Homework V - Group 56

I. DESCRIPTION OF THE NETWORKS USED IN THE EXPERIMENTS

For this Homework, we used Feed-Forward Neural Networks and Convolutional Neural Networks.

Feed-Forward Neural Networks:

For the Feed-Forward Neural Networks, we started by testing 3 different models, one with 128 neurons in the hidden layer, other with 256 neurons in the hidden layer and the last one with 512 neurons in the hidden layer. At the beginning, we had our models without regularization with one hidden layer having the sigmoid function as the activation function.

Then we choose the number of neurons that give us the best results and test 2 different models, one with the Sigmoid activation function in the hidden layer and the other with the Tanh activation function in the hidden layer.

We then choose the activation function that gave us the best results. We now try 4 different models, with 1, 2, 3 and 4 hidden layers respectively, having those hidden layers both number of neurons and activation function that gave us the best result.

After choosing the number of hidden layers that give us the best results, we test 3 different models, one without regularization, other with L1 regularization and other with L2 regularization. All of those 3 models have in the hidden layers, the number of neurons, the activation function with the best results, being the number of hidden layers also the one with the best results.

Convolutional Neural Networks:

For the Convolutional Neural Networks, we started by testing 3 different models, one with 8 filters, other with 16 filters and the last one with 32 filters. At the beginning, we had our models with a kernel size of 4, with Pooling, 1 convolutional layer and without regularization.

Then we choose the number of filters that give us the best results and test 4 different models with 2, 4, 8 and 16 kernels with that number of filters.

We then select the number of kernels that give us the best results, and use that number of kernels to test 2 different models, one with pooling and the other without it.

Then we check if the results are better with or without pooling and use the best result to test 4 models with 1, 2, 3 and 4 convolutional layers.

We then choose the number of convolutional layers that give us the best results and test the model with the fully connected layer between the convolutional and output layers, and check if it's more useful to use it or not.

We finally tested 2 models, one without regularization and the other with Dropout regularization.

II. DESCRIPTION OF THE EVALUATION APPROACH

To evaluate the results, for every configuration from both Feed-Forward and Convolutional Neural Networks, we checked what was the model with the highest accuracy and lowest loss values.

In cases where a model with a higher accuracy also has a higher loss (for example in Overfitting models), we choose the one with the highest accuracy in Validation.

In cases where both loss and accuracy are very similar, we use the simplest model.

III. PRESENTATION AND DISCUSSION OF THE RESULTS

Feed-Forward Neural Networks:

For the Feed-Forward Neural Networks, the number of neurons that gave us the best results was having the hidden layer with 128 neurons, that gave us both the lowest loss, with 0.1608, and the highest accuracy, with 0.9526. After that we had the model with 256 neurons, which gave us 0.2375 as the loss and 0.9320 as the accuracy. Finally the worst results were with 512 neurons, with 0.2677 as the loss and 0.9245 as the accuracy.

For the hidden layer activation function, the Sigmoid function gave us a loss of 0.1914 and an accuracy of 0.9446, while the Tanh function gave us a loss of 0.1076 and an accuracy of 0.9670.

For the number of hidden layers, having 1 hidden layer gave us a loss of 0.1043 and an accuracy of 0.9680. Having 2 hidden layers gave us a loss of 0.1096 and an accuracy of 0.9665. Having 3 hidden layers gave us a loss of 0.0849 and an accuracy of 0.9731. Having 4 hidden layers gave us a loss of 0.0852 and an accuracy of 0.9726.

Finally, for the regularization applied, not having regularization gave us a loss of 0.0881 and an accuracy of 0.9747. Having a L1 regularization gave us a loss of 71.4172 and an accuracy of 0.9755. Having a L2 regularization gave us a loss of 6.0901 and an accuracy of 0.9758.

Convolutional Neural Networks:

For the Convolutional Neural Networks, having 8 filters gave us a loss of 0.0487, and an accuracy of 0.9846.The model with 16 filters gave us 0.0487 as the loss and 0.9848 as the accuracy. Finally, the model with 32 filters gave us a loss of 0.0483 and an accuracy of 0.9844.

For the number of kernels, having 2 kernels gave us a loss of 0.0512 and an accuracy of 0.9842. Having 4 kernels gave us a loss of 0.0495 and an accuracy of 0.9839. Having 8 kernels gave us a loss of 0.0491 and an accuracy of 0.9843. Having 16 kernels gave us a loss of 0.0495 and an accuracy of 0.9847.

The model with pooling gave us a loss of 0.0479 and an accuracy of 0.9845. The model without pooling gave us a loss of 0.0477 and an accuracy of 0.9848.

About the number of convolutional layers, having 1 convolutional layer gave us a loss of 0.0476 and an accuracy of 0.9847. Having 2 convolutional layers gave us a loss of 0.0476 and an accuracy of 0.9852. Having 3 convolutional layers gave us a loss of 0.0475 and an accuracy of 0.9844. Having 4 convolutional layers gave us a loss of 0.0471 and an accuracy of 0.9854.

Having a fully connected layer between the convolutional and output layer gave us a loss of 0.0475 and an accuracy of 0.9844, while not having it gave us a loss of 0.0471 and an accuracy of 0.9854.

The Dropout regularization gave us a loss of 0.0465 and an accuracy of 0.9856. On the other hand, no regularization gave us a loss of 0.0464 and an accuracy of 0.9860

IV. THE MOST IMPORTANT CONCLUSIONS

Feed-Forward Neural Networks:

For the Feed-Forward Neural Networks, the number of neurons with the best results was having 128 neurons in the hidden layer, and after that was having 256 neurons and the worst was having 512 neurons. So we can conclude that lower numbers of neurons in the hidden layer contribute for a better accuracy of the model.

The activation function that gave us the best results was the Tanh activation function, and so we concluded that the Tanh function is the best for the hidden layers of Feed-Forward Neural Networks models.

For the number of hidden layers, the best results were with 3 hidden layers, after that were with 4 hidden layers, 1 hidden layer and the worst results were with 2 hidden layers. The values were very similar between both the models with 1 and 2 hidden layers and the models with 3 and 4 hidden layers. With this we concluded that having more hidden layers contributes to a better accuracy of the model.

For the regularization testing, we verified that both L1 and L2 regularizations had the Train accuracy much higher than the Validation accuracy. With this we verified that these two models were in Overfitting and so we concluded that both L1 and L2 regularizations tend to Overfit the model. With this we concluded that the best results were without regularization, and so we concluded that we shouldn't apply regularization in Feed-Forward Neural Networks models.

Convolutional Neural Networks:

For the Convolutional Neural Networks, we decided that the most adequate number of filters to use was 8. Having 16 filters gave us an higher accuracy, but only a difference of 0.0002 didn't make it worth using more filters.

The same happened to the number of kernels, since all of them had similar results, we decided to use the lowest number of kernels tested, that was 2 kernels.

The model without pooling gave us the best result, although they had both very similar results.

The model with 4 convolutional layers gave us the highest accuracy, with 0.9854, and a loss of 0.0471. However, the results from the model with 2 convolutional layers were almost the same, with an accuracy of 0.9852 and a loss of 0.0476, and so we decided to use 2 convolutional layers.

Between the two models, one with a fully connected layer between the convolutional and output layer and other without it, the model without it gave us the best results, and so we decided to not use it.

Since the model without regularization gave us better results than the model with Dropout regularization, we decided that it is better to use a model without regularization in Convolutional Neural Networks.

V. LINK TO THE GOOGLE COLABORATORY CODE

https://colab.research.google.com/drive/1SNXpXhSYcU0y9D29F96IQRIdFPZMLCT1?usp=sharing

REFERENCES

https://colab.research.google.com/drive/1GuOWptT1DEo1qdZgIJ78TLPGDso_JWJb

Pedro Leitão
Machine Learning, Bologna Master Degree in Computer Science and Engineering
Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal
pedro.o.leitao@gmail.com

Eduardo Bártolo
Machine Learning, Bologna Master Degree in Data Science and Engineering
Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal
eduardo.bartolo@gmail.com