Contents

Abstract

Machine learning is a growing area in computer science. Neural networks is very popular subdomain of it. This project generally covers the work of M. Nielsen and his book Neural Networks and Deep Learning[1]. We will first introduce neual networks and then aplly it to a generic problem which is called classifying handwritten digits.

2. Introduction

The idea of neural networks is to take a large number of dataset known as training examples, and then develop a system which can learn from those training examples. In other words, the neural network uses the examples to automatically infer rules for recognizing other unknown examples. Furthermore, by increasing the number of training examples, the network can learn more about handwriting, and so improve its accuracy.

There are just 100 training digits below, perhaps we could build a better handwriting recognizer by using thousands or even millions or billions of training examples.

Figure 1: 100 Handwritten Digits

This project is concerned with write a computer program implementing a neural network that learns to recognize handwritten digits.

Along the way there are many key ideas about neural networks, including two important types of artificial neuron (the perceptron and the sigmoid neuron), and the standard learning algorithm for neural networks, known as stochastic gradient descent.

1.1 Perceptrons

Perceptron is a type of artificial neuron. Perceptrons were developed in the 1950s and 1960s by the scientist Frank Rosenblatt, inspired by earlier work by Warren McCulloch and Walter Pitts. Today, it's more common to use other models of artificial neurons - in this book, and in much modern work on neural networks, the main neuron model used is one called the sigmoid neuron.

A perceptron takes several binary inputs, $x_1, x_2, \ldots$, and produces a single binary output:

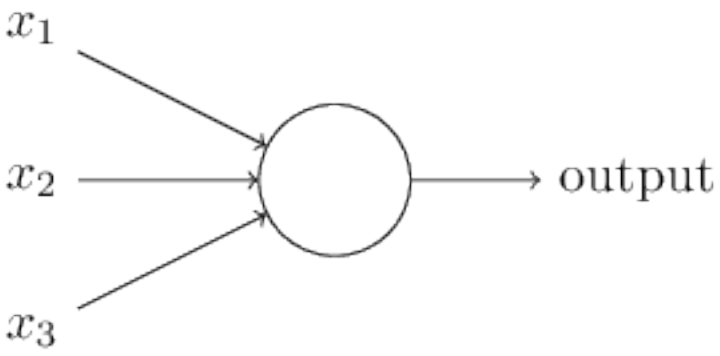

Figure 2: Perceptron

In general it could have more or fewer inputs. Rosenblatt proposed a simple rule to compute the output. He introduced weights, $w_1, w_2, \ldots$, real numbers expressing the importance of the respective inputs to the output. The neuron's output, 0 or 1, is determined by whether the weighted sum $\sum_j w_j x_j$ is less than or greater than some threshold value. Just like the weights, the threshold is a real number which is a parameter of the neuron. To put it in more precise algebraic terms:

$$output = \begin{cases} 0, & if \quad \sum_j \ w_j x_j \leqslant threshold \\ 1, & if \quad \sum_j \ w_j x_j > threshold \end{cases} \tag{1}$$