-\sigma(u_1)) = \sigma(w_1^T x)(1-\sigma(w_1^T x))$$

$$V_2 = \sigma(u_2)$$

$$u_1 = w_{11}' x_1 + w_{21}' x_2 = w_1^T x$$

$$u_2 = w_{12}' x_1 + w_{22}' x_2 = w_2^T x \qquad \frac{\partial V_2}{\partial u_2} = \sigma(u_2)(1-\sigma(u_2)) = \sigma(w_2^T x)(1-\sigma(w_2^T x))$$

$$\to W'^T = \begin{bmatrix} w_1^T \\ w_2^T \end{bmatrix} = \begin{bmatrix} w_{11}' & w_{21}' \\ w_{12}' & w_{22}' \end{bmatrix}$$

$$\sigma(u)(1-\sigma(u)) \begin{bmatrix} x_1 & 0 \\ 0 & x_2 \end{bmatrix} \begin{bmatrix} w_{11}^2 & 0 \\ 0 & w_{12}^2 \end{bmatrix} \begin{bmatrix} o_1(1-o_1) & 0 \\ 0 & o_2(1-o_2) \end{bmatrix} \begin{bmatrix} o_1-y_1 \\ o_2-y_2 \end{bmatrix}$$

$$\frac{\partial L}{\partial w_{11}'} = \left(\frac{\partial L_1}{\partial v_1} + \frac{\partial L_2}{\partial v_1}\right) \frac{\partial v_1}{\partial u_1} \cdot \frac{\partial u_1}{\partial w_{11}'} = \left(w_{11}^2\, o_1(1-o_1)(o_1-y_1) + w_{12}^2 \cdot o_2(1-o_2)\cdot(o_2-y_2)\right) \cdot \sigma(w_{11}'x_1 + w_{21}'x_2)\cdot(1-\sigma(w_{11}'x_1 + w_{21}'x_2))x_1$$

$$\frac{\partial L}{\partial w_{21}'} = \left(\frac{\partial L_1}{\partial v_1} + \frac{\partial L_2}{\partial v_1}\right) \frac{\partial v_1}{\partial u_1} \cdot \frac{\partial u_1}{\partial w_{21}'} = \left(w_{11}^2\, o_1(1-o_1)(o_1-y_1) + w_{12}^2 \cdot o_2(1-o_2)\cdot(o_2-y_2)\right) \cdot \sigma(w_{11}'x_1 + w_{21}'x_2)\cdot(1-\sigma(w_{11}'x_1 + w_{21}'x_2))x_2$$

$$\frac{\partial L}{\partial w_{12}'} = \left(\frac{\partial L_1}{\partial v_2} + \frac{\partial L_2}{\partial v_2}\right) \frac{\partial v_2}{\partial u_2} \cdot \frac{\partial u_2}{\partial w_{12}'} = \left(w_{21}^2 \cdot o_1(1-o_1)\cdot(o_1-y_1) + w_{22}^2\, o_2(1-o_2)(o_2-y_2)\right) \cdot \sigma(w_{12}'x_1 + w_{22}'x_2)\cdot(1-\sigma(w_{12}'x_1 + w_{22}'x_2))x_1$$

$$\frac{\partial L}{\partial w_{22}'} = \left(\frac{\partial L_1}{\partial v_2} + \frac{\partial L_2}{\partial v_2}\right) \frac{\partial v_2}{\partial u_2} \cdot \frac{\partial u_1}{\partial w_{11}'} = \left(w_{21}^2 \cdot o_1(1-o_1)\cdot(o_1-y_1) + w_{22}^2\, o_2(1-o_2)(o_2-y_2)\right) \cdot \sigma(w_{12}'x_1 + w_{22}'x_2)\cdot(1-\sigma(w_{12}'x_1 + w_{22}'x_2))x_2$$

Ans

(b)

```
20180716@eelab5:~/20180716_hw4$ python hw4_1_p3.py training.csv testing.csv
0.949
0.78
5000
0.01
```

Answer to Problem 2

(a)

Exercise 4.4.1 and 4.4.2

Stream 3, 1, 4, 1, 5, 9, 2, 6, 5

$h(x) = (ax+b) \bmod 32$

— : Tail length

	$2x+1 \bmod 32$	$3x+7 \bmod 32$	$4x \bmod 32$
3	7 = 00111 ⇒ 0	16 = 10000 ⇒ 4 V	12 = 01100 ⇒ 2
1	3 = 00011 ⇒ 0	10 = 01010 ⇒ 1	4 = 00100 ⇒ 2
4	9 = 01001 ⇒ 0	19 = 10011 ⇒ 0	16 = 10000 ⇒ 4 V
1	3 = 00011 ⇒ 0	10 = 01010 ⇒ 1	4 = 00100 ⇒ 2
5	11 = 01011 ⇒ 0	22 = 10110 ⇒ 1	20 = 10100 ⇒ 2
9	19 = 10011 ⇒ 0	2 = 00010 ⇒ 1	4 = 00100 ⇒ 2
2	5 = 00101 ⇒ 0	13 = 01101 ⇒ 0	8 = 01000 ⇒ 3
6	13 = 00111 ⇒ 0	25 = 11001 ⇒ 0	24 = 11000 ⇒ 3
5	11 = 01011 ⇒ 0	22 = 10110 ⇒ 1	20 = 10100 ⇒ 2
	estimate : $2^0 = 1$	estimate : $2^4 = 16$	estimate : $2^4 = 16$

Discussion

- $ax+b \bmod 2^5$ 꼴의 함수 사용시 실제 length 9와는 다소 차이가 있는 2, 16 이라는 estimate을 얻었다.

- 이러한 현상을 해결하기 위한 방법은 여러가지가 있다.

  1. 더 많은 $ax+b \bmod 2^k$ 꼴의 hash func을 사용한 후 몇몇개끼리 Grouping 한다.

     그 그룹의 estimate Average를 각각 구한 뒤 여러 그룹의 Avg 값의 median 값을 취한다면 오차를 줄일 수 있을 것이다.

  2. a, b, 2^k 에서의 k 값을 알맞게 조절해 $2^3=8$ 이 나올 확률을 높일 수는 있다.

Exercise 4.5.3

3 1 4 1 3 4 2 1 2

$\xrightarrow{\text{stream}}$

i	1	2	3	4	5	6	7	8	9
X_i.element	3	1	4	1	3	4	2	1	2
X_i.value	2	3	2	2	1	1	2	1	1

(b) when k’s are 1, 10, 100
0.5
2.0
23.0