Proof upon the Equivalence between Multi-layered Network Model with Linear Activation Function for Hidden Layers and Single-layered Network Model

Shaun Leo

School of Public Affairs and Administration, University of Electronic Science and Technology

1 Processing

Considering the activation function as the following

$$\varphi(\mathrm{x}) = k \cdot \mathrm{x} + h \tag{1}$$

To calculate the next layer from the present i-th layer, the function can be spread as

$$\mathrm{a}^{(i+1)} = \varphi(\mathrm{W}_{n_{i+1} \times n_i} \mathrm{a}^{(i)} + \mathrm{b}^{(i)}) \tag{2}$$

$$\mathrm{a}^{(i+2)} = \varphi(\mathrm{W}_{n_{i+2} \times n_{i+1}} \mathrm{a}^{(i+1)} + \mathrm{b}^{(i+1)}) \tag{3}$$

Replace the term $\mathrm{a}^{(i+1)}$ in the equation (3) with the right part in the equation (2) to generate

$$\mathrm{a}^{(i+2)} = \varphi\big(\mathrm{W}_{n_{i+2} \times n_{i+1}} \varphi(\mathrm{W}_{n_{i+1} \times n_i} \mathrm{a}^{(i)} + \mathrm{b}^{(i)}) + \mathrm{b}^{(i+1)}\big) \tag{4}$$

$$\mathrm{a}^{(i+2)} = k^{(i+1)} \cdot k^{(i)} \cdot \mathrm{W}_{n_{i+2} \times n_i} \cdot \mathrm{a}^{(i)} + o(\mathrm{b}^{(i+2)}, \mathrm{b}^{(i+1)}, \mathrm{b}^{(i)}) \tag{5}$$

Take the epoch from the equation (2) to the equation (5) for $j - i$ times to get the following consequence

$$\mathrm{a}^{(j)} = \prod_{\iota=i}^{j} k^{(\iota)} \cdot \mathrm{W}_{n_j \times n_i} \cdot \mathrm{a}^{(i)} + o(\mathrm{b}^{(j)}, \ldots, \mathrm{b}^{(i)}) \tag{6}$$

The term $o(\mathrm{b}^{(j)}, \ldots, \mathrm{b}^{(i)})$ in the equation (6), as well as a function, appears short for all the other arguments amid the network model besides of the input data a. The equation (6) states linear in the same architecture as the initial equation, the activation function.

2 Conclusion

Multi-layered network with linear activation functions appears equivalent to single-layer network mathematically and consequentially.