E412 Foundation of Big Data Analytics, Fall 2021

HW4

20170616 정희진

1. Neural Nets and Deep Learning

(a) Compute the gradients using the chain rule

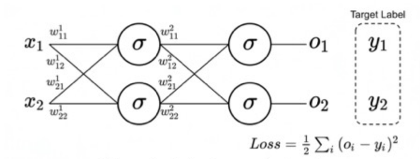


Figure 1: A simple neural network

Forward

$$\begin{bmatrix} w_{11}^1 & w_{12}^1 \\ w_{21}^1 & w_{22}^1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \quad \sigma\left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}\right) = \begin{bmatrix} x_3 \\ x_4 \end{bmatrix},$$

$$\begin{bmatrix} w_{11}^2 & w_{12}^2 \\ w_{21}^2 & w_{22}^2 \end{bmatrix}\begin{bmatrix} x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} u_3 \\ u_4 \end{bmatrix}, \quad \sigma\left(\begin{bmatrix} u_3 \\ u_4 \end{bmatrix}\right) = \begin{bmatrix} o_1 \\ o_2 \end{bmatrix},$$

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

Backward

$$\frac{\partial L}{\partial o_1} = o_1 - y_1$$

$$\frac{\partial L}{\partial o_2} = o_2 - y_2$$

$$\frac{\partial L}{\partial u_3} = \frac{\partial L}{\partial o_1}\cdot\frac{\partial o_1}{\partial u_3} = o_1(1-o_1)\frac{\partial L}{\partial o_1} = o_1(1-o_1)(o_1-y_1)$$

$$\frac{\partial L}{\partial u_4} = \frac{\partial L}{\partial o_2}\cdot\frac{\partial o_2}{\partial u_4} = o_2(1-o_2)\frac{\partial L}{\partial o_2} = o_2(1-o_2)(o_2-y_2)$$

$$\frac{\partial L}{\partial w_{11}^2} = \frac{\partial L}{\partial u_3}\cdot\frac{\partial u_3}{\partial w_{11}^2} = x_3\frac{\partial L}{\partial u_3} = x_3 o_1(1-o_1)(o_1-y_1)$$

$$\frac{\partial L}{\partial w_{12}^2} = \frac{\partial L}{\partial u_3}\cdot\frac{\partial u_3}{\partial w_{12}^2} = x_4\frac{\partial L}{\partial u_3} = x_4 o_1(1-o_1)(o_1-y_1)$$

$$\frac{\partial L}{\partial w_{21}^2} = \frac{\partial L}{\partial u_4}\cdot\frac{\partial u_4}{\partial w_{21}^2} = x_3\frac{\partial L}{\partial u_4} = x_3 o_2(1-o_2)(o_2-y_2)$$

$$\frac{\partial L}{\partial w_{22}^2} = \frac{\partial L}{\partial u_4}\cdot\frac{\partial u_4}{\partial w_{22}^2} = x_4\frac{\partial L}{\partial u_4} = x_4 o_2(1-o_2)(o_2-y_2)$$

$$\frac{\partial L}{\partial x_3} = \frac{\partial L}{\partial u_3}\cdot\frac{\partial u_3}{\partial x_3} + \frac{\partial L}{\partial u_4}\cdot\frac{\partial u_4}{\partial x_3} = w_{11}^2\frac{\partial L}{\partial u_3} + w_{21}^2\frac{\partial L}{\partial u_4} = w_{11}^2 o_1(1-o_1)(o_1-y_1) + w_{21}^2 o_2(1-o_2)(o_2-y_2)$$

$$\frac{\partial L}{\partial x_4} = \frac{\partial L}{\partial u_3}\cdot\frac{\partial u_3}{\partial x_4} + \frac{\partial L}{\partial u_4}\cdot\frac{\partial u_4}{\partial x_4} = w_{12}^2\frac{\partial L}{\partial u_3} + w_{22}^2\frac{\partial L}{\partial u_4} = w_{12}^2 o_1(1-o_1)(o_1-y_1) + w_{22}^2 o_2(1-o_2)(o_2-y_2)$$

$$\frac{\partial L}{\partial u_1} = \frac{\partial L}{\partial x_3}\cdot\frac{\partial x_3}{\partial u_1} = x_3(1-x_3)\frac{\partial L}{\partial x_3} = x_3(1-x_3)\{w_{11}^2 o_1(1-o_1)(o_1-y_1) + w_{21}^2 o_2(1-o_2)(o_2-y_2)\}$$

$$\frac{\partial L}{\partial u_2} = \frac{\partial L}{\partial x_4}\cdot\frac{\partial x_4}{\partial u_2} = x_4(1-x_4)\frac{\partial L}{\partial x_4} = x_4(1-x_4)\{w_{12}^2 o_1(1-o_1)(o_1-y_1) + w_{22}^2 o_2(1-o_2)(o_2-y_2)\}$$

$$\frac{\partial L}{\partial w_{11}^1} = \frac{\partial L}{\partial u_1}\cdot\frac{\partial u_1}{\partial w_{11}^1} = x_1\frac{\partial L}{\partial u_1} = x_1 x_3(1-x_3)\{w_{11}^2 o_1(1-o_1)(o_1-y_1) + w_{21}^2 o_2(1-o_2)(o_2-y_2)\}$$

$$\frac{\partial L}{\partial w_{12}^1} = \frac{\partial L}{\partial u_1}\cdot\frac{\partial u_1}{\partial w_{12}^1} = x_2\frac{\partial L}{\partial u_1} = x_2 x_3(1-x_3)\{w_{11}^2 o_1(1-o_1)(o_1-y_1) + w_{21}^2 o_2(1-o_2)(o_2-y_2)\}$$

$$\frac{\partial L}{\partial w_{21}^1} = \frac{\partial L}{\partial u_2}\cdot\frac{\partial u_2}{\partial w_{21}^1} = x_1\frac{\partial L}{\partial u_2} = x_1 x_4(1-x_4)\{w_{12}^2 o_1(1-o_1)(o_1-y_1) + w_{22}^2 o_2(1-o_2)(o_2-y_2)\}$$

$$\frac{\partial L}{\partial w_{22}^1} = \frac{\partial L}{\partial u_2}\cdot\frac{\partial u_2}{\partial w_{22}^1} = x_2\frac{\partial L}{\partial u_2} = x_2 x_4(1-x_4)\{w_{12}^2 o_1(1-o_1)(o_1-y_1) + w_{22}^2 o_2(1-o_2)(o_2-y_2)\}$$

Answer

$$\frac{\partial L}{\partial w_{11}^1} = x_1 x_3(1-x_3)\{w_{11}^2 o_1(1-o_1)(o_1-y_1) + w_{21}^2 o_2(1-o_2)(o_2-y_2)\}, \quad \frac{\partial L}{\partial w_{11}^2} = x_3 o_1(1-o_1)(o_1-y_1)$$

$$\frac{\partial L}{\partial w_{12}^1} = x_2 x_3(1-x_3)\{w_{11}^2 o_1(1-o_1)(o_1-y_1) + w_{21}^2 o_2(1-o_2)(o_2-y_2)\}, \quad \frac{\partial L}{\partial w_{12}^2} = x_4 o_1(1-o_1)(o_1-y_1)$$

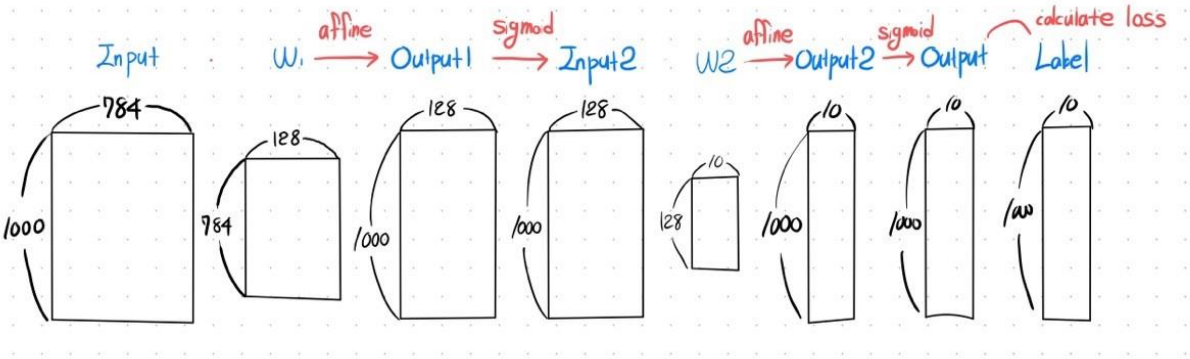
$$\frac{\partial L}{\partial w_{21}^1} = x_1 x_4(1-x_4)\{w_{12}^2 o_1(1-o_1)(o_1-y_1) + w_{22}^2 o_2(1-o_2)(o_2-y_2)\}, \quad \frac{\partial L}{\partial w_{21}^2} = x_3 o_2(1-o_2)(o_2-y_2)$$

$$\frac{\partial L}{\partial w_{22}^1} = x_2 x_4(1-x_4)\{w_{12}^2 o_1(1-o_1)(o_1-y_1) + w_{22}^2 o_2(1-o_2)(o_2-y_2)\}, \quad \frac{\partial L}{\partial w_{22}^2} = x_4 o_2(1-o_2)(o_2-y_2)$$

(이때 $x_3 = \sigma(w_{11}^1 x_1 + w_{12}^1 x_2)$
$x_4 = \sigma(w_{21}^1 x_1 + w_{22}^1 x_2)$)

(b) Implement a fully-connected network to distinguish digits using Python.

The structure of FCN (without bias) is below figure. However, in hw4_1_p2.py, bias is added to that figure.



Result is below figure.

```
0.965
0.782
10000
1.0
```

2. Mining Data Streams

(a) Solve the following problems, which are based on the exercises in the Mining of Massive Datasets 3rd edition (MMDS) textbook.

Exercise 4.4.1 and 4.4.2

(a)

x	h(x)	h(x)	tail length
3	7	00111	0
1	3	00011	0
4	9	01001	0
1	3	00011	0
5	11	01011	0
9	19	10011	0
2	5	00101	0
6	13	01101	0
5	11	01011	0

Estimate: $2^0 = 1$

(b)

x	h(x)	h(x)	tail length
3	16	10000	4
1	10	01010	1
4	19	10011	0
1	10	01010	1
5	22	10110	1
9	2	00010	1
2	13	01101	0
6	25	11001	0
5	22	10110	1

Estimate: $2^4 = 16$

(c)

x	h(x)	h(x)	tail length
3	12	01100	2
1	4	00100	2
4	16	10000	4
1	4	00100	2
5	20	10100	2
9	2	00010	1
2	8	01000	3
6	24	11000	3
5	20	10100	2

Estimate: $2^4 = 16$

(a)에서처럼 2x는 무조건 짝수를 만들어내고 거기다 +1을 하면 홀수를 만들어 내므로 마지막 bit가 무조건 1이되어 tail length가 무조건 0이 된다. 이러한 상황은 미리 방지해야 한다.

Exercise 4.5.3

i	1	2	3	4	5	6	7	8	9
X_i	3	1	4	1	3	4	2	1	2
X_i.value	2	3	2	2	1	1	2	1	1
Estimate	27	45	27	27	9	9	27	9	9

Average of Estimate $= \frac{27 \times 4 + 45 + 9 \times 4}{9} = 21$

(b) Implement the DGIM algorithm.

When k1=1, k2=10, k3=100, The result is

```
0.5
2.0
23.0
```