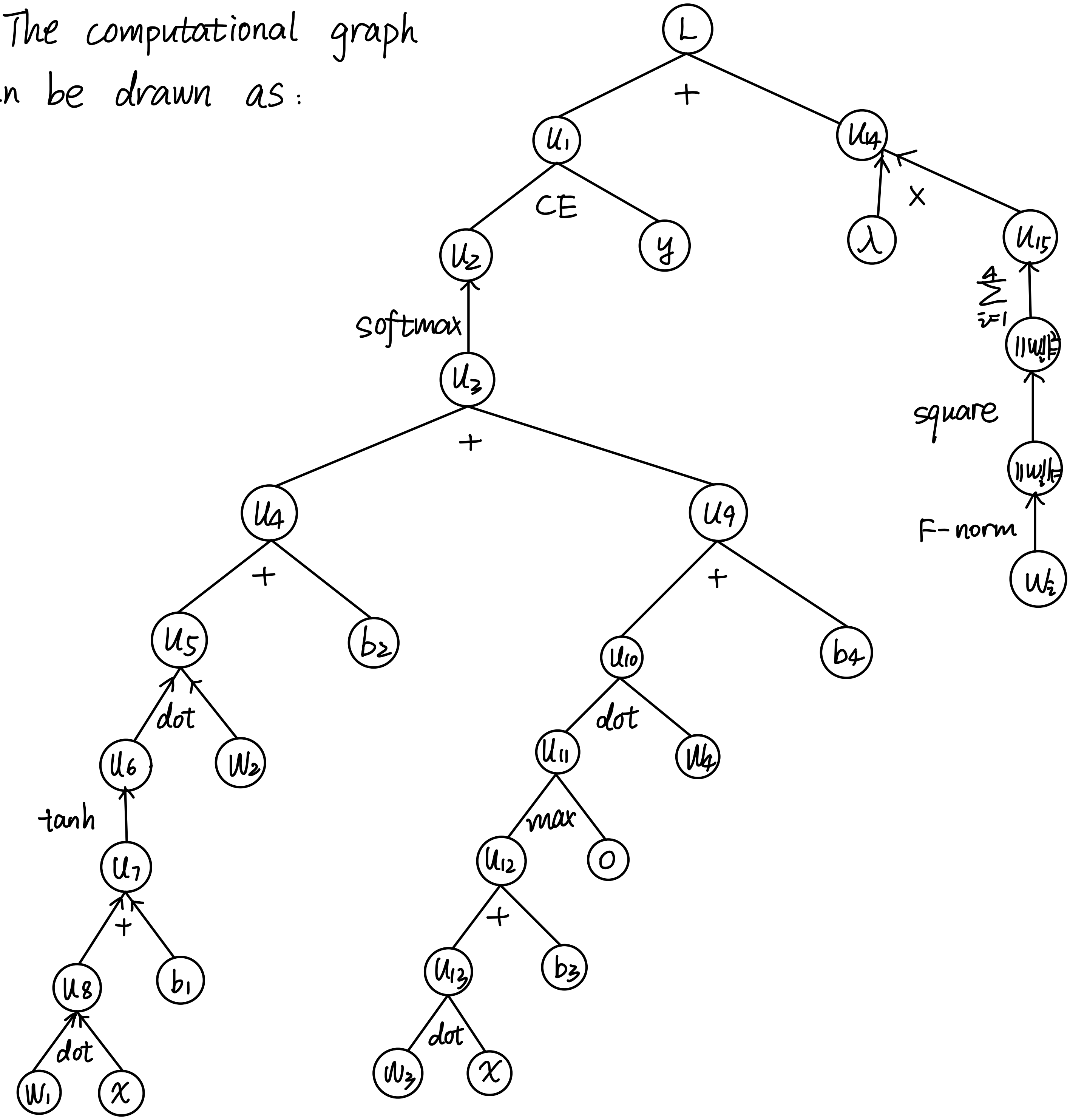
1 Written Problems

1、The computational graph can be drawn as:



$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial u_{14}} \cdot \frac{\partial u_{14}}{\partial u_{15}} \cdot \frac{\partial u_{15}}{\partial \|W\|_F^2} \cdot 2W_1 + \frac{\partial L}{\partial u_1} \cdot \frac{\partial u_1}{\partial u_2} \cdot \frac{\partial u_2}{\partial u_3} \cdot \frac{\partial u_3}{\partial u_4} \cdot \frac{\partial u_4}{\partial u_5} \cdot \frac{\partial u_5}{\partial u_6} \cdot \frac{\partial u_6}{\partial u_7} \cdot x$$

$$\frac{\partial L}{\partial W_2} = \frac{\partial L}{\partial u_{14}} \cdot \frac{\partial u_{14}}{\partial u_{15}} \cdot \frac{\partial u_{15}}{\partial \|W_3\|_F^2} \cdot 2W_2 + \frac{\partial L}{\partial u_1} \cdot \frac{\partial u_1}{\partial u_2} \cdot \frac{\partial u_2}{\partial u_3} \cdot \frac{\partial u_3}{\partial u_4} \cdot \frac{\partial u_4}{\partial u_5} \cdot \frac{\partial u_5}{\partial W_2}$$

$$\frac{\partial L}{\partial W_3} = \frac{\partial L}{\partial u_{14}} \cdot \frac{\partial u_{14}}{\partial u_{15}} \cdot \frac{\partial u_{15}}{\partial \|W_3\|_F^2} \cdot 2W_3 + \frac{\partial L}{\partial u_1} \cdot \frac{\partial u_1}{\partial u_2} \cdot \frac{\partial u_2}{\partial u_3} \cdot \frac{\partial u_3}{\partial u_9} \cdot \frac{\partial u_9}{\partial u_{10}} \cdot \frac{\partial u_{10}}{\partial u_{11}} \cdot \frac{\partial u_{11}}{\partial u_{12}} \cdot x$$

$$\frac{\partial L}{\partial W_4} = \frac{\partial L}{\partial u_{14}} \cdot \frac{\partial u_{14}}{\partial u_{15}} \cdot \frac{\partial u_{15}}{\partial \|W_4\|_F^2} \cdot 2W_4 + \frac{\partial L}{\partial u_1} \cdot \frac{\partial u_1}{\partial u_2} \cdot \frac{\partial u_2}{\partial u_3} \cdot \frac{\partial u_3}{\partial u_9} \cdot \frac{\partial u_9}{\partial u_{10}} \cdot \frac{\partial u_{10}}{\partial W_4}$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial u_1} \cdot \frac{\partial u_1}{\partial u_2} \cdot \frac{\partial u_2}{\partial u_3} \cdot \frac{\partial u_3}{\partial u_4} \cdot \frac{\partial u_4}{\partial u_5} \cdot \frac{\partial u_5}{\partial u_6} \cdot \frac{\partial u_6}{\partial u_7}$$

$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial u_1} \cdot \frac{\partial u_1}{\partial u_2} \cdot \frac{\partial u_2}{\partial u_3} \cdot \frac{\partial u_3}{\partial u_4}$$

$$\frac{\partial L}{\partial b_3} = \frac{\partial L}{\partial u_1} \cdot \frac{\partial u_1}{\partial u_2} \cdot \frac{\partial u_2}{\partial u_3} \cdot \frac{\partial u_3}{\partial u_9} \cdot \frac{\partial u_9}{\partial u_{10}} \cdot \frac{\partial u_{10}}{\partial u_{11}} \cdot \frac{\partial u_{11}}{\partial u_{12}}$$

$$\frac{\partial L}{\partial b_4} = \frac{\partial L}{\partial u_1} \cdot \frac{\partial u_1}{\partial u_2} \cdot \frac{\partial u_2}{\partial u_3} \cdot \frac{\partial u_3}{\partial u_9}$$

2. (1) For Conv$_1$: $N=100$, $F=5$, stride $=3$

We need $N+2P_1-F/\text{stride}$ to be an integer

$(95+2P_1) \bmod 3 = 0 \Rightarrow P_1 = 2.$

For Maxpool$_1$: $N=(100+4-5)/3+1 = 34$, $F=2$, stride $=2$

$(34+2P_2-2) \bmod 2 = 0 \Rightarrow P_2 = 0$

For Conv$_2$: $N=(34-2)/2+1=17$, $F=3$, stride $=2$

$(17+2P_3-3) \bmod 2 = 0 \Rightarrow P_3 = 0$

For Maxpool$_2$: $N=(17-3)/2+1 = 8$, $F=3$, stride $=3$

$(8+2P_4-3) \bmod 3 = 0 \Rightarrow P_4 = 2$

(2) For Conv$_1$: $(100+4-5)/3+1 = 34$

Shape: $34 \times 34 \times 8$

Parameters: $(5\times5\times3+1)\times 8 = 608$

For MaxPool$_1$: $(34+0-2)/2+1 = 17$

Shape: $17\times17\times8$

Parameters: 0

For Conv$_2$: $(17+0-3)/2+1 = 8$

Shape: $8\times8\times16$

Parameters: $(3\times3\times8+1)\times16 = 1168$

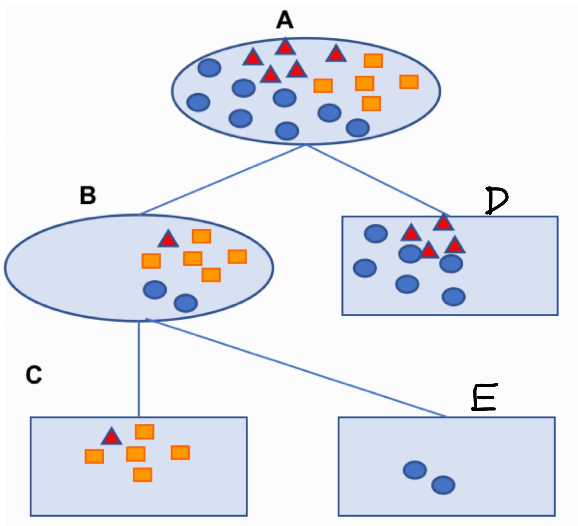
For MaxPool2: $(8+2\times2-3)/3+1=4$

Shape: $4\times4\times16$

Parameters: 0

Total number of parameters: $608+0+1168+0=1776$

3.



A: $P_\triangle=\frac{5}{18}$ $P_\square=\frac{5}{18}$ $P_\circ=\frac{8}{18}=\frac{4}{9}$

Gini index:

$$\varphi(p)=1-\left(\frac{5}{18}\right)^2-\left(\frac{5}{18}\right)^2-\left(\frac{4}{9}\right)^2=0.648$$

Entropy:

$$\varphi(p)=-\left(\frac{5}{18}\log_2\frac{5}{18}+\frac{5}{18}\log_2\frac{5}{18}+\frac{4}{9}\log_2\frac{4}{9}\right)$$

$=1.547$

Classification Error:

$$\varphi(p)=1-\max(p_i)=1-\frac{4}{9}=0.556$$

B: $P_\triangle=\frac{1}{8}$, $P_\square=\frac{5}{8}$, $P_\circ=\frac{1}{4}$

Gini index:

$$\varphi(p)=1-\left(\frac{1}{8}\right)^2-\left(\frac{5}{8}\right)^2-\left(\frac{1}{4}\right)^2=0.531$$

Entropy:

$$\varphi(p)=-\left(\frac{1}{8}\log_2\frac{1}{8}+\frac{5}{8}\log_2\frac{5}{8}+\frac{1}{4}\log_2\frac{1}{4}\right)$$

$=1.299$

Classification Error:

$$\varphi(p)=1-\max(p_i)=1-\frac{5}{8}=0.375$$

C: $P_\triangle=\frac{1}{6}$ $P_\square=\frac{5}{6}$

Gini index:

$$\varphi(p)=1-\left(\frac{1}{6}\right)^2-\left(\frac{5}{6}\right)^2=0.278$$

Entropy:

$$\varphi(p)=-\left(\frac{1}{6}\log_2\frac{1}{6}+\frac{5}{6}\log_2\frac{5}{6}\right)$$

$=0.650$

Classification Error:

$$\varphi(p)=1-\max(p_i)=0.167$$

D: $P_\triangle=\frac{4}{10}=\frac{2}{5}$, $P_\circ=\frac{3}{5}$

Gini index:

$$\varphi(p)=1-\left(\frac{2}{5}\right)^2-\left(\frac{3}{5}\right)^2=0.48$$

Entropy:

$$\varphi(p)=-\left(\frac{2}{5}\log_2\frac{2}{5}+\frac{3}{5}\log_2\frac{3}{5}\right)$$

$=0.971$

Classification Error:

$$\varphi(p)=1-\max(p_i)=\frac{2}{5}$$

E: $P_\circ=1$

Gini index: $1-1^2=0$

Entropy: $-1\log_2 1=0$

Classification Error: $1-1=0$

4. (a) $\hat{MSE} = \frac{1}{10}(1^2+1^2+2^2+2^2+2^2+2^2+3^2+1^2+2^2+3^2) = 4.1$

Avg prediction $= \frac{1}{10}(6+8+9+5+9+5+4+8+9+4) = 6.7$

$\text{Bias}^2 = (\bar{h}_D(x) - t(x))^2 = (6.7 - 7.2)^2 = 0.25$

$\text{Variance} = \frac{1}{10}(0.7^2 + 1.3^2 + 2.3^2 + 1.7^2 + 2.3^2 + 1.7^2 + 2.7^2 + 1.3^2 + 2.3^2 + 2.7^2)$

$= 4.01$

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

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

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

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

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

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

$(t(x) - y)^2$

$= \text{Variance} + \text{Bias}^2 + \varepsilon^2 + 2(\bar{h}(x) - t(x))(t(x) - y)$

Since $\hat{MSE} = \text{Variance} + \text{Bias}^2 + \varepsilon^2 + 2(\bar{h}(x) - t(x))(t(x) - y)$,

$\varepsilon^2 + 2(\bar{h}(x) - t(x))(t(x) - y) = 0.04 - 0.2 = -0.16 \neq 1 = \sigma^2$

So we conclude that $\hat{MSE} \neq \text{Variance} + \text{Bias}^2 + \sigma^2$ under these 10 models.

5. $\sigma(a) = \frac{1}{1+e^{-a}}$, $\tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$

$1 - 2\sigma(a) = \frac{1+e^{-a}}{1+e^{-a}} - \frac{2}{1+e^{-a}} = \frac{e^{-a} - 1}{1+e^{-a}} = -\frac{1-e^{-a}}{1+e^{-a}}$

$$= -\frac{e^{\frac{a}{2}} - e^{-\frac{a}{2}}}{e^{\frac{a}{2}} + e^{-\frac{a}{2}}} = -\tanh\left(\frac{a}{2}\right)$$

Thus, we obtain that $1 - 2\sigma(a) = -\tanh\left(\frac{a}{2}\right)$

$$\Rightarrow \tanh(a) = 2\sigma(2a) - 1$$

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

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

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

Compare it with the original $y_k(x, w)$, we can get:

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

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

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

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

Thus, there exists linear transformation between these w, $\hat{w}$, that enable $y_k(x, w) = \hat{y}_k(x, \hat{w})$ for all x.