DEEP LEARNING

Lecture notes

Simone Carletti

A.Y. 2024/2025

Contents

1 Feed Forward Neural Networks 2

Chapter 1

Feed Forward Neural Networks

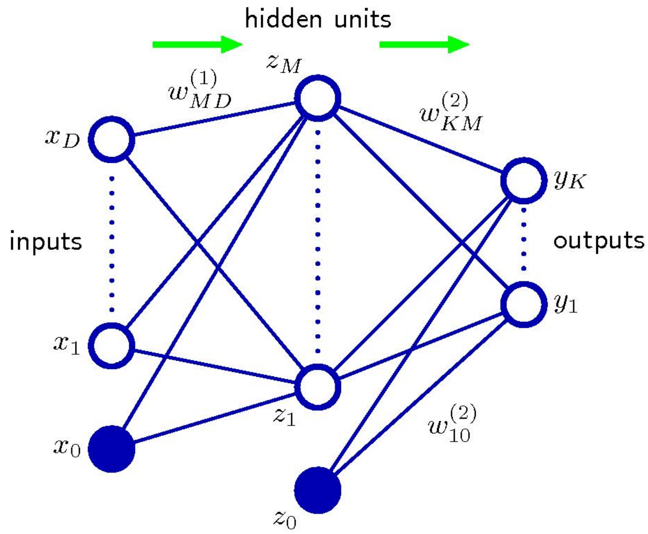

Figure 1.1

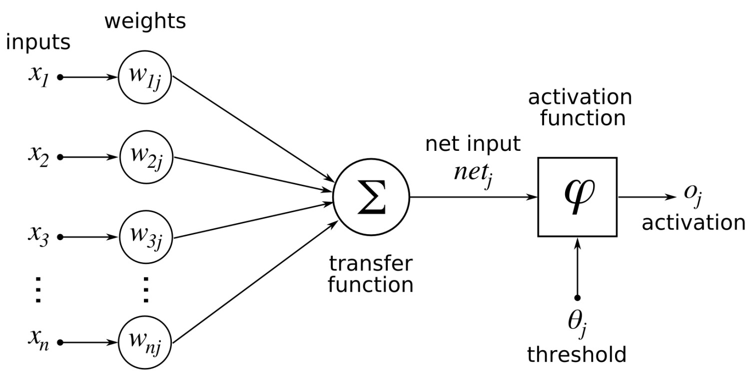

Figure 1.2

Each layer first computes what is equivalent to the output a_j of a linear statistical model and then applies a non-linear function to the linear model output. For the **first layer**, which takes the network input x (that is a vector $x = (x_0, \ \ldots, \ x_D)$) as input, these two steps look like this:

$$
\begin{aligned}
a_j^{(1)} &= \sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)} \\
z_j^{(1)} &= h_1(a_j^{(1)})
\end{aligned}
$$

where h_1 is the non-linear (activation) function in the first layer. The superscript $^{(k)}$ indicates the index of the layer.
We can get rid of writing the so-called **bias** $w_{j0}^{(1)}$ explicitly by adding an extra input x_0 that is always set to one and extending the sum to go from zero:

$$
a_j^{(1)} = \sum_{i=0}^{D} w_{ji}^{(1)} x_i
$$

We will use this notation in the following for all layers and just write for example $\sum_i \ldots$ to be understood as $\sum_{i=0}^{D} \ldots$

The **second layer** takes the output the of the first layer as input:

$$
a_j^{(2)} = \sum_{i}^{M} w_{ji}^{(2)} z_i^{(1)}
$$

The second layer non-linear function is denoted by h_2 so the output of the network is

$$
y_j = h_2(a_j^{(2)})
$$

This gives an example of how the neural network model input to output mapping can be specified.

$$
h_2\left(\sum_{i}^{M} w_{ji}^{(2)} h_1\left(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right)\right)
$$