Question 2

For network 2, given any input vector $\overrightarrow{a_2}$, we can get the output vector $\overrightarrow{o_2}$ which equals:

$$\overrightarrow{o_2} = \widetilde{W}\,\overrightarrow{a_2} + \tilde{b} \qquad (1)$$

For network 1, given any input vector $\overrightarrow{a_1}$, we can also get the hidden layer outputs $\overrightarrow{h_{11}}, \overrightarrow{h_{12}}$, as well as the final output $\overrightarrow{o_1}$:

$$\overrightarrow{h_{11}} = W^{(1)}\overrightarrow{a_1} + \vec{b}^{(1)} \qquad (2)$$

$$\begin{aligned} \overrightarrow{h_{12}} &= W^{(2)}\,\overrightarrow{h_{11}} + \vec{b}^{(2)} \\ &= W^{(2)}\left(W^{(1)}\overrightarrow{a_1} + \vec{b}^{(1)}\right) + \vec{b}^{(2)} \\ &= (W^{(2)}W^{(1)})\overrightarrow{a_1} + W^{(2)}\vec{b}^{(1)} + \vec{b}^{(2)} \qquad (3) \end{aligned}$$

$$\begin{aligned} \overrightarrow{o_1} &= W^{(3)}\overrightarrow{h_{12}} + \vec{b}^{(3)} \\ &= W^{(3)}((W^{(2)}W^{(1)})\overrightarrow{a_1} + W^{(2)}\vec{b}^{(1)} + \vec{b}^{(2)}) + \vec{b}^{(3)} \\ &= (W^{(3)}W^{(2)}W^{(1)})\overrightarrow{a_1} + W^{(3)}W^{(2)}\vec{b}^{(1)} + W^{(3)}\,\vec{b}^{(2)} + \vec{b}^{(3)} \qquad (4) \end{aligned}$$

Since network 1 and network2 are equivalent, $\overrightarrow{o_2}$ is identical to $\overrightarrow{o_1}$.
Combining equation (1) and equation (4), we get the following equation:

$$\widetilde{W}\,\overrightarrow{a_2} + \tilde{b} = (W^{(3)}W^{(2)}W^{(1)})\overrightarrow{a_1} + (W^{(3)}W^{(2)}\vec{b}^{(1)} + W^{(3)}\,\vec{b}^{(2)} + \vec{b}^{(3)})$$

Then we get the result:

$$\begin{cases} \widetilde{W} = W^{(3)}W^{(2)}W^{(1)} \\ \tilde{b} = (W^{(3)}W^{(2)}\vec{b}^{(1)} + W^{(3)}\,\vec{b}^{(2)} + \vec{b}^{(3)} \end{cases}$$