A Short Introduction to Neural Networks

Tasadduk Chowdhury

July 21, 2020

1 Abstract

Write me please.

2 A toy example

A neural network is a complex mapping of the space of inputs to the space of outputs. A neural network which maps R^2 to R^2 can be constructed the following way. I will use a mathematical convention and write $f(x_1, x_2) = (\hat{y}_1, \hat{y}_2)$ to refer to this mapping. How one arrives at $(\hat{y}_1, \hat{y}_2)$ is entirely up to the developer implementing the network f. With just a single hidden layer with 3 neurons, f will be defined as follows.

The network has 3 layers: input layer, 1 hidden layer, output layer. The elements of each layer are called *neurons*. Our input layer has 2 neurons,

3 Description

One training example consists of an input vector $x = (x_1, \dots, x_{n^1})$ and an output vector $y = (y_1, \dots, y_{n^L})$.

The are L layers in the network. The inputs to the network are in layer 1 and the outputs are in layer L. Layers $2, 3, \dots, L-1$ are the hidden layers.

In the beginning, layer ℓ has a number of inputs. Then, layer ℓ passes linear combinations of the inputs (aka. *weighted inputs*) to each neuron in layer $\ell + 1$. After that, layer $\ell + 1$ *activates* the received signals. The activated signals serve as "inputs" to layer $\ell + 1$. This process is applied sequentally to all the hidden layers $\ell = 2, \dots, L-1$.

To proceed further, we first need to define a list of symbols. The subscripts k and j represent the indeces for neurons.

- $w^{\ell}_{j,k}$: weight for the connection from the kth neuron in the $(\ell - 1)$th layer to the jth neuron in the ℓth layer.

- b^{ℓ}_{j}: bias of the jth neuron in the ℓth layer.

- a_j^ℓ: activation of the jth neuron in the ℓth layer.
- z_j^ℓ: weighted input for the jth neuron in the ℓth layer.

The weights $w_{j,k}^\ell$ and the biases b_j^ℓ are the parameters of a neural network. The jth neuron in layer ℓ receives a weighted input,

$$z_j^\ell = \sum_k w_{j,k}^\ell a_k^{\ell-1} + b_j^\ell, \tag{1}$$

from layer $\ell - 1$. The weighted input z_j^ℓ is then transformed or *activated* by a function $g : R \to R$,

$$a_j^\ell = g(z_j^\ell). \tag{2}$$

In layer 1, the input neurons do not get activated and so we have $a_k^1 = x_k$ for $k = 1, \ldots, n^1$.

4 Logistic regression as a simple network

Write me please.