Q2

Wang Yuge

2021-03-07

Assume that the input vector is $\vec{a}^{(0)}$ and the output at the layers in Figure 2a are $\vec{a}^{(1)}$, $\vec{a}^{(2)}$, and $\vec{a}^{(3)}$. For the network in Figure 2a, we have

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

For the network in Figure 2b, we have

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

Then,

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