Application No.: (210485)

Name: Yang Yuxin

Question No.: 2

Method 1:

One layer MLP:

$a_0^{(1)} = W_{0,0}^{(1)} a_0^{(0)} + W_{0,1}^{(1)} a_1^{(0)} + b_0^{(1)}$

$a_1^{(1)} = W_{1,0}^{(1)} a_0^{(0)} + W_{1,1}^{(1)} a_1^{(0)} + b_1^{(1)}$

General:

$$\vec{a}^{(n)} = W^{(n)} \vec{a}^{(n-1)} + \vec{b}^{(n)}$$
$$= W^{(n)} [W^{(n-1)} \vec{a}^{(n-2)} + b^{(n-1)}] + \vec{b}^{(n)}$$
$$= X_1^n \vec{a}^{(0)} + X_2^n \vec{b}^{(1)} + X_3^n \vec{b}^{(3)} + \dots + X_n^n \vec{b}^{(n-1)} + \vec{b}^n$$
$$= \sum_{i=1}^{n} X_i^n \vec{a}^{(i-1)} + \vec{b}^n$$

where $X_i^n = W^n W^{n-1} W^{n-2} \dots W^1$

Network 1 is equivant with Network 2.

$\widetilde{W} = X_1^n = W^3 \times W^1 \times W^2 \quad (n=3)$

$$\widetilde{b} = \sum_{i=1}^{3} X_i^n \vec{b}^{(i-1)} - X_1^n \vec{a}^{(0)} + \vec{b}^{(n)} \quad (n=3)$$
$$= X_2^3 \vec{b}^{(1)} + X_3^3 \vec{b}^{(2)} + \vec{b}^{(3)}$$
$$= W^3 W^2 \vec{b}^{(1)} + W^3 \vec{b}^{(2)} + \vec{b}^{(3)}$$

Method 2:

For Network 1 (MLP with Muptiple hidden layers)

The represtion of the first hidden layer:

$$\begin{cases} \vec{a}^1 = W^1 \vec{a}^0 + \vec{b}^1 & ① \\ \vec{a}^2 = W^2 \vec{a}^1 + \vec{b}^2 & ② \\ \vec{a}^3 = W^3 \vec{a}^2 + \vec{b}^3 & ③ \end{cases}$$

$\Rightarrow \vec{a}^3 = W^3 W^2 \vec{a}^2 + W^3 \vec{b}^2 + \vec{b}_3$ (②③)

$\Rightarrow \vec{a}^3 = W^3 W^2 W^1 \vec{a}^0 + \underline{W^3 W^2 \vec{b}^1 + W^3 \vec{b}^2 + \vec{b}^3}$ (①②③)

Since $W^3 W^2 W^1$ is a 5×5 matrix

$\vec{b}^1, \vec{b}^2, \vec{b}^3$ is a 5×1 matrix

W^3, W^2, W^1 is a 5×5 matrix.

Thus $W^3 W^2 \vec{b}_1 + W^3 \vec{b}^2 + \vec{b}^3$ is a 5×1 matrix.

$\widetilde{W} = W^3 W^2 W^1$

$\widetilde{b} = W^3 W^2 \vec{b}^1 + W^3 \vec{b}^2 + \vec{b}^3$