Question 2

2.2) Based on network 1, the output is

$$\begin{aligned}
\vec{a}^{(3)} &= W^{(3)}\vec{a}^{(2)} + \vec{b}^{(3)} \\
&= W^{(3)}[W^{(2)}\vec{a}^{(1)} + \vec{b}^{(2)}] + \vec{b}^{(3)} \\
&= W^{(3)}W^{(2)}\vec{a}^{(1)} + W^{(3)}\vec{b}^{(2)} + \vec{b}^{(3)} \\
&= W^{(3)}W^{(2)}[W^{(1)}\vec{a}^{(0)} + \vec{b}^{(1)}] + W^{(3)}\vec{b}^{(2)} + \vec{b}^{(3)} \\
&= W^{(3)}W^{(2)}W^{(1)}\vec{a}^{(0)} + W^{(3)}W^{(2)}\vec{b}^{(1)} + W^{(3)}\vec{b}^{(2)} + \vec{b}^{(3)}
\end{aligned}$$

Thus, for network 1 and 2 to be equivalent, the weights and bias for network 2 must be

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