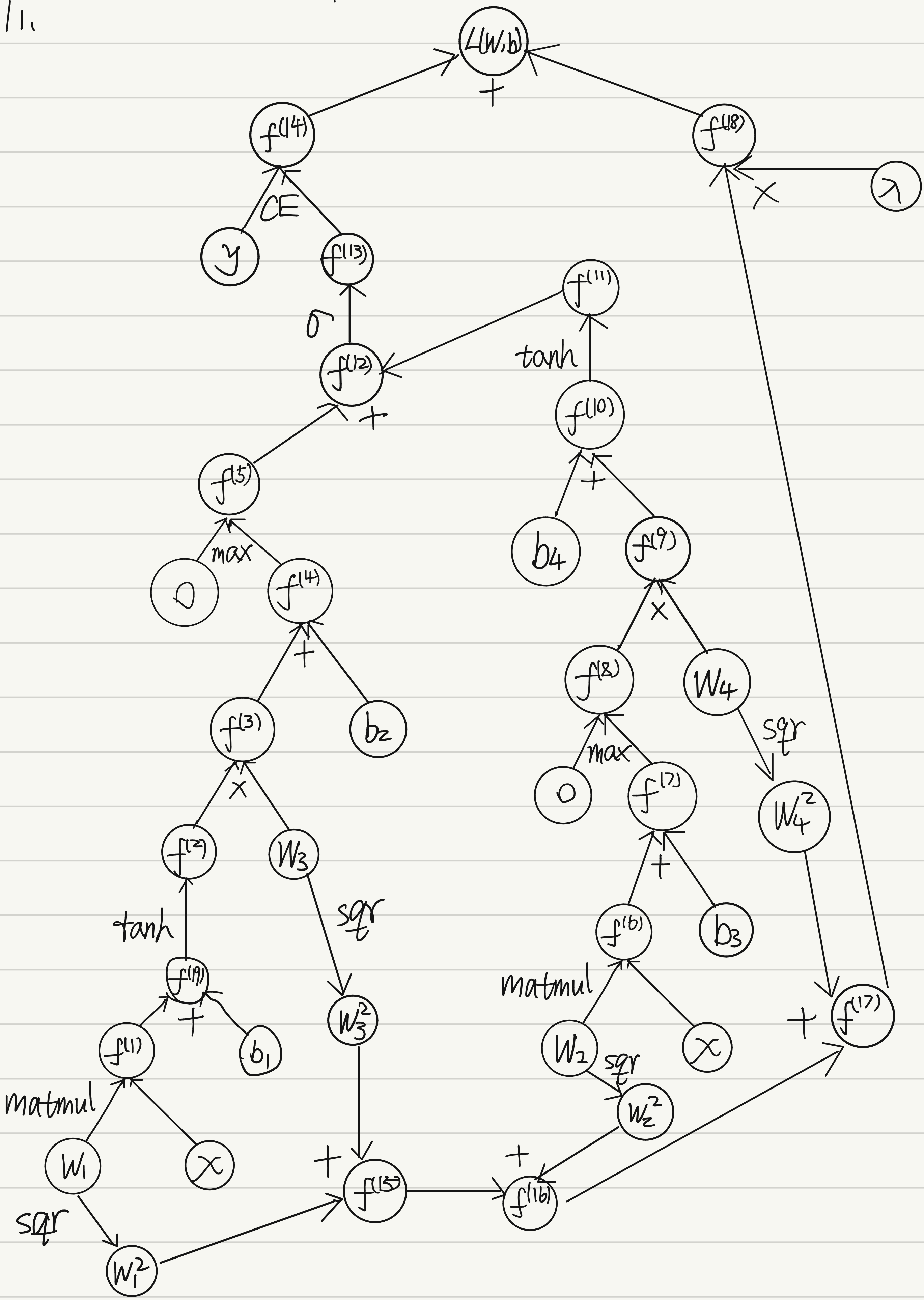
DDA2020 Assignment 3

Name：费祥　　Student ID：120090414

T1.



$$\frac{\partial L(w,b)}{\partial f^{(14)}} = \frac{dL(w,b)}{dL} \cdot \frac{\partial L}{\partial f^{(14)}} = \frac{\partial L}{\partial f^{(14)}}$$

$$\frac{\partial L(w,b)}{\partial f^{(13)}} = \frac{\partial L}{\partial f^{(14)}} \cdot \frac{\partial f^{(14)}}{\partial f^{(13)}}$$

$$\frac{\partial L(w,b)}{\partial f^{(12)}} = \frac{\partial L}{\partial f^{(13)}} \cdot \frac{\partial f^{(13)}}{\partial f^{(12)}}$$

$$\frac{\partial L(w,b)}{\partial f^{(5)}} = \frac{\partial L}{\partial f^{(12)}} \cdot \frac{\partial f^{(12)}}{\partial f^{(5)}}$$

$$\frac{\partial L}{\partial f^{(4)}} = \frac{\partial L}{\partial f^{(5)}} \cdot \frac{\partial f^{(5)}}{\partial f^{(4)}}, \quad \frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial f^{(4)}} \cdot \frac{\partial f^{(4)}}{\partial b_2}$$

$$\frac{\partial L}{\partial f^{(3)}} = \frac{\partial L}{\partial f^{(4)}} \cdot \frac{\partial f^{(4)}}{\partial f^{(3)}}, \quad \frac{\partial L}{\partial f^{(2)}} = \frac{\partial L}{\partial f^{(3)}} \cdot \frac{\partial f^{(3)}}{\partial f^{(2)}}$$

$$\frac{\partial L}{\partial f^{(19)}} = \frac{\partial L}{\partial f^{(2)}} \cdot \frac{\partial f^{(2)}}{\partial f^{(19)}}, \quad \frac{\partial L}{\partial f^{(1)}} = \frac{\partial L}{\partial f^{(19)}} \cdot \frac{\partial f^{(19)}}{\partial f^{(1)}}$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial f^{(19)}} \cdot \frac{\partial f^{(19)}}{\partial b_1}, \quad \frac{\partial L}{\partial f^{(11)}} = \frac{\partial L}{\partial f^{(12)}} \cdot \frac{\partial f^{(12)}}{\partial f^{(11)}}$$

$$\frac{\partial L}{\partial f^{(10)}} = \frac{\partial L}{\partial f^{(11)}} \cdot \frac{\partial f^{(11)}}{\partial f^{(10)}}, \quad \frac{\partial L}{\partial b_4} = \frac{\partial L}{\partial f^{(10)}} \cdot \frac{\partial f^{(10)}}{\partial b_4}$$

$$\frac{\partial L}{\partial f^{(9)}} = \frac{\partial L}{\partial f^{(10)}} \cdot \frac{\partial f^{(10)}}{\partial f^{(9)}}, \quad \frac{\partial L}{\partial f^{(8)}} = \frac{\partial L}{\partial f^{(9)}} \cdot \frac{\partial f^{(9)}}{\partial f^{(8)}}$$

$$\frac{\partial L}{\partial f^{(7)}} = \frac{\partial L}{\partial f^{(8)}} \cdot \frac{\partial f^{(8)}}{\partial f^{(7)}}, \quad \frac{\partial L}{\partial b_3} = \frac{\partial L}{\partial f^{(7)}} \cdot \frac{\partial f^{(7)}}{\partial b_3}$$

$$\frac{\partial L}{\partial f^{(6)}} = \frac{\partial L}{\partial f^{(7)}} \cdot \frac{\partial f^{(7)}}{\partial f^{(6)}}, \quad \frac{\partial L}{\partial f^{(18)}} = \frac{dL(w,b)}{dL} \cdot \frac{\partial L}{\partial f^{(18)}} = \frac{\partial L}{\partial f^{(18)}}$$

$$\frac{\partial L}{\partial f^{(17)}} = \frac{\partial L}{\partial f^{(18)}} \cdot \frac{\partial f^{(18)}}{\partial f^{(17)}}, \quad \frac{\partial L}{\partial (W_4^2)} = \frac{\partial L}{\partial f^{(17)}} \cdot \frac{\partial f^{(17)}}{\partial (W_4^2)}$$

$$\frac{\partial L}{\partial W_4} = \frac{\partial L}{\partial f^{(9)}} \cdot \frac{\partial f^{(9)}}{\partial W_4} + \frac{\partial L}{\partial (W_4^2)} \cdot 2W_4, \quad \frac{\partial L}{\partial f^{(16)}} = \frac{\partial L}{\partial f^{(17)}} \cdot \frac{\partial f^{(17)}}{\partial f^{(16)}}$$

$$\frac{\partial L}{\partial (W_2^2)} = \frac{\partial L}{\partial f^{(16)}} \cdot \frac{\partial f^{(16)}}{\partial (W_2^2)}, \quad \frac{\partial L}{\partial W_2} = \frac{\partial L}{\partial f^{(6)}} \cdot \frac{\partial f^{(6)}}{\partial W_2} + \frac{\partial L}{\partial (W_2^2)} \cdot 2W_2$$

$$\frac{\partial L}{\partial f^{(15)}} = \frac{\partial L}{\partial f^{(16)}} \cdot \frac{\partial f^{(16)}}{\partial f^{(15)}}, \quad \frac{\partial L}{\partial (W_3^2)} = \frac{\partial L}{\partial f^{(15)}} \cdot \frac{\partial f^{(15)}}{\partial (W_3^2)}$$

$$\frac{\partial L}{\partial W_3} = \frac{\partial L}{\partial f^{(3)}} \cdot \frac{\partial f^{(3)}}{\partial W_3} + \frac{\partial L}{\partial (W_3^2)} \cdot 2W_3$$

$$\frac{\partial L}{\partial (W_1^2)} = \frac{\partial L}{\partial f^{(15)}} \cdot \frac{\partial f^{(15)}}{\partial (W_1^2)}, \quad \frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial f^{(1)}} \cdot \frac{\partial f^{(1)}}{\partial W_1} + \frac{\partial L}{\partial (W_1^2)} \cdot 2W_1$$

$\therefore$ $W_1 = W_1 - \alpha \frac{\partial L}{\partial W_1}$, $W_2 = W_2 - \alpha \frac{\partial L}{\partial W_2}$

$W_3 = W_3 - \alpha \frac{\partial L}{\partial W_3}$, $W_4 = W_4 - \alpha \frac{\partial L}{\partial W_4}$

$b_1 = b_1 - \alpha \frac{\partial L}{\partial b_1}$, $b_2 = b_2 - \alpha \frac{\partial L}{\partial b_2}$

$b_3 = b_3 - \alpha \frac{\partial L}{\partial b_3}$, $b_4 = b_4 - \alpha \frac{\partial L}{\partial b_4}$

T2,

Conv_1: $(63+2\times2-5)/2+1 = 32$

$\therefore$ the shape of activation map: $32\times32\times1$

the number of parameters: $(5\times5\times3+1)\times10 = 760$

the computational cost: $[5\times5\times3+(5\times5\times3-1)+1]\times32\times32\times10 = 1536000$

Maxpool_1: $(32-2)/3+1 = 11$

$\therefore$ the shape of activation map: $11\times11\times1$

the number of parameters: 0

the computational cost: $2\times2\times11\times11\times10 = 4840$

Conv_2: $(11+2\times1-3)/2+1 = 6$

$\therefore$ the shape of activation map: $6\times6\times1$

the number of parameters: $(3\times3\times10+1)\times20 = 1820$

the computational cost: $[3\times3\times10+(3\times3\times10-1)+1]\times6\times6\times20 = 129600$

Maxpool_2: $(6+2\times1-2)/2+1 = 4$

$\therefore$ the shape of activation map: $4\times4\times1$

the number of parameters: 0

the computational cost: $2\times2\times4\times4\times20 = 1280$

total number of parameters: 2580

total computational cost: 1671720

T3.

Gini index:

$$G_A = 1-\left[\left(\frac{5}{18}\right)^2+\left(\frac{5}{18}\right)^2+\left(\frac{8}{18}\right)^2\right] = 0.648$$

$$G_{B_1} = 1-\left[\left(\frac{1}{8}\right)^2+\left(\frac{2}{8}\right)^2+\left(\frac{5}{8}\right)^2\right] = 0.531$$

$$G_{B_2} = 1-\left[\left(\frac{4}{10}\right)^2+\left(\frac{6}{10}\right)^2\right] = 0.48$$

$$G_{C_1} = 1-\left[\left(\frac{1}{6}\right)^2+\left(\frac{5}{6}\right)^2\right] = 0.278$$

$$G_{C_2} = 1-1^2 = 0$$

Entropy:

$$E_A = -\frac{5}{18}\log_2\frac{5}{18}-\frac{5}{18}\log_2\frac{5}{18}-\frac{8}{18}\log_2\frac{8}{18} = 1.547$$

$$E_{B_1} = -\frac{1}{8}\log_2\frac{1}{8}-\frac{2}{8}\log_2\frac{2}{8}-\frac{5}{8}\log_2\frac{5}{8} = 1.299$$

$$E_{B_2} = -\frac{4}{10}\log_2\frac{4}{10}-\frac{6}{10}\log_2\frac{6}{10} = 0.971$$

$$E_{C_1} = -\frac{1}{6}\log_2\frac{1}{6}-\frac{5}{6}\log_2\frac{5}{6} = 0.65$$

$$E_{C_2} = -1\cdot\log_2(1) = 0$$

Misclassification error:

$$M_A = 1-\frac{8}{18} = \frac{5}{9}$$

$$M_{B_1} = 1-\frac{5}{8} = \frac{3}{8}$$

$$M_{B_2} = 1-\frac{6}{10} = \frac{2}{5}$$

$$M_{C_1} = 1-\frac{5}{6} = \frac{1}{6}$$

$$M_{C_2} = 1-1 = 0$$

T4.

(a) empirical MSE:

$$\widehat{MSE} = \frac{1}{10}\times\left[(6-7)^2+(8-7)^2+(9-7)^2+(5-7)^2+(10-7)^2+(5-7)^2 + (4-7)^2+(8-7)^2+(9-7)^2+(3-7)^2\right] = 5.3$$

$$\overline{h}_D(X=3) = \frac{1}{10}\times(6+8+9+5+10+5+4+8+9+3) = 6.7$$

$$Bias^2 = \left[\overline{h}_D(X=3) - t(X=3)\right]^2 = (6.7-6.7)^2 = 0$$

$$Variance = \frac{1}{10}\sum_{i}^{10}(h_{D_i}(X=3) - \overline{h}_D(X=3))^2$$

$$= \frac{1}{10}\times\left[(6-6.7)^2+(8-6.7)^2+(9-6.7)^2+(5-6.7)^2+(10-6.7)^2+(5-6.7)^2 + (4-6.7)^2+(8-6.7)^2+(9-6.7)^2+(3-6.7)^2\right] = 5.21$$

(b) reason ①:

in fact, the theoretical equation is:

$$E_{(x,y),D}\left[(h_D(x)-y)^2\right] = E_{(x,y),D}\left[(h_D(x)-\overline{h}(x))^2\right] + E_{(x,y),D}\left[(\overline{h}(x)-t(x))^2\right] + E_{x,y}\left[(t(x)-y)^2\right]$$

and we need to use integral and the probability distribution to compute each expectation term. But in reality, we don't know the distribution and our sample is limited within finite data point.

reason ②:

here we suppose the error $\varepsilon \in N(0,\sigma^2)$ and $\sigma^2 = 0.5$ but in fact, the error may not satisfy the normal distribution we have supposed.

reason ③:

even in the discrete case, we have:

$$\widehat{MSE}(x,y) = \frac{1}{10}\sum_{i=1}^{10}(h_{D_i}(x) - y)^2$$

$$= \frac{1}{10}\sum_{i=1}^{10}(h_{D_i}(x) - \bar{h}(x) + \bar{h}(x) - y)^2$$

$$= \frac{1}{10}\sum_{i=1}^{10}[2(h_{D_i}(x) - \bar{h}(x))(\bar{h}(x) - y)] + \frac{1}{10}\sum_{i=1}^{10}[h_{D_i}(x) - \bar{h}(x)]^2$$

$$+ \frac{1}{10}\sum_{i=1}^{10}[\bar{h}(x) - y]^2 = \frac{1}{10}\sum_{i=1}^{10}[h_{D_i}(x) - \bar{h}(x)]^2 + \frac{1}{10}\sum_{i=1}^{10}[\bar{h}(x) - y]^2$$

$$= \frac{1}{10}\sum_{i=1}^{10}[h_{D_i}(x) - \bar{h}(x)]^2 + \frac{1}{10}\sum_{i=1}^{10}[\bar{h}(x) - t(x) + t(x) - y]^2$$

$$= \frac{1}{10}\sum_{i=1}^{10}[h_{D_i}(x) - \bar{h}(x)]^2 + \frac{1}{10}\sum_{i=1}^{10}(\bar{h}(x) - t(x))^2 + \frac{1}{10}\sum_{i=1}^{10}[t(x) - y]^2$$

$$+ \frac{1}{10}\sum_{i=1}^{10}[2(\bar{h}(x) - t(x))(t(x) - y)]$$

$$= \frac{1}{10}\sum_{i=1}^{10}[h_{D_i}(x) - \bar{h}(x)]^2 + \frac{1}{10}\sum_{i=1}^{10}(\bar{h}(x) - t(x))^2 + \frac{1}{10}\sum_{i=1}^{10}[t(x) - y]^2$$

$$= \text{bias}^2 + \text{Variance} + \frac{1}{10}\sum_{i=1}^{10}[t(x) - y]^2$$

and $\sigma^2 \neq \frac{1}{10}\sum_{i=1}^{10}[t(x) - y]^2$

in fact, it is similar to limited sample reason

T5、

$$\sigma(a) = \frac{1}{1+e^{-a}}$$

$$\tanh(a) = \frac{e^{a}-e^{-a}}{e^{a}+e^{-a}}$$

$$\therefore \quad \tanh(a) = 2\sigma(2a) - 1$$

$$y_k(x, w) = \sigma\left(\sum_{j=1}^{M} w_{kj}^{(2)} \sigma\left(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right) + w_{k0}^{(2)}\right)$$

$$\hat{y}_k(x, \hat{w}) = \sigma\left(\sum_{j=1}^{M} \hat{w}_{kj}^{(2)} \left[2\sigma\left(2\left(\sum_{i=1}^{D} \hat{w}_{ji}^{(1)} x_i + \hat{w}_{j0}^{(1)}\right)\right) - 1\right] + \hat{w}_{k0}^{(2)}\right)$$

$$= \sigma\left(\sum_{j=1}^{M} 2\hat{w}_{kj}^{(2)} \sigma\left(\sum_{i=1}^{D} 2\hat{w}_{ji}^{(1)} x_i + 2\hat{w}_{j0}^{(1)}\right) - \sum_{j=1}^{M} \hat{w}_{kj}^{(2)} + \hat{w}_{k0}^{(2)}\right)$$

we let $2\hat{w}_{kj}^{(2)} = w_{kj}^{(2)}$

$$2\hat{w}_{ji}^{(1)} = w_{ji}^{(1)}$$

$$2\hat{w}_{j0}^{(1)} = w_{j0}^{(1)}$$

$$\hat{w}_{k0}^{(2)} - \sum_{j=1}^{M} \hat{w}_{kj}^{(2)} = w_{k0}^{(2)}$$

then $y_k(x, w) = \hat{y}_k(x, \hat{w})$, and the above is linear transformation between w and $\hat{w}$.

$\therefore$ Q.E.D.