CNN

We used Pytorch for implementation for the network. There were 2 convolutional layers with 64 feature maps each. After this we flatten the output feature map from the last convolutional layer, we then have a dense layer of 256 neuron and finally and dense layer with 10 neurons, giving the output. We used cross entropy as our loss function.

We use 3 optimization algorithm: SGD, RMSprop, Adam; and in each algorithm we used Dropout and weight regularization.

For weight regularization we use weight decay in the optimization algorithm. Since in regularization we want to minimize $J(W)+0.5\lambda\|W\|^2$, for this our weight update will be

$$W^{(t+1)}=W^{(t)}-\eta\frac{\delta J}{\delta W}-\lambda W^{(t)}$$

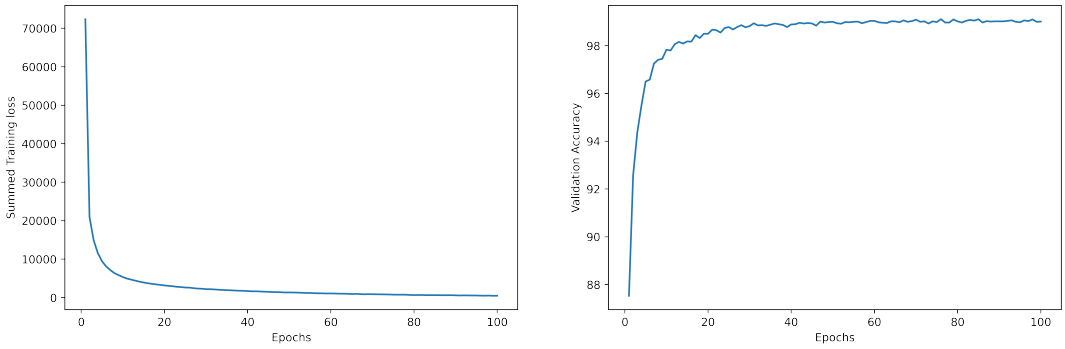
$$W^{(t+1)}=(1-\lambda)W^{(t)}-\eta\frac{\delta J}{\delta W}$$

and this is similar to weight decay $W^{(t+1)}=(1-\lambda)W^{(t)}$

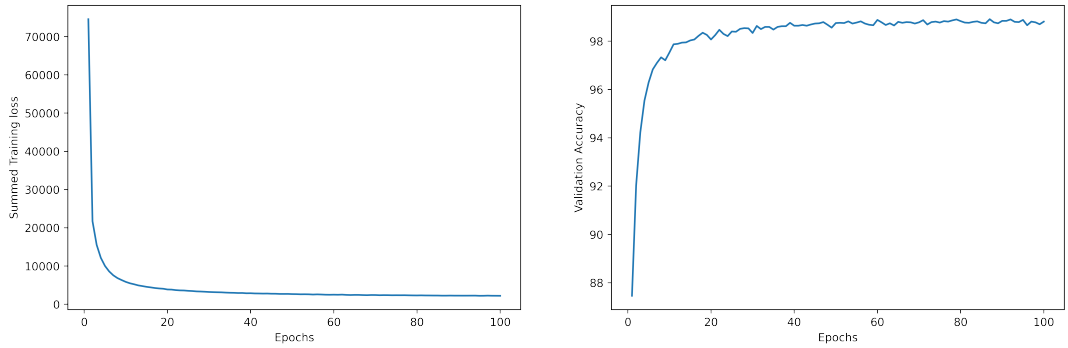
There were 12 cases as follows (entries of table represent the maximum accuracy obtained in each case):

	SGD	**RMSprop**	**Adam**
Without Dropout and Without Regularization	99.11%	99.37%	99.32%
With Dropout and without regularization	99.06%	99.27%	99.32%
Without Dropout and with regularization	98.91%	98.99%	99.01%
With dropout and with regularization	98.96%	99.0%	98.95%

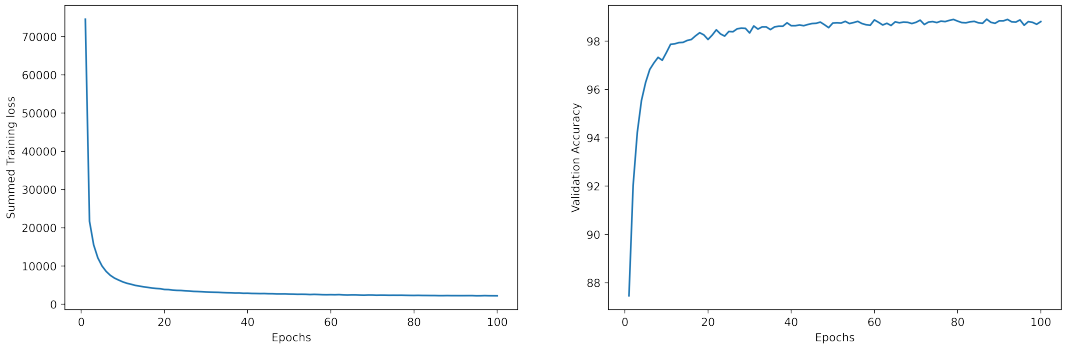
Here are graphs for loss function on training data or Validation curacy for each of the case.



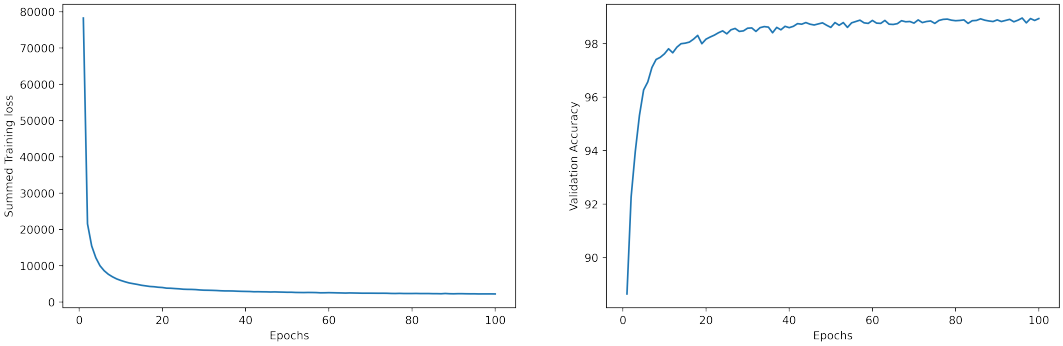
SGD without Dropout and without regularization



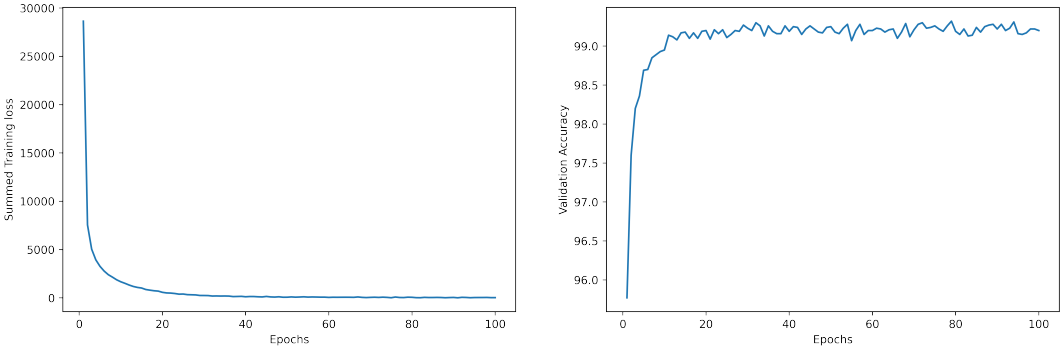
SGD with Regularization only



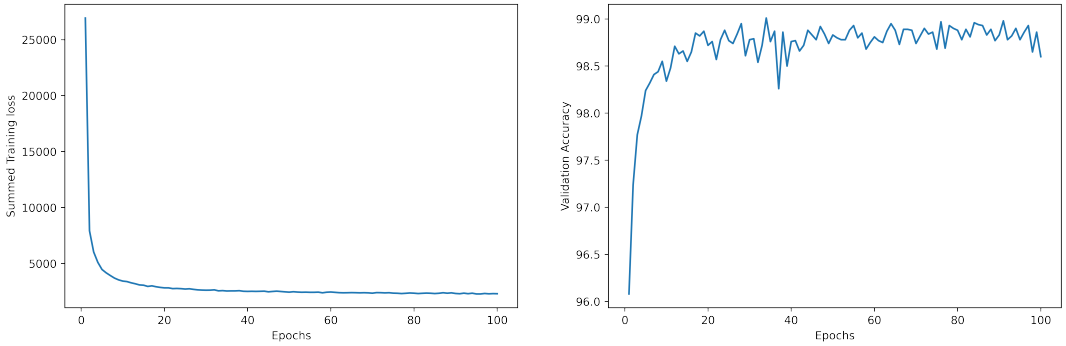
SGD with Dropout only



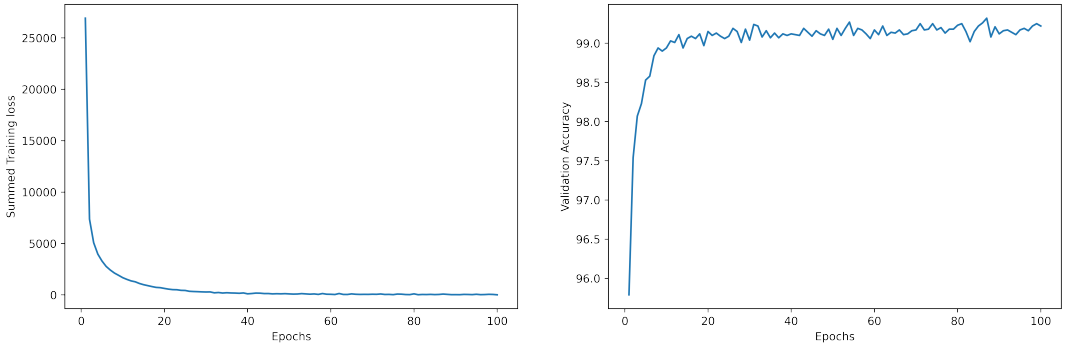
SGD with Regularization and Dropout



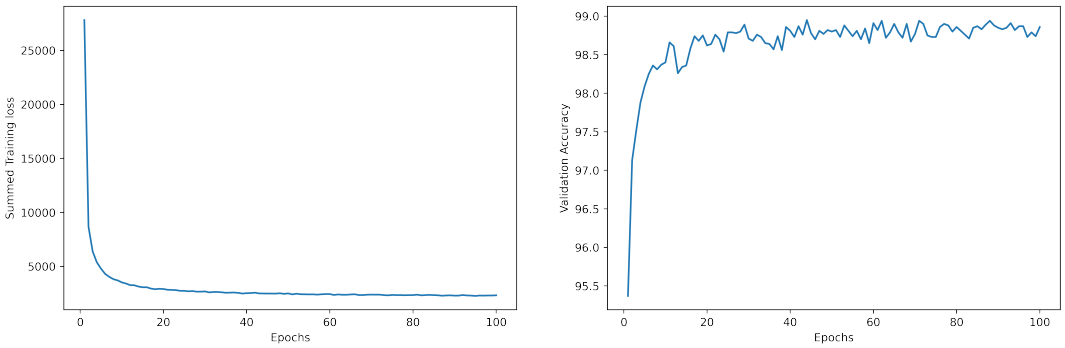
Adam without Regularization and without Dropout



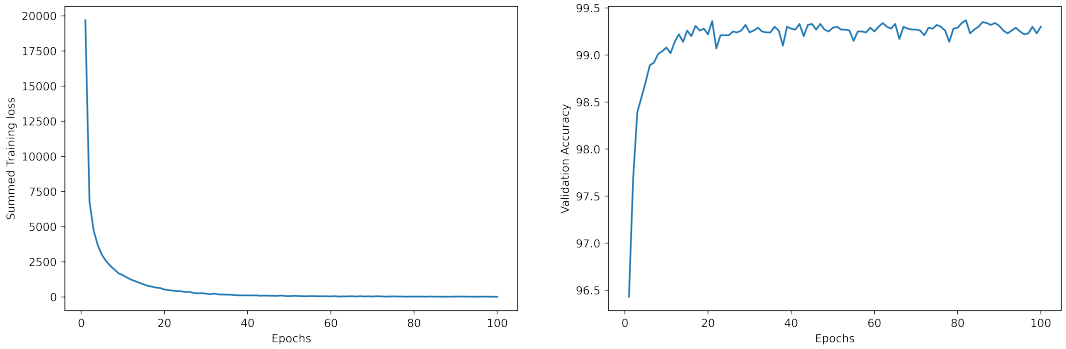
Adam with Regularization



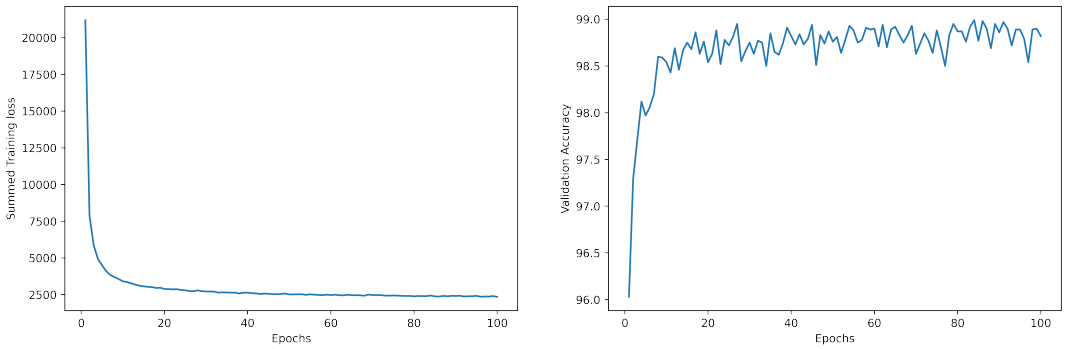
Adam with Dropout



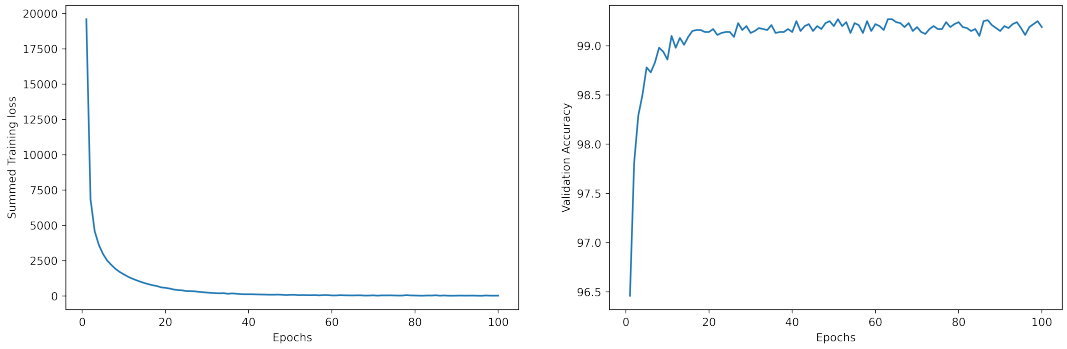
Adam with Dropout and Regularization



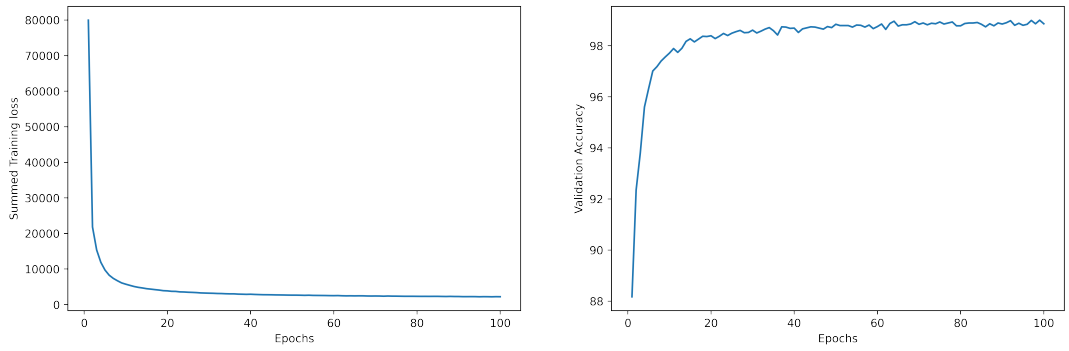
RMSprop without regularization and without dropout



RMSprop with Regularization

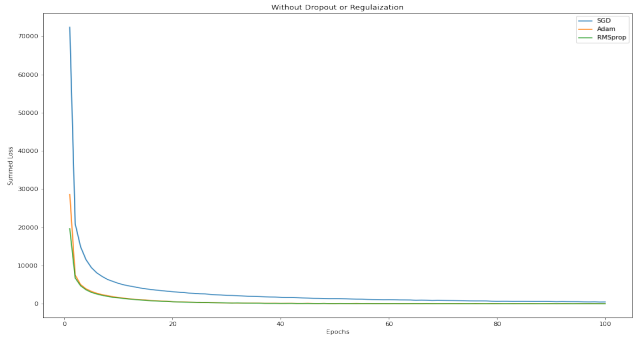


RMSprop with Drop-out

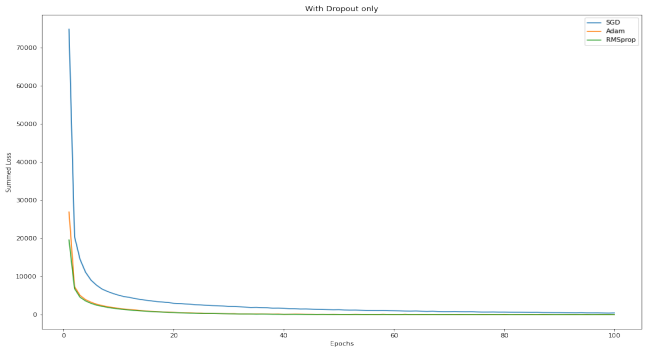


RMSprop with Dropout and Regularization

We also compared each of the Optimization algorithm, and how are they converging, the graph shows the comparison:

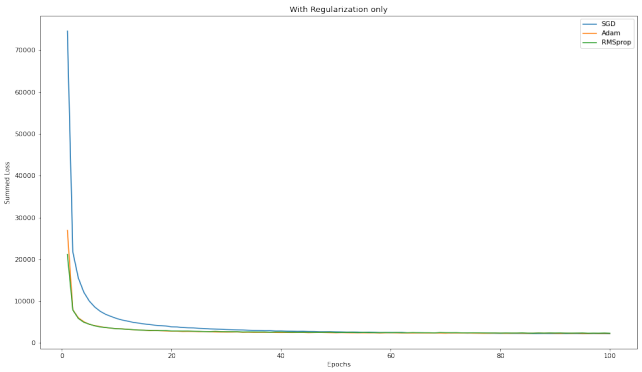


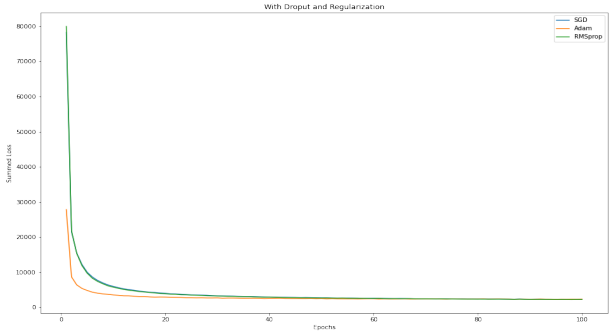
Without Dropout and Without Regularization



With Dropout only

With Regularization only





With Dropout and with Regularization

It was found out that Adam was converging faster than RMSprop and SGD, and SGD was the slowest to converge.

The Accuracy depends upon the initialization points which may vary whenever we try to run the algorithm.