

# EE412 Foundation of Big Data Analytics, Fall 2021

## HW4

Name: 노현섭

Student ID: 20190220

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

(a)

$$W1 = \begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{bmatrix}, W2 = \begin{bmatrix} w_{11}^2 & w_{21}^2 \\ w_{12}^2 & w_{22}^2 \end{bmatrix}, p = W1 * x, q = \sigma(p), r = W2 * q, o = \sigma(r)$$

(x := Input, o := Output)

$$loss = \frac{1}{2} \sum_i (o_i - y_i)^2, g(o) = o - y$$

$$\rightarrow g(o_1) = o_1 - y_1, g(o_2) = o_2 - y_2$$

$$g(r_i) = J_r(o_i) g(o_i) = o_i(1 - o_i) * g(o_i) = o_i(1 - o_i)(o_i - y_i)$$

$$g(w_{ij}^2) = J_{w_{ij}^2}(r_j) g(r_j) = q_j * g(r_j) = q_j o_j(1 - o_j)(o_j - y_j) \quad (\because r_i = w_{1i}^2 q_1 + w_{2i}^2 q_2)$$

$$g(q_i) = J_{q_i}(r_1) g(r_1) + J_{q_i}(r_2) g(r_2) = w_{1i}^2 g(r_1) + w_{2i}^2 g(r_2)$$

$$g(p_i) = J_{p_i}(q_i) g(q_i) = q_i(1 - q_i) g(q_i)$$

$$\therefore g(w_{ij}^1) = J_{w_{ij}^1}(p_j) g(p_j) = p_j * g(p_j) = p_j q_j(1 - q_j) g(q_j)$$

$$= p_j q_j(1 - q_j) \{ w_{1j}^2 o_1(1 - o_1)(o_1 - y_1) + w_{2j}^2 o_2(1 - o_2)(o_2 - y_2) \}$$

## Answer to Problem 2

### Exercise 4.4.1 and Exercise 4.4.2

(a)

|             |   |   |   |   |    |    |   |    |    |
|-------------|---|---|---|---|----|----|---|----|----|
| Stream      | 3 | 1 | 4 | 1 | 5  | 9  | 2 | 6  | 5  |
| $h(x)$      | 7 | 3 | 9 | 3 | 11 | 19 | 5 | 13 | 11 |
| Tail length | 0 | 0 | 0 | 0 | 0  | 0  | 0 | 0  | 0  |

Estimate of the number of distinct elements =  $2^0 = 1$

(b)

|             |    |    |    |    |    |   |    |    |    |
|-------------|----|----|----|----|----|---|----|----|----|
| Stream      | 3  | 1  | 4  | 1  | 5  | 9 | 2  | 6  | 5  |
| $h(x)$      | 16 | 10 | 19 | 10 | 22 | 2 | 13 | 25 | 22 |
| Tail length | 4  | 1  | 0  | 1  | 1  | 1 | 0  | 0  | 1  |

Estimate of the number of distinct elements =  $2^4 = 16$

(c)

|             |    |   |    |   |    |   |   |    |    |
|-------------|----|---|----|---|----|---|---|----|----|
| Stream      | 3  | 1 | 4  | 1 | 5  | 9 | 2 | 6  | 5  |
| $h(x)$      | 12 | 4 | 16 | 4 | 20 | 4 | 8 | 24 | 20 |
| Tail length | 2  | 2 | 4  | 2 | 2  | 2 | 3 | 3  | 2  |

Estimate of the number of distinct elements =  $2^4 = 16$

Prob. If 'a' is even and 'b' is odd, the result of the hash is always odd, so the tail length is always 0. Therefore, it is not possible to properly estimate the number of individual elements. We should avoid such situations.

### Exercise 4.5.3

Stream: 3, 1, 4, 1, 3, 4, 2, 1, 2

|                      |   |   |   |   |   |   |   |   |   |
|----------------------|---|---|---|---|---|---|---|---|---|
| i                    | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| $X_i.\text{element}$ | 3 | 1 | 4 | 1 | 3 | 4 | 2 | 1 | 2 |
| $X_i.\text{value}$   | 2 | 3 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |