Deep Learning Lab

Assignment 2

Name: Aleksandar Petrov

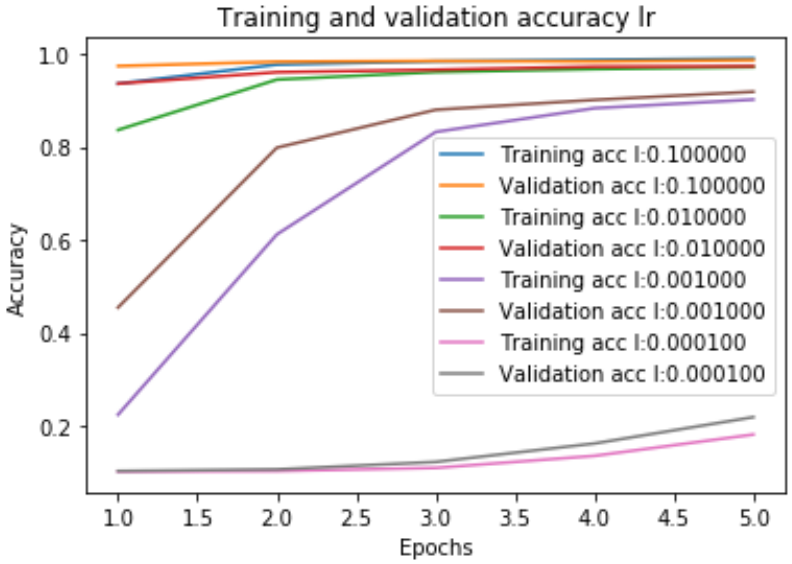
Immatriculation number: 3743542

I had some difficulties coming to terms with the python scripts and understanding how to run those on the pc-pool computers. Sadly someone interrupted my training session on pooltf54 and I couldn't run random search for 50 iterations. However I managed to run it for 10 on my laptop and used those parameters for the final task. I also made a jupyter implementation of the code just because I find it easier to experiment. I've also changed the mnist() function so that it does not output one_hot_enocded vectors but I used the appropriate loss function from keras and it worked fine.

Task 1

We see that the high learning rates (0.1 & 0.01) reach the >90% accuracy relatively fast and the learning curve is flat - meaning that the NN doesn't learn that much in later epochs. The actual learning happens in the first couple of epochs. The low learning rate (0.0001) shows that our network doesn't learn at all.

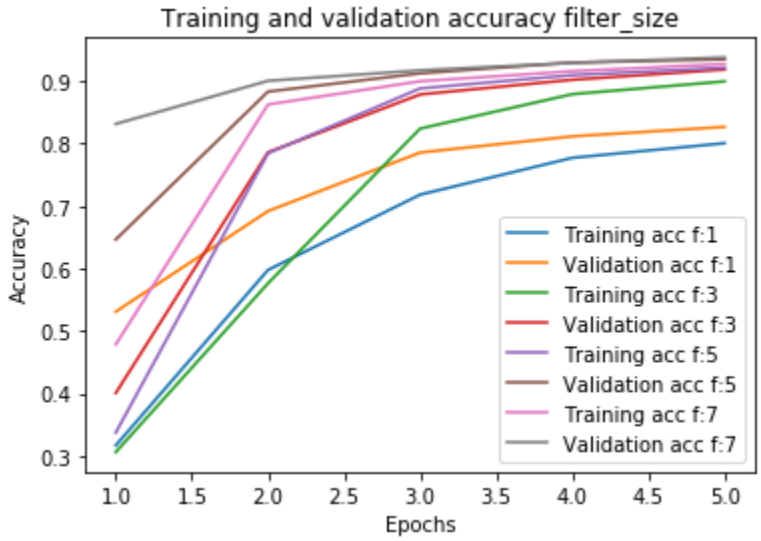
The steps we make during gradient descent are too small for us to reach the minimum. The middle learning rate (0.001) has a nice smooth curve, although it doesn't reach validation accuracy higher than ~92%. Furthermore it doesn't exhibit any overfitting (the validation accuracy is better than the training).



Task 2

Small filter size (1x1) doesn't have enough capacity to learn the features we need in this case. 3x3 filter shows best performance in aspect of training time and validation accuracy. The two bigger filters (5x5, 7x7) need a lot more time for training and do not show a significantly better performance.

My guess would be that bigger filters would make sense for bigger images, and vice versa for small filters. However we see that too small of a filter or too big can be equally as bad but in a different way.



Task 3

The result of my random search is:

```
Best found configuration: {'num_filters': 22, 'learning_rate': 0.037206934
81704873, 'filter_size': (5, 5), 'batch_size': 95}
```

I find it strange that there is a batch size of 95 and a number of filters 22 those aren't logarithmically distributed between the boundaries I set up and I also set the log=True label. So I would appreciate if you tell me where I did wrong in my code. The test error I achieved is 1.45 %.