AStar BII Test
Question 2

Joel Ang

31 May, 2020

A multilayer perceptron can be rewritten the following way.

$$\begin{aligned}
\mathbf{a}^{(l)} &= W^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^{(l)} \\
&= W^{(l)} \cdot (W^{(l-1)}\mathbf{a}^{(l-2)} + \mathbf{b}^{(l-1)}) + \mathbf{b}^{(l)} \\
&= W^{(l)}W^{(l-1)}\mathbf{a}^{(l-2)} + (W^{(l)}\mathbf{b}^{(l-1)} + \mathbf{b}^{(l)}) \\
&= ... \\
&= \Omega^{(l)}_{(1)}\mathbf{a}^{(0)} + (\Omega^{(l)}_{(2)}\mathbf{b}^{(1)} + \Omega^{(l)}_{(3)}\mathbf{b}^{(2)} + ... + \Omega^{(l)}_{(l)}\mathbf{b}^{(l-1)} + \mathbf{b}^{(l)}) \\
&= \Omega^{(l)}_{(1)}\mathbf{a}^{(0)} + \beta^{(l)} \\
\text{where } \ \Omega^{(l)}_{(i)} &= W^{(l)}W^{(l-1)}...W^{(i)} \\
\beta^{(l)} &= \Omega^{(l)}_{(2)}\mathbf{b}^{(1)} + \Omega^{(l)}_{(3)}\mathbf{b}^{(2)} + ... + \Omega^{(l)}_{(l)}\mathbf{b}^{(l-1)} + \mathbf{b}^{(l)}
\end{aligned}$$

So, an equivalent single layer network would simply be such that.

$$\begin{aligned}
\tilde{W} &= \Omega^{(l)}_{(1)} \\
\tilde{\mathbf{b}} &= \beta^{(l)}
\end{aligned}$$

and in this case, $l = 3$, so

$$\begin{aligned}
\tilde{W} &= \Omega^{(3)}_{(1)} \\
&= W^{(3)}W^{(2)}W^{(1)} \\
\tilde{\mathbf{b}} &= \beta^{(3)} \\
&= \Omega^{(3)}_{(2)}\mathbf{b}^{(1)} + \Omega^{(3)}_{(3)}\mathbf{b}^{(2)} + \mathbf{b}^{(3)} \\
&= W^{(3)}W^{(2)}\mathbf{b}^{(1)} + W^{(3)}\mathbf{b}^{(2)} + \mathbf{b}^{(3)}
\end{aligned}$$