2.2

Network 1 is given that, $\vec{a}^{(1)} = W^{(1)}\vec{a}^{(0)} + \vec{b}^{(1)}$

$$\vec{a}^{(2)} = W^{(2)}\vec{a}^{(1)} + \vec{b}^{(2)}$$

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

, where input $= \vec{a}^{(0)}$

output $= \vec{a}^{(3)}$

To make network 2 is equivalent to network 1, all inputs and outputs of both networks need to be identical.

Alternatively, for Network 2, the input is $\vec{a}^{(0)}$, and the output is $\vec{a}^{(3)}$

$\vec{a}^{(3)} = \tilde{W}\vec{a}^{(0)} + \tilde{b}$, find $\tilde{W}$ and $\tilde{b}$.

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

$$= W^{(3)}\left[W^{(2)}\left(W^{(1)}\vec{a}^{(0)} + \vec{b}^{(1)}\right) + \vec{b}^{(2)}\right] + \vec{b}^{(3)}$$

$$= W^{(3)}\left[W^{(2)}W^{(1)}\vec{a}^{(0)} + W^{(2)}\vec{b}^{(1)} + \vec{b}^{(2)}\right] + \vec{b}^{(3)}$$

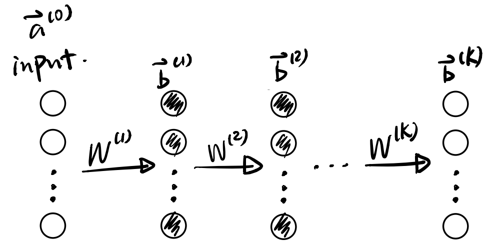
$$= W^{(3)}W^{(2)}W^{(1)}\vec{a}^{(0)} + W^{(3)}W^{(2)}\vec{b}^{(1)} + W^{(3)}\vec{b}^{(2)} + \vec{b}^{(3)}$$

$\longrightarrow$ Therefore,

$$\boxed{\begin{aligned} \tilde{W} &= W^{(3)}W^{(2)}W^{(1)} \\ \tilde{b} &= W^{(3)}W^{(2)}\vec{b}^{(1)} + W^{(3)}\vec{b}^{(2)} + \vec{b}^{(3)} \end{aligned}}$$

, where $W^{(1)}$, $W^{(2)}$, $W^{(3)}$, $\vec{b}^{(1)}$, $\vec{b}^{(2)}$, and $\vec{b}^{(3)}$ are given from Network 1.

※ <u>Generalization:</u>



$$\begin{bmatrix} \vec{a}^{(1)} \\ \vec{a}^{(2)} \\ \vdots \\ \vec{a}^{(k)} \end{bmatrix} = \begin{bmatrix} W^{(1)} & & & 0 \\ & W^{(2)} & & \\ & & \ddots & \\ 0 & & & W^{(k)} \end{bmatrix} \begin{bmatrix} \vec{a}^{(0)} \\ \vec{a}^{(1)} \\ \vdots \\ \vec{a}^{(k-1)} \end{bmatrix} + \begin{bmatrix} \vec{b}^{(1)} \\ \vec{b}^{(2)} \\ \vdots \\ \vec{b}^{(k)} \end{bmatrix}$$

→ the network above is MLP with $(k-1)$ hidden layers.

→ Tranfer it to MLP with no hidden layers that is equivalent to the MLP with $(k-1)$ hidden layers:

Output $= \tilde{W}\vec{a}^{(0)} + \tilde{b}$, where input $= \vec{a}^{(0)}$

Find $\tilde{W}$ and $\tilde{b}$.

$$\vec{a}^{(k)} = W^{(k)}\vec{a}^{(k-1)} + \vec{b}^{(k)}$$

$$= W^{(k)}W^{(k-1)}W^{(k-2)}\cdots W^{1}\vec{a}^{(0)} + W^{(k)}W^{(k-1)}\cdots W^{(2)}\vec{b}^{(1)} + \cdots + W^{(k)}\vec{b}^{(k-1)} + \vec{b}^{(k)}$$

Thus, $\tilde{W} = W^{(k)}W^{(k-1)}\cdots W^{(1)}$

$\tilde{b} = W^{(k)}W^{(k-1)}\cdots W^{(2)}\vec{b}^{(1)} + \cdots + W^{(k)}\vec{b}^{(k-1)} + \vec{b}^{(k)}$

For example, if $k=5$, there will be 4 hidden layers.

$$\tilde{W} = W^{(5)}W^{(4)}W^{(3)}W^{(2)}W^{(1)}\vec{a}^{(0)}$$

$$\tilde{b} = W^{(5)}W^{(4)}W^{(3)}W^{(2)}\vec{b}^{(1)} + W^{(5)}W^{(4)}W^{(3)}\vec{b}^{(2)} + W^{(5)}W^{(4)}\vec{b}^{(3)} + W^{(5)}\vec{b}^{(4)} + \vec{b}^{(5)}$$