Linear activation functions in a Neural Network equate to a single layer Perceptron: Demonstration

Luis Ramírez Guzmán

November 19, 2021

The goal of this exercise is to prove that if every activation function in a deep feed-forward neural network is linear, as well as the activation functions for the output layer, the network serves as a simple linear model for any input.

Starting wih proper definition: Let there be N input neurons, M output neurons, P hidden layers with Q neurons each. Let F be the set of linear activation functions where $f_i \in F$ and z_{pq} any input for any of this functions, in the form $f(z)$.

Let $v = [v1, v2, ...v_N]$ be values of the input neurons with corresponding weights $w = [w_1, w_2, ...w_N]$.
Let the weights of the network be

$$W = \begin{bmatrix} w_{11} & w_{21} & \dots & w_{P1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1Q} & w_{2Q} & \dots & w_{PQ} \end{bmatrix}$$

where W_p is a column corresponding to the weights of a single layer.

Let $p \in P$, $q \in Q$, so $z_{pq} = w_1 v_1 + w_2 v_2 + \cdots + w_N v_N$ given that $p = 1$ is the input to a neuron of the first hidden layer.

Notice that for any other of the hidden layers and the output layer, the input to the neuron's activation function has the same form, but replacing $v_n | n \in N$ for the evaluation of the activtion function of a previous layer's neuron $f_{(p-1)q}(z)$.

Lastly we define $f_{pq}(z)$ to be a linear combination, thus a linear function.

$$f_{pq}(z) = W_{p-1} \cdot f_{p-1}(z)$$

so, it follows

$$f_{(p-1)q}(z) = W_{p-2} \cdot f_{p-2}(z)$$

For the output layer we have the expression $f_M(z) = W_p \cdot f_p(z)$ and with it we can finally express the goal: the expression for $f_M(z)$ is a linear function. It then becomes an issue of determining wether $f_p(z)$ is a linear function, which is in turn defined in terms of $f_{p-1}(z)$.
Finally , it can be expressed:

$$f_M(z) = f_M(W_p \cdot f_p(W_{p-1} \cdot f_{p-1}(\ldots W_1 \cdot f_1(w \cdot v)\ldots))) \quad \square$$