Question 2 – Equivalent networks

Network 1 (two hidden layers):

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

Network 2 (no hidden layers):

$\vec{a}^{(out)} = \widetilde{W}\vec{a}^{(in)} + \tilde{b}$ ---------- (4)

Given that Network 1 and Network 2 are equivalent:

$\vec{a}^{(0)} = \vec{a}^{(in)}$ ---------- (5)

$\vec{a}^{(3)} = \vec{a}^{(out)}$ ---------- (6)

Substitute Equation (6) into Equation (3):

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

Substitute Equation (2):

$$\begin{aligned}\vec{a}^{(out)} &= W^{(3)}\left(W^{(2)}\vec{a}^{(1)} + \vec{b}^{(2)}\right) + \vec{b}^{(3)} \\ &= W^{(2)}W^{(3)}\vec{a}^{(1)} + W^{(3)}\vec{b}^{(2)} + \vec{b}^{(3)}\end{aligned}$$

Substitute Equation (1):

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

Substitute Equation (5):

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

Substitute Equation (4):

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

Network 2's weights $(\widetilde{W})$ and bias $(\tilde{b})$:

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

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