DL Report Exercise 2

Edward Kreutzarek

November 2018

1 Architecture

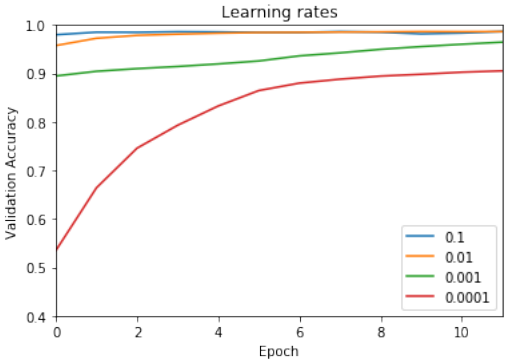
For this Convolutional Neural Network two convolutional layers were used. Both were followed by a ReLu layer and a Max-Pooling layer

1. Input Layer [50000,28,28,1]
2. Convolutional Layer (16 3 3 filters, stride = 1) + ReLu + Max Pooling (pool size = 2)
3. Convolutional Layer (16 3 3 filters, stride = 1) + ReLU + Max Pooling (pool size = 2)
4. Fully Connected Layer (128 Units)
5. Fully Connected Layer (10 Units)

Loss function: Cross Entropy

2 Learning Rates

As one can see in the picture, there is not really much difference between the learning rates 0.1, 0.01 and 0.001. 0.1 is the best learning rate and 0.0001 is by far the worst. 0.0001 is a too small learning rate so it takes too long to even reach a somewhat good accuracy.

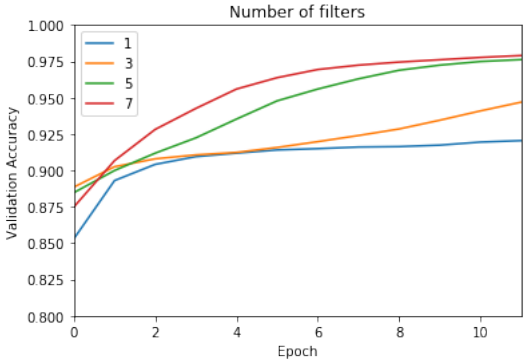


3 Number of filters

The number of filters had a more obvious effect on the validation accuracy. Choosing only one filter led by far to the worst result.

The higher the number of filters the better the result for the validation accuracy got. However, as the plot shows, the accuracy seems to start plateauing.

It seems to be preferred to choose smaller filters, but have bigger number of those, although i think that it's just a hyperparameter that needs to be tuned specifically for the model.



4 Difficulties

The most difficult thing was to get the whole stuff running without getting an memory error or something along the lines of '..when allocating'.

At first i tried to build my CNN by using methods on the tensorflow tutorial. But i got nowhere and was confused by stuff like tf.estimator or estimatorSpec

At first i used tf.nn.conv2d instead of tf.layers.conv2d. Then i got an error when i tried to pass the filtersize as an array directly into the parameters, so i had to initialise the filters.

I know, that i should have used an extra test function, but i found out too late how to properly save and restore my model. For the next time, i definitely will keep this in mind.

5 Random Search

*possible values for the filter size were [3,4,5,6,7], like in Exercise 2, as opposed to just 3 and 5.

Best found configuration:
filter size: 6, batch size: 17, num filters: 33, learning rate: 0.012480532457181262

Error: 0.99029