

Practical backpropagation revision

Rafael

August 12, 2018

Abstract

Based on Matt Mazur weblog

1 Introduction

There are quite a number of scientific articles that detail the formalities of artificial neural networks (ANNs). While some background knowledge about derivatives (gradient descent), chain rule, and vector algebra is arguably enough to understand the nuts and bolts of the training of ANNs, there are underlying aspects that could remain obscure to grasp for some readers.

This document is focused on starting on a specific example of an ANN for its training and testing. Based on the Matt Mazur's weblog¹, we go beyond and provide both the practical and technical background to integrally understand the process.

The rest of this document employs the nomenclature shown in Table 1.

2 Sample ANN architecture

Figure 1 shows the sample ANN architecture that will be employed during this tutorial. Notice the following:

¹<https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/>

Symbol	Meaning
L	The number of layers of the feedforward multilayer network
$w_{ji}^{(l)}$	The individual weight in layer l coming from neuron i that arrives to the neuron j of layer $(l + 1)$.
v_j^l	The local field (<i>net</i>) of a given network. This is the dot product between the activations and weights coming from neurons of the immediate previous layer.
φ	The activation (squash) function applied to the local field (<i>net</i>) of a neuron. Let it be the sigmoid function.
$\varphi(v_j^l)$	The output of the neuron j of layer l .
...	...

Table 1: Symbols employed on the description

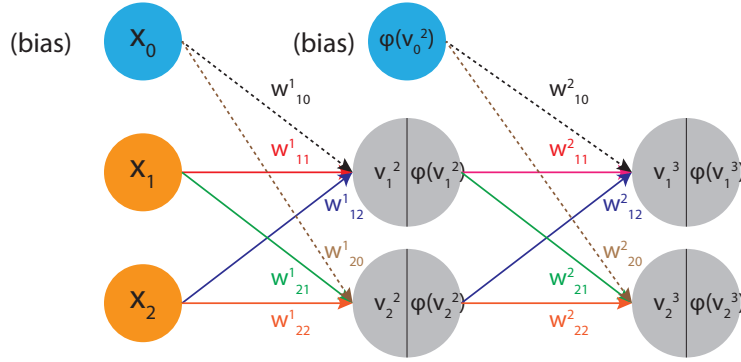


Figure 1: The sample ANN to be employed during this tutorial

- This is a 2-2-2 architecture, since there are two inputs (or attributes), two neurons in a hidden layer, and two neurons in the output layer.
- The employed notation allows to see that indexing could be employed for easily referencing each weight, local field (v) and activation φ .
- Biases are represented with a 0 subindex (the top separated circles). Biases are included in the calculation of the local field v .
- There are initial values for input (x) and weights (w). We will thoroughly employ them in the following description.
- ...

3 Feedforward stage

If you have arrived to this step, presumably you known some of the following points:

- The *heavy* work of an ANN is carried out by the set of *weights*.
- You can use different input data x but they are actually somehow fixed.
- The performance of the ANN is modified by the adjustments the weights.
- Somehow, every labeled example-pattern should improve the performance of the ANN; as a result, somehow the errors found should modify the weights.
- The derivative could be employed as a way to determine whether weight values should be increased-reduced; this is linked to the gradient descend algorithm.

The rest of this section will try to describe the feedforward step, on which we evaluate the performance (accuracy) of the network predictions in order to later perform the weights adjustments (backpropagation).