Notes on Neural Networks

Marcio Woitek

May 24, 2020

Model Representation

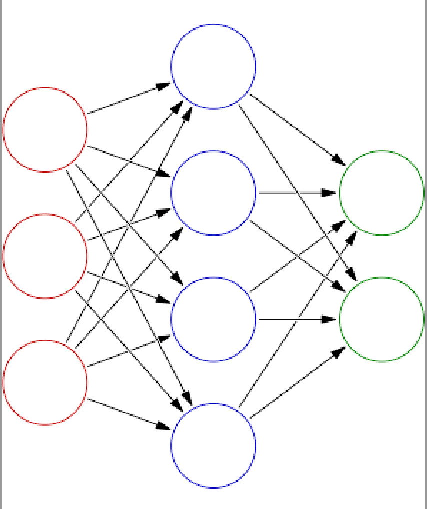

Figure: Basic Structure of a Neural Network

Model Representation

Fundamental Ideas

- A neural network is formed by nodes or *units*.
- The units are organized in *layers*.
- There are at least two layers, the *input layer* and the *output layer*.
- However, very often, between these layers there are *hidden layers*.
- In Fig. 1, the red nodes form the input layer, the blue nodes form the hidden layer, and the green nodes form the output layer.
- For simplicity, in the next slides, we discuss the particular case of the neural network in this figure.

Model Representation

Underlying Mathematical Concepts

- First, consider the input layer.
- We assume there are n_1 input units. In the case of Fig. 1, we have $n_1 = 3$.
- We denote the input values by $x_1, x_2, \ldots, x_{n_1}$.
- There can also be an extra unit, the so-called *bias unit*. Its value is represented by x_0.
- It is convenient to organize all these values in a single $(n_1 + 1)$-dimensional column vector:

$$\mathbf{a}^{(1)} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n_1} \end{bmatrix}. \tag{1}$$

Model Representation

Underlying Mathematical Concepts

- We can use a similar idea to represent the values related to the other layers.
- To be more precise, we assume the neural network has L layers.
- Then the values associated with the j-th layer will be organized in a column vector $\mathbf{a}^{(j)}$ ($j = 1, \ldots, L$).
- The vector $\mathbf{a}^{(j)}$ is $(n_j + 1)$-dimensional, where n_j denotes the number of units in the j-th layer.
- This vector has an extra dimension because the value associated with the bias unit of this layer is also included in $\mathbf{a}^{(j)}$.

Example of Fig. 1

- As an example, we discuss the neural network in Fig. 1.
- Clearly, this network has $L = 3$ layers.
- Then the unit values are organized in 3 vectors, $\mathbf{x} = \mathbf{a}^{(1)}$, $\mathbf{a}^{(2)}$ and $\mathbf{y} = \mathbf{a}^{(3)}$.
- From Fig. 1, we can see that the numbers of units are given by $n_1 = 3$, $n_2 = 4$ and $n_3 = 2$.
- Therefore, taking into account the values of the bias units, we can state the following: $\mathbf{a}^{(1)} \in \mathbb{R}^4$, $\mathbf{a}^{(2)} \in \mathbb{R}^5$ and $\mathbf{a}^{(3)} \in \mathbb{R}^3$.

Forward Propagation

- Next, we explain how the input values are propagated in the network.
- Essentially, we start from the input vector $\mathbf{x} = \mathbf{a}^{(1)}$, and map $\mathbf{a}^{(j)}$ to $\mathbf{a}^{(j+1)}$ until we obtain the output vector $\mathbf{y} = \mathbf{a}^{(L)}$.
- In our example, the idea is the following: first, the input vector $\mathbf{x} = \mathbf{a}^{(1)}$ is mapped to $\mathbf{a}^{(2)}$, and then this vector is mapped to $\mathbf{a}^{(3)}$, producing the output $\mathbf{y}$.
- To be specific about how these mappings work, we introduce 2 vectors, $\mathbf{z}^{(2)}$ and $\mathbf{z}^{(3)}$.
- In general, we introduce $L-1$ vectors $\mathbf{z}^{(j)}$, where $j = 2, \ldots, L$.
- We also define a set with $L-1$ functions $g^{(j)}$, where $j = 2, \ldots, L$. The function $g^{(j)}$ is called the *activation function* of the j-th layer. Notice that there isn't an activation function associated with the input layer.
- To discuss our example, we consider the particular case in which all activation functions are the same. We denote this function by h.

Forward Propagation

Activation Function

- For the activation functions, there are many choices.
- To discuss our example, we consider the particular case in which h is the *sigmoid* or *logistic function*:

$$h(x) = \frac{1}{1 + \exp(-x)}. \tag{2}$$

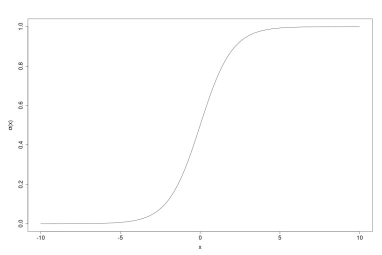

Figure: Graph of the sigmoid function.

Forward Propagation

Forward Propagation Equations

- We continue by explaining how the vector $\mathbf{z}^{(j)}$ is computed from the unit values in $\mathbf{a}^{(j-1)}$.
- We imagine all units in the $(j-1)$-th layer (including the bias unit) are connected to every unit in the j-th layer (excluding the bias unit).
- Then between these layers there are $(n_{j-1}+1)\, n_j$ connections.
- Each of these connections has a different *weight*. We will organize all these weights in a matrix denoted by $\Theta^{(j-1)}$.
- The element $\Theta_{ik}^{(j-1)}$ in the i-th row and k-th column of this matrix is interpreted as follows: it is the weight of the connection between the k-th unit in the $(j-1)$-th layer and the i-th unit in the j-th layer.
- Then the dimension of the matrix $\Theta^{(j-1)}$ is $n_j \times (n_{j-1}+1)$.

Forward Propagation

Forward Propagation Equations

- Finally, we can write down the equations that determine the vector $\mathbf{z}^{(j)}$.
- The i-th component of this vector is given by the *weighted sum* of the components of $\mathbf{a}^{(j-1)}$:

$$\begin{aligned} z_i^{(j)} &= \Theta_{i0}^{(j-1)} a_0^{(j-1)} + \Theta_{i1}^{(j-1)} a_1^{(j-1)} + \ldots + \Theta_{i,n_{j-1}}^{(j-1)} a_{n_{j-1}}^{(j-1)} \\ &= \sum_{k=0}^{n_{j-1}} \Theta_{ik}^{(j-1)} a_k^{(j-1)}. \end{aligned} \tag{3}$$

- The last expression can be written in matrix form as follows:

$$\mathbf{z}^{(j)} = \Theta^{(j-1)} \mathbf{a}^{(j-1)}. \tag{4}$$

Forward Propagation

Forward Propagation Equations

- Now, to obtain the vector $\mathbf{a}^{(j)}$, we only need to apply the activation function $g^{(j)}$ to every component of $\mathbf{z}^{(j)}$:

$$\mathbf{a}^{(j)} = g^{(j)}\left(\mathbf{z}^{(j)}\right) = \begin{bmatrix} g^{(j)}\left(z_1^{(j)}\right) \\ g^{(j)}\left(z_2^{(j)}\right) \\ \vdots \\ g^{(j)}\left(z_{n_j}^{(j)}\right) \end{bmatrix}. \tag{5}$$

- .