Deep Learning Lab WSE 2018 Exercise 1

Assigment: Implementing a CNN in Tensorflow

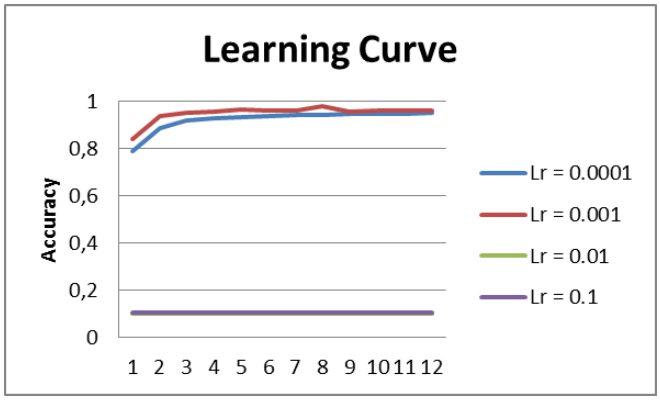
Implementation: Implement a CNN consisting of two convolutional layers

- Number of filters: 16
- Filter size: 3x3
- Each followed by ReLU activation function and a max pooling layer
- After convolution layers add fully connected layer with 128 units and softmax layer for classification
- Train the network by optimizing the cross-entropy loss with stochastic gradient descent
- Save validation performance after each epoch
- Perform experiments changing Learning Rate:[.1, .01, .001, .0001]
- Perform experiments changing Filters size:[1, 3, 5, 7]
- Implement random search to automatically tune hyper parameters.

Results: The following parameters were obtained with the iteration of the parameters listed above.

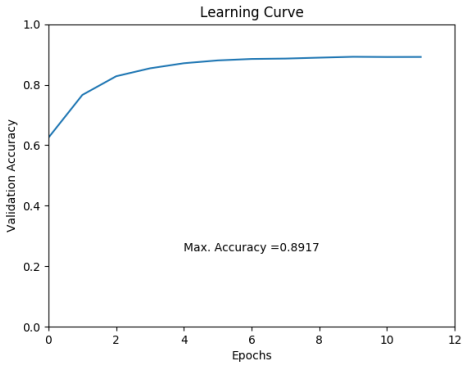
Variable conditions: ***Learning Rate***

Fixed conditions: Neural network of two convolutional layers, Filter Size, Number of Filters
ReLu as activation function, Stochastic Gradient Descent, Softmax
Batch size 64units

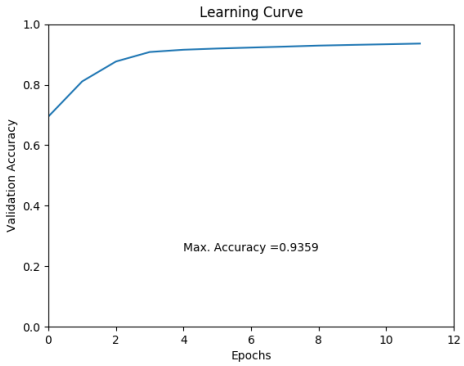


Variable conditions: ***Filter Size***

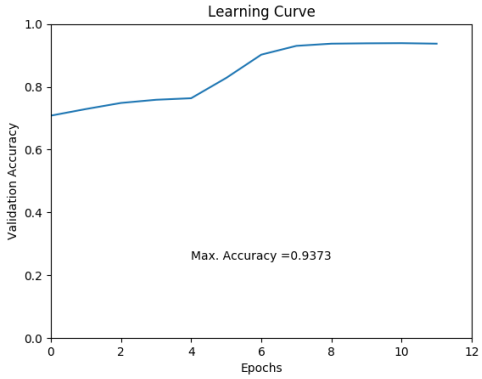
Fixed conditions: Neural network of two convolutional layers, Number of Filters, Learning rate
ReLu as activation function, Stochastic Gradient Descent, Softmax
Batch size 64units



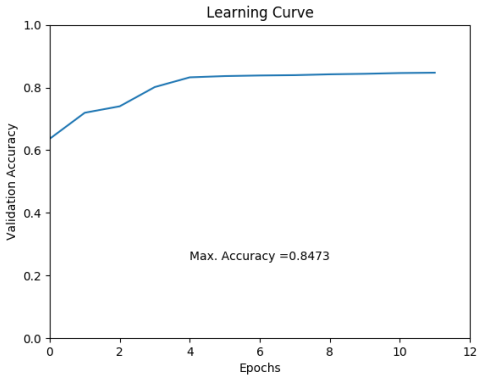
Filter size: 1x1



Filter size: 3x3



Filter size: 5x5



Filter size: 7x7

Conclusion:

Learning Rate

- Which conclusions could be drawn from these figures?

- There is a huge dependency in the performance of the network regarding to this parameter

Which value for the learning rate works best?

-The performance of the network resulted similar and at its best with 0.001 and 0.0001 learning rates

- Do you have an idea what is happening if the learning rate is too small / high?

-If the learning rate is too big doesn't learn properly, but we have an improvement in the processing time.

If the learning rate is too small we get a better improvement on the learning accuracy, but we have a considerable increase in processing time cost.

Filter Size

- Do you have an idea in which scenarios smaller filters might be better than larger filters and vice versa?

- Since we are working with image recognition problem, when a big amount of pixels are necessary for the network recognize the object it's better to use large filters.

If what differentiates objects are some small and local features is better to use smaller filters.

Feedback:

Really confusing figure it out how to run random search to automatically tune hyper parameters.
Computer always ran out of memory.
Implementation/Solution of the code was over 30hours.