

Exercise, May 1st

This exercise is built around the Google program located on:

<https://playground.tensorflow.org>

There are four classification datasets available. The three simplest ones have already been discussed during the course on Wednesday morning.

In this exercise we will focus on the fourth classification dataset, the spiral.

The dataset is composed of 500 points. The “ratio of training to test data” is simply the ratio of training data you want to use: for example, if the ratio is 90% it means that you have 450 training data and 50 test data. Using 70% training data (350 points) and 30% test data (150 points) is a common configuration. This is what we will use in the following.

The goal of the exercise is to help you understand the impact of changing certain hyper-parameters. In this exercise we will leave the regularization option to “none” (and the regularization parameter to zero) as regularization will be discussed in details on Friday.

First, keep the input as just the two coordinates X_1 and X_2 . We will use 0.03 as the default learning rate. We will assume that data have no noise and that the default batch size is 10.

We examine the behaviour of the networks until about 2500 iterations.

1. Start by using a linear activation function and any number of hidden layers and neurons. What do you observe? How do you interpret it?
2. Now use the sigmoid activation function and no hidden layer. What do you observe? How do you interpret it?
3. Now put just one hidden layer, a sigmoid activation function, and make the number of neurons in the hidden layer equal to 4. Then make it equal to the maximum offered by the application, that is 8. What do you observe in both cases?
4. Now do the same as in question 3, but using the ReLU activation function. What do you observe?
5. Now use two hidden layers each with four neurons. What do you observe?
6. Now try two hidden layers with 8 neurons each and a sigmoid activation function. What do you observe?
7. Now try three hidden layers with 8 neurons each. What happens if you compare sigmoid and ReLU?

8. Now keep the three layers with 8 neurons and change the batch size to 30. What happens if you use ReLU?
9. Now that we have obtained a good model see what happens if you introduce all the 7 input variables : $X_1, X_2, X_1^2, X_2^2, X_1X_2, \sin X_1, \sin X_2$.