NNDL:

Problem Set #5

Table of Contents

Alejandro C Parra Garcia

Problem 1

Activation Function

Use the Relu activation function since is faster to train than the tanh for the hidden layers, according to the book. For the output layer also use the Relu activation function since we are trying to predict a salary, so the Relu return a value from 0 to ∞. And thus, is better for this case. Since the Sigmoid return 0 to 1 and the tanh return -1 to 1, and thus unsuitable for a salary.

Architecture of the Net

The idea of the architecture is to compress the information to try to predict a Salary. Thus, the number of nodes in each layer get reduced the deeper the neural network gets.

10-10-8-8-6-4-1

Loss Function

The loss function is set to MSE or Mean Squared Error, since we are trying to predict a number. And this loss function penalizes higher errors more instantly than smaller deviations.

Optimizer

According to the book page 105, Adam outperforms SGD. So, I'm going to use Adam

Result

- We can see from the Graph that the results of the model are really bad. They get worse with the higher salaries, but that can be explain because there aren't many datapoints in the higher salaries.

We can also see it with the metrics. The RMSE is 337.35, the MAE is 210.81 and the R2 is 0.42. Overall, a bad model.

Problem 2

Activation Function

Use the Relu activation function since is faster to train than the tanh for the hidden layers, according to the book.

For the output layer also use the Sigmoid activation function since we are trying to predict a class, so the sigmoid return a value from 0 to 1. That's perfect for what we need for a class.

Architecture of the Net

The idea of the architecture is to compress the information to try to predict a Class. Thus, the number of nodes in each layer get reduced the deeper the neural network gets.

10-10-8-8-6-4-3

Loss Function

The loss function is set categorical_crossentropy, since we are trying to predict a Class of 3 different values

Optimizer

According to the book page 105, Adam outperforms SGD. So, I'm going to use Adam

Result

The model is really good, as it achieves an accuracy of 0.916 on the test data. We also see that the Precision and Recall are really high. Being the lowest 0.82.

We can also see on the Confusion Matrix that the model is really good missing only 5 predictions.

Compare

We can also see that the model using keras achieve a higher accuracy than in the Mini Project 1 and in the Mini project 2.

In the case of the Mini project 2:

For the different Learning Rates, Keras got a better accuracy of 0.916 compare with the 0.9 accuracy of the learning rates 0.05 and 0.01

For nº of nodes in the hidden layer we find the same results. Keras is better

For the activation function, we see that Keras keeps winning.