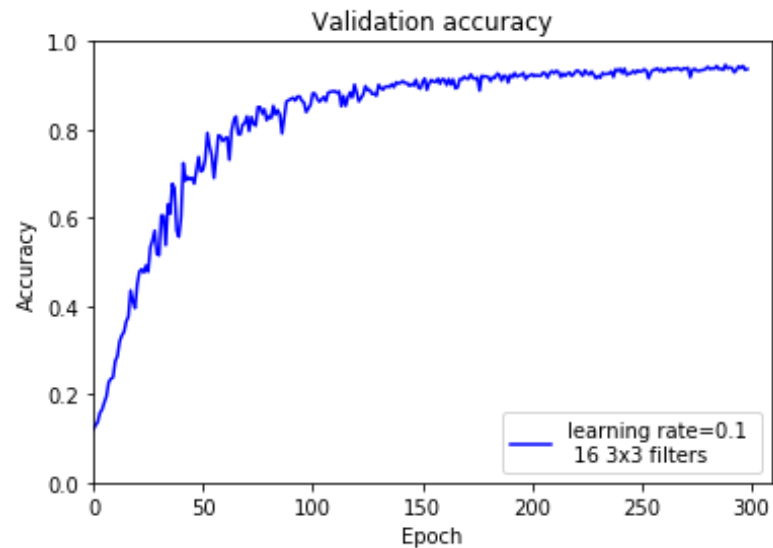
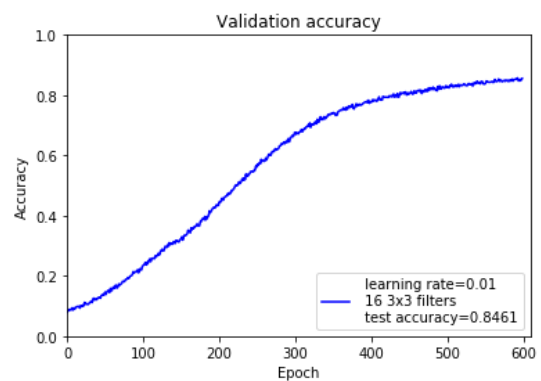
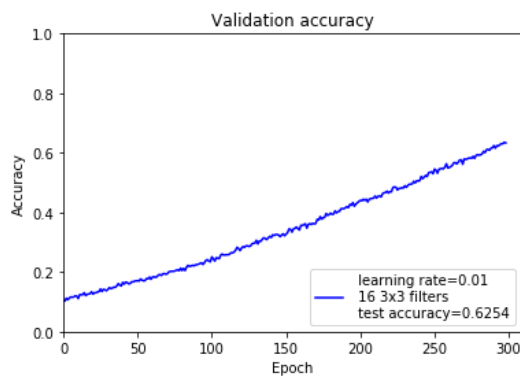


RESULTS

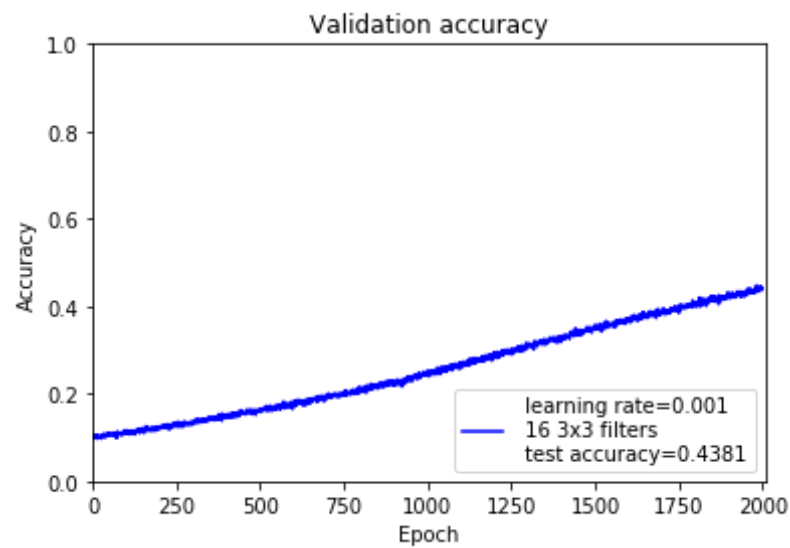
1. Converges at ~200 epochs



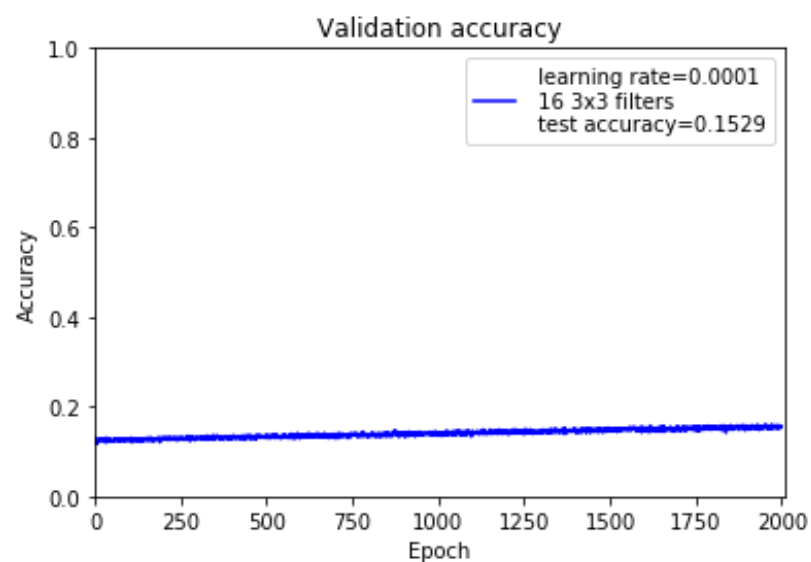
2. Learns very slow- requires more no. of epoch compared to 1. Better accuracy ~600epochs



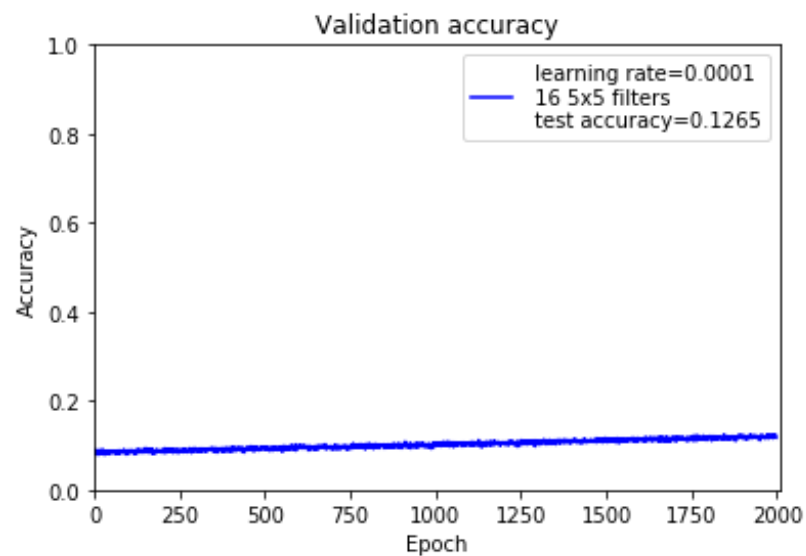
3. Learns very slow- requires more no. of epoch compared to 2. Poor accuracy even after 2000 epochs



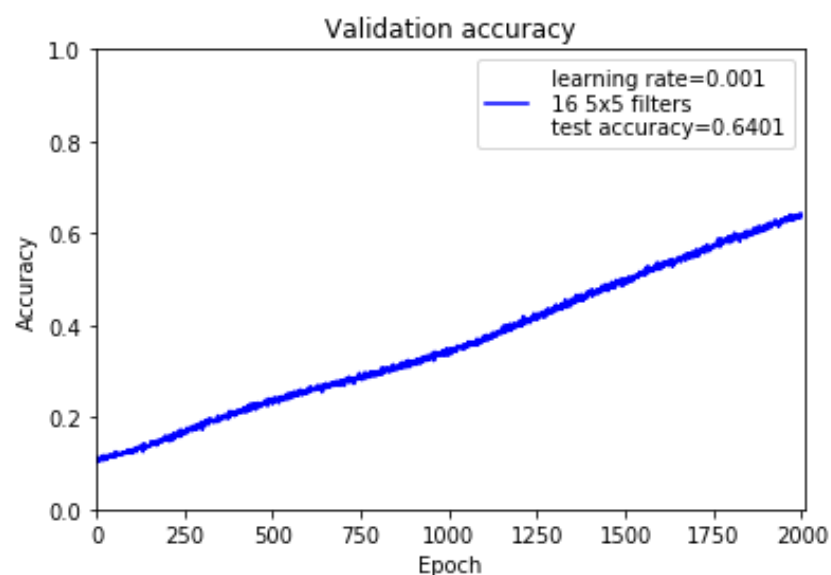
4. Learns very slow- Poor accuracy even after 2000 epochs(vanishing gradient problem)(no learning at all)



5. Learns very slow- Poor accuracy even after 2000 epochs(vanishing gradient problem)(no learning at all)

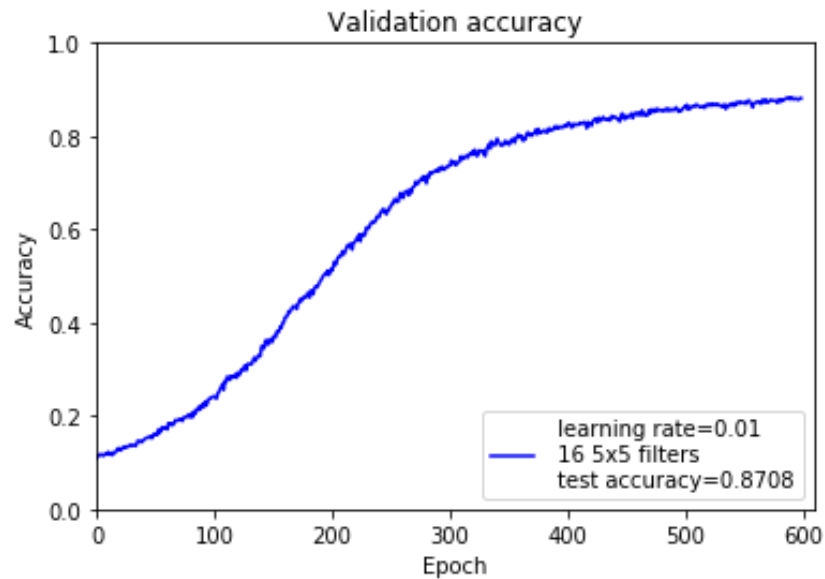


6. Better results than learning rate = 0.0001 for same config of filters.

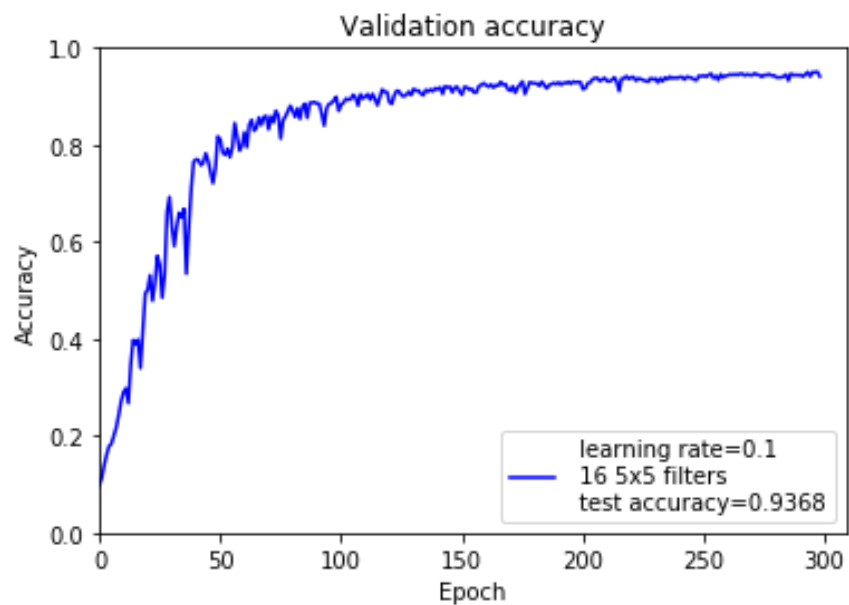


Conclusion: We do not see any learning at all, for learning rate=0.0001

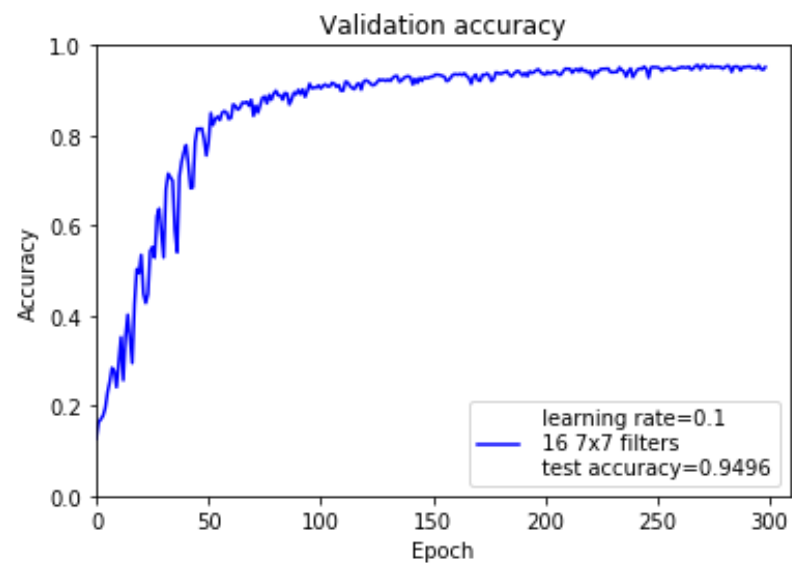
7. Better results than learning rate = 0.001 for same config of filters. And requires less epochs than 6. For convergence



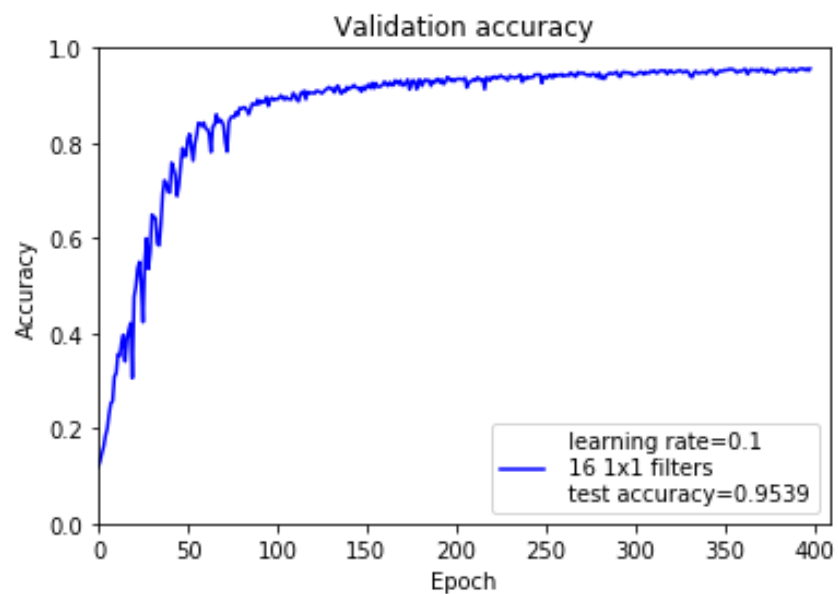
8. Converges very early ~170epochs



9. Converges very early ~150epochs



10. Converges very early ~250epochs



Conclusion: As the filter size increases for a given learning rate and other fixed hyperparameters, convergence time(no. of epochs) is reduced

Observation: As the filter size increases, the run time for training reduces.(Due to less number of computation required for dot product)

Random Search:

Best found configuration:

batch_size: 57

filter_size: 5

learning_rate: 0.001526956778859322

num_filters: 25

Test Accuracy: 0.989 => Test Error: 0.011

