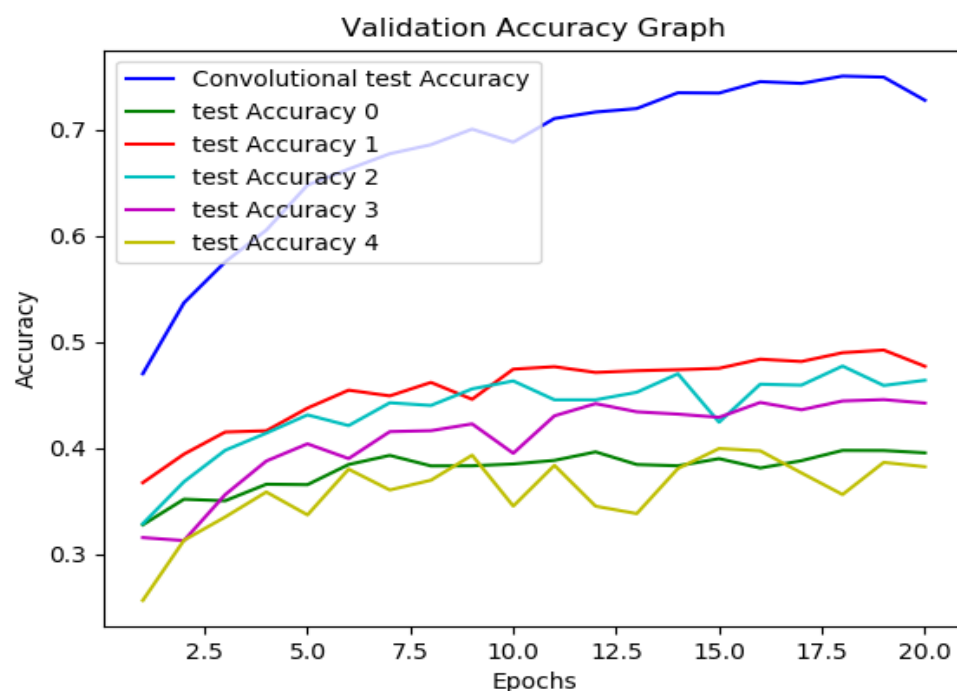
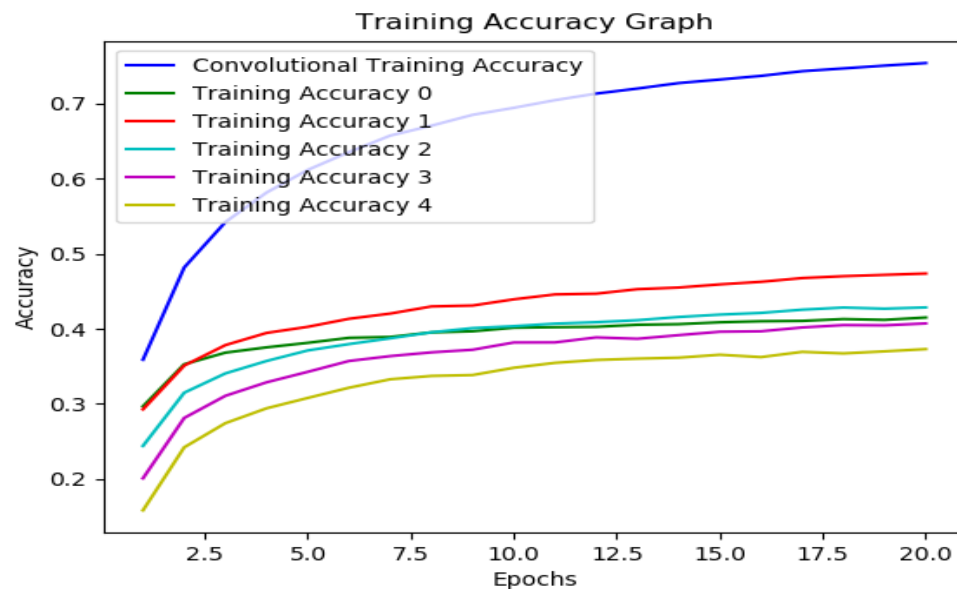


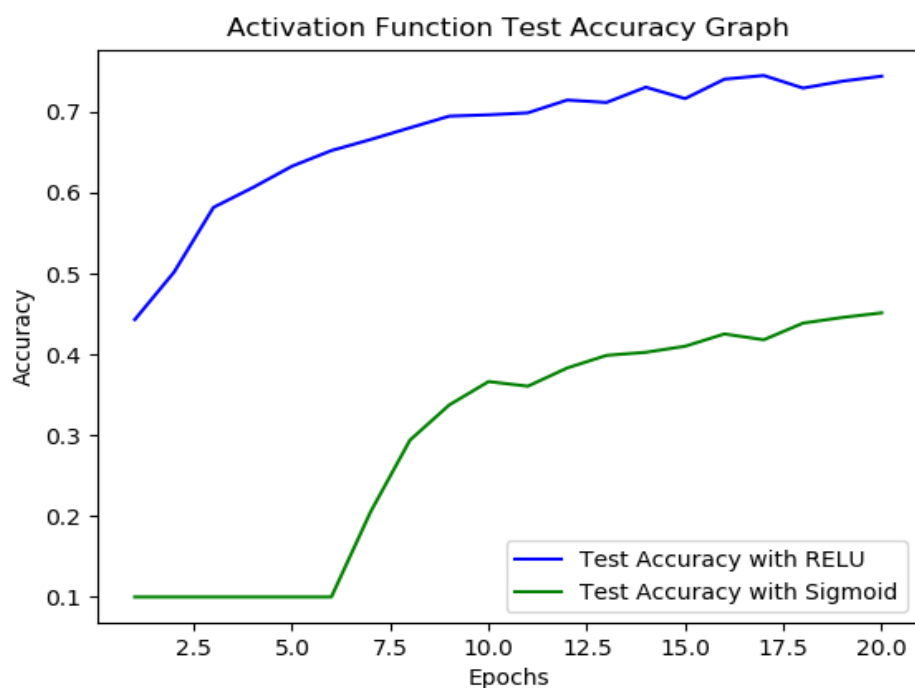
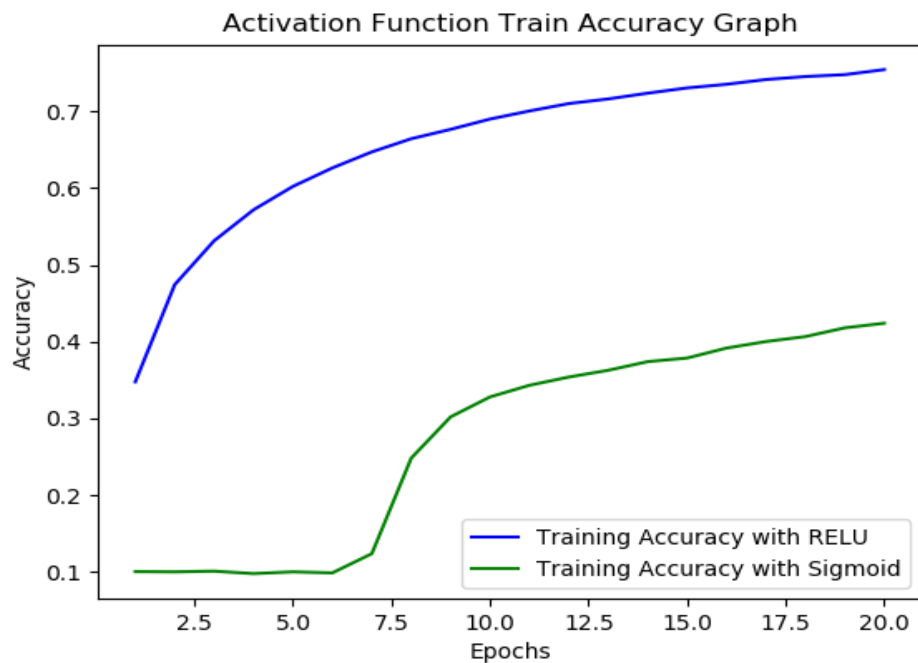
Question 1:

Part A:



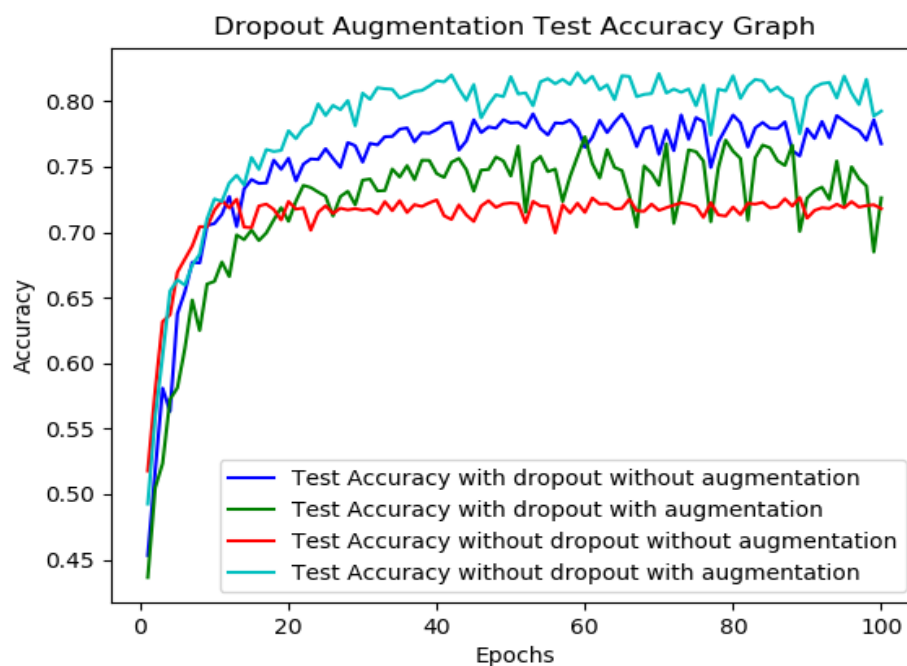
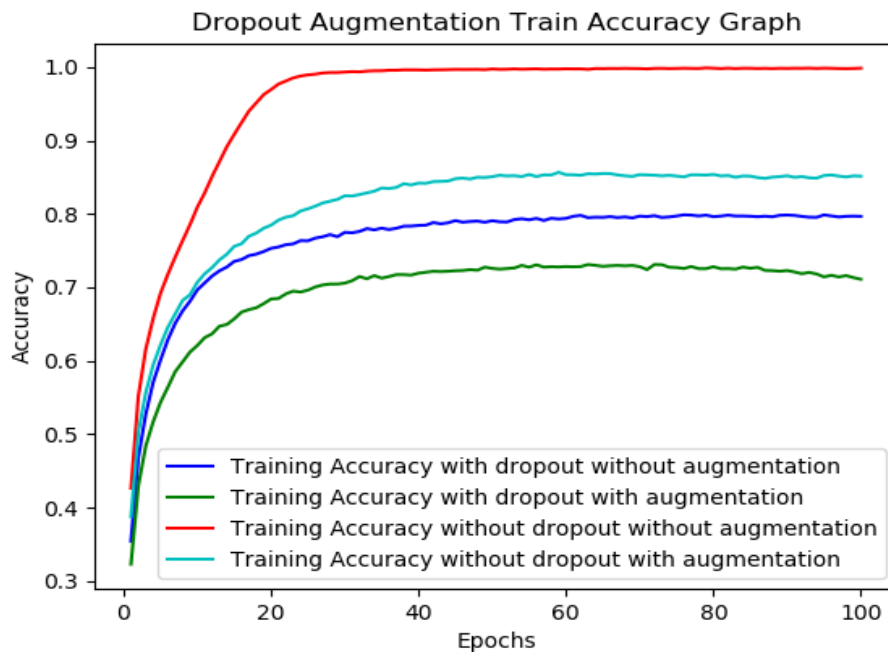
As we can see in the above graphs convolution layer always perform better than dense layers, this is because convolution layers are better feature extractor than dense layer since dense layer disregards spatial relationship.

Part B:



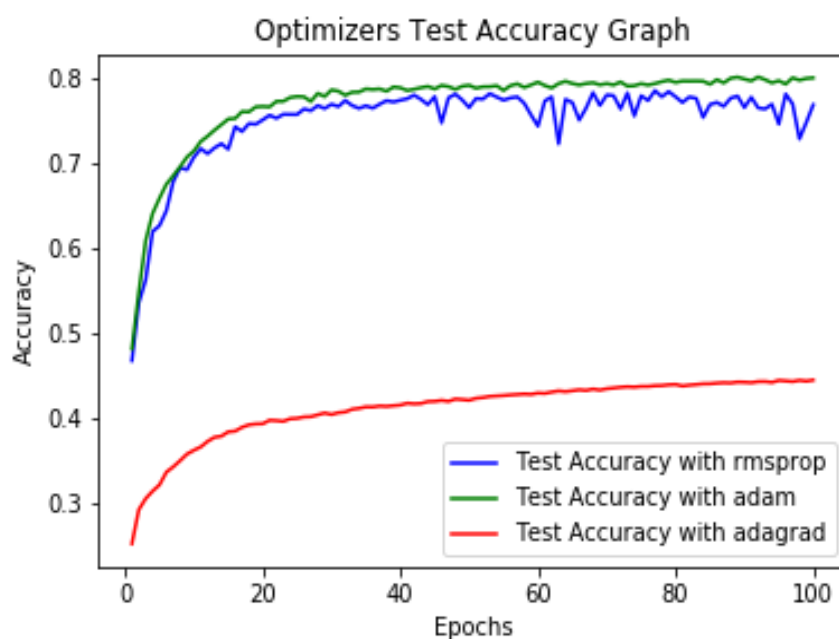
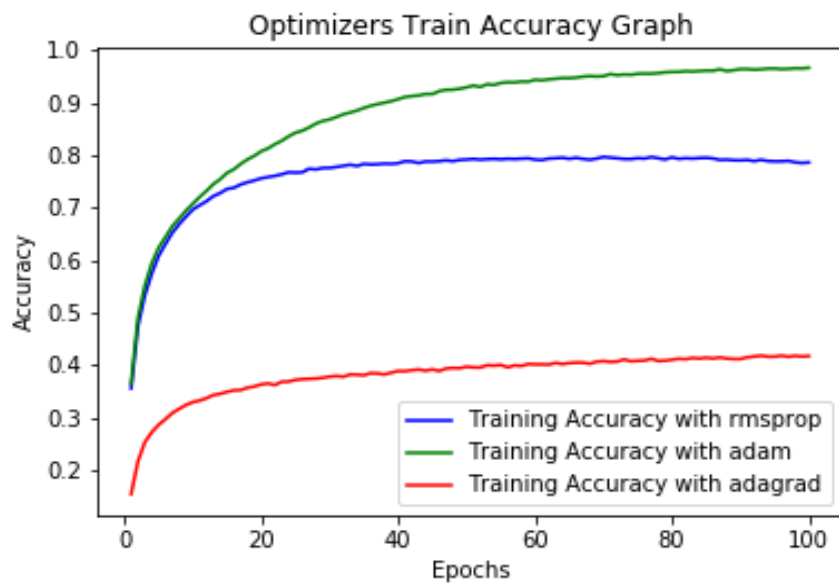
As we can see in the above graph with sigmoid activation function, it is constant for some regions because it is having vanishing gradient problem, while ReLU performs better because it does not have vanishing gradient problem and updates weights faster. Also ReLU produces sparse representation which is beneficial than dense representation produced by Sigmoid.

Part C:



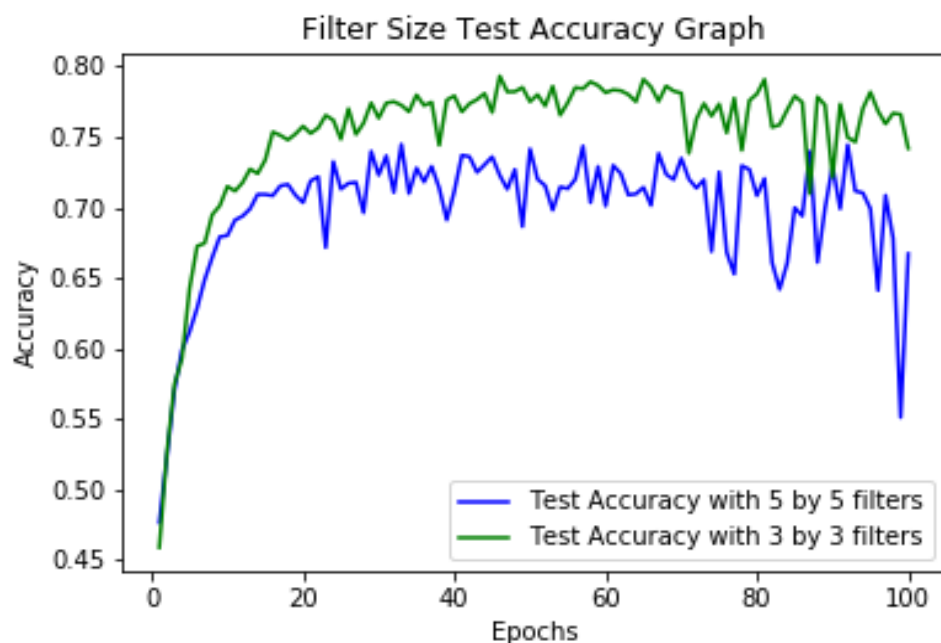
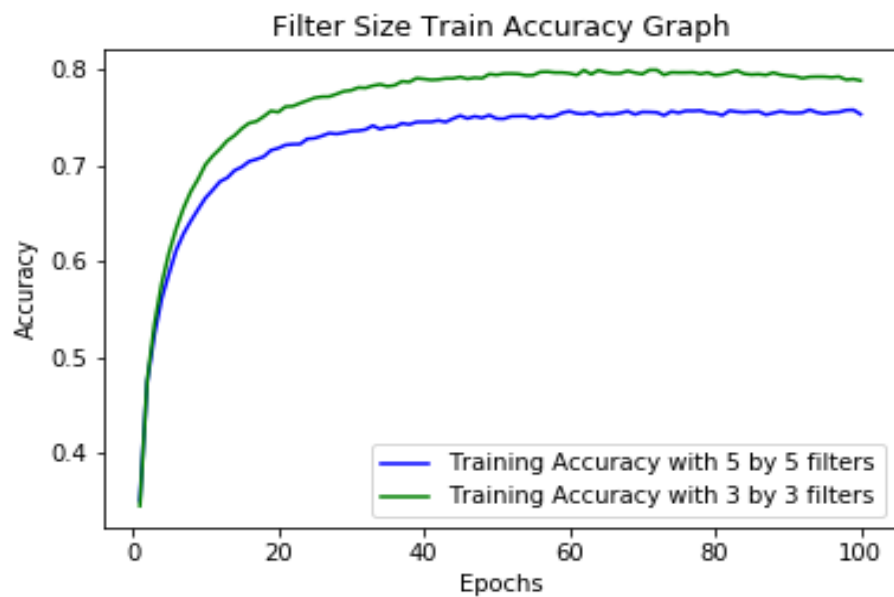
Dropout and augmentation both are regularization techniques, as we can see in the above graphs it is clearly visible that over fitting occurs when we do not use dropout and regularization. Also Augmentation without dropout is more powerful because it exposes the model to more variation of data which can be seen in test time as it generalizes better. When we use both augmentation and dropout, model is under fitting as can be seen with the graph.

Part D:



RMSProp is an extension of Adagrad that deals with its radically diminishing learning rate, as we can see in the above graphs it performs better than Adagrad. Also Adam slightly outperforms RMSProp because of its Bias correction and even Adam is a combination of RMSProp and momentum optimizers.

Part E:



With two 3X3 kernels we will have more layers than one 5X5 kernel. Which makes two 3X3 kernels learn complex and more non-linear features than one 5X5 kernels. That's why two 3X3 filter performs better than one 5X5 filter. Even one 5X5 filter is computationally expensive than two 3X3 filters.