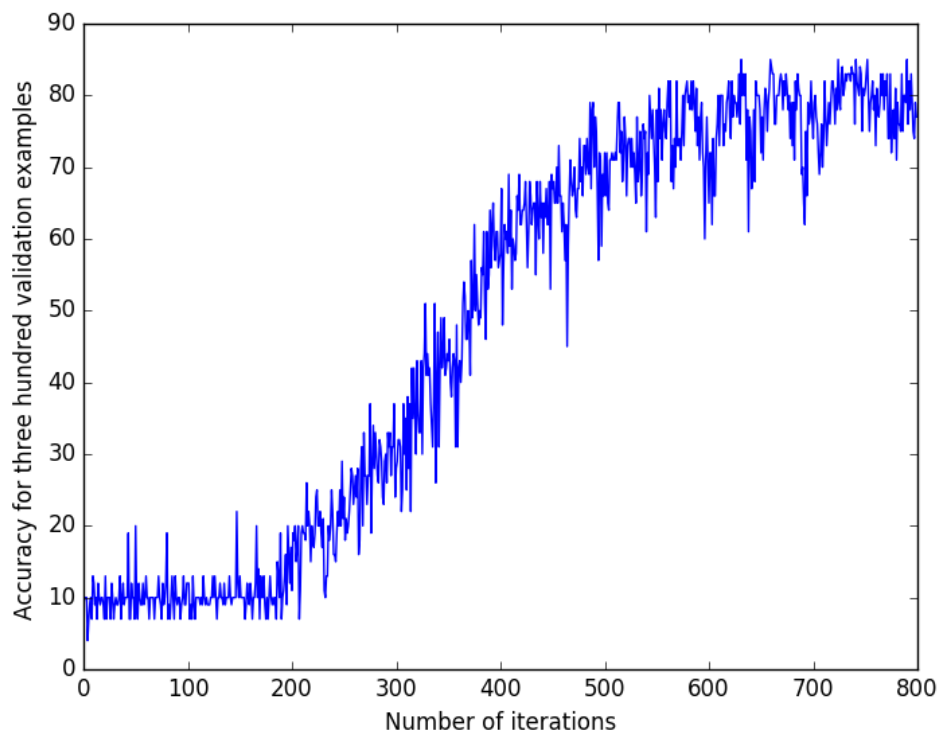
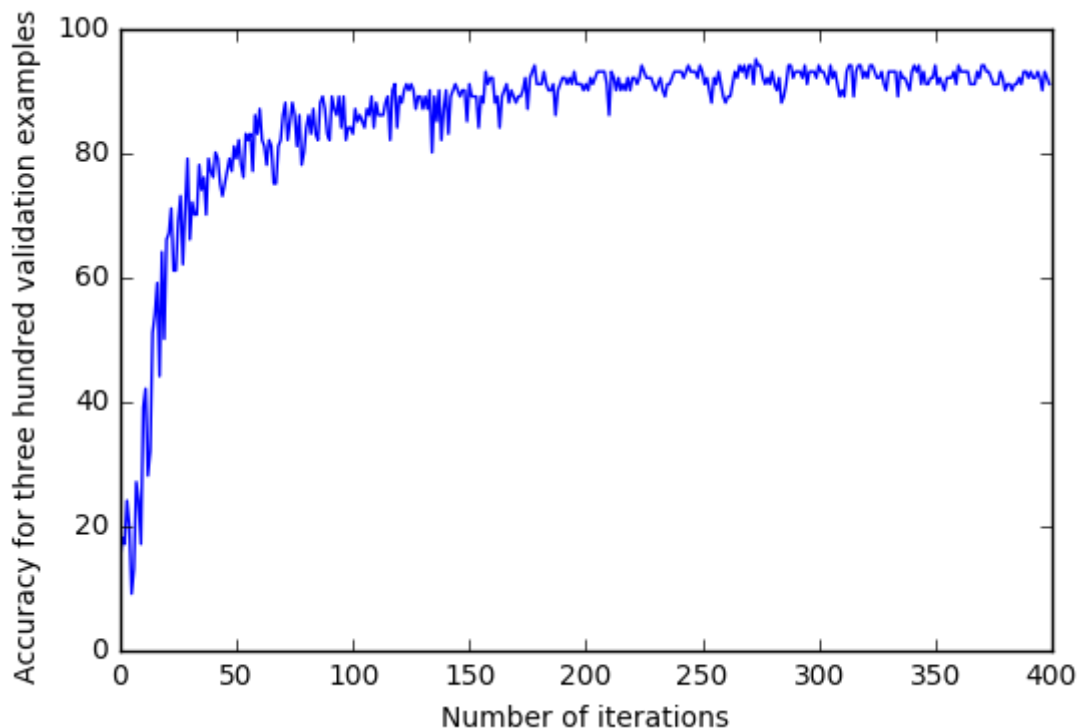


Validation Accuracy

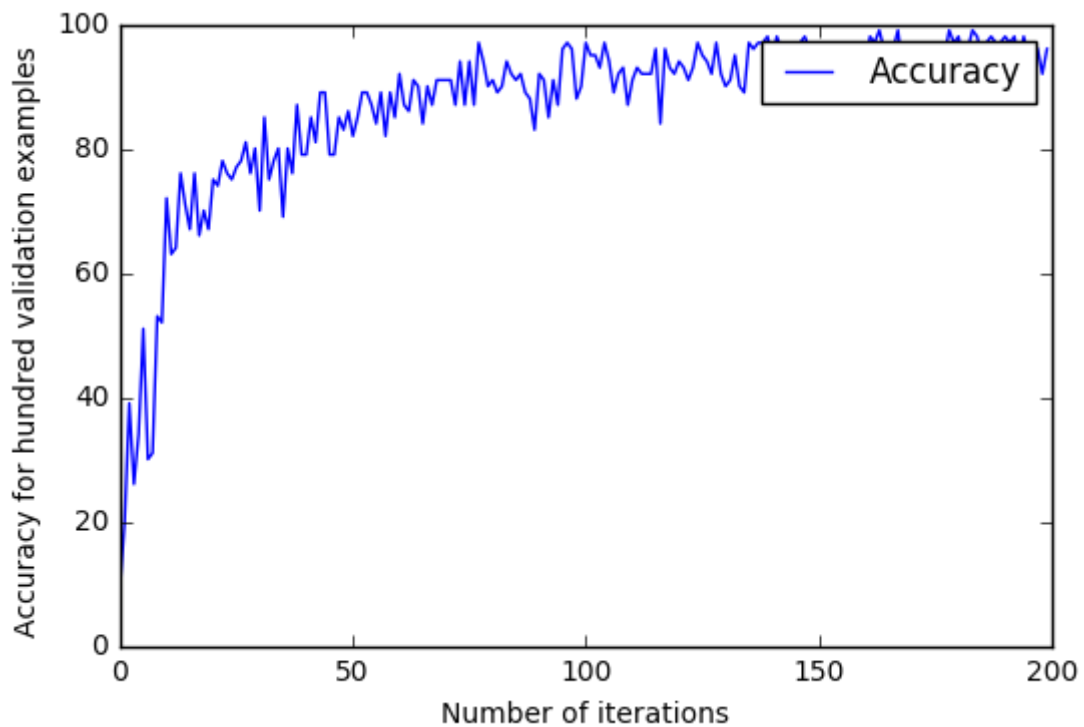
1. Using 6400 examples with a mini batch of 16 using a learning rate of 0.1 and validation on 300 examples. Accuracy $\geq 86\%$



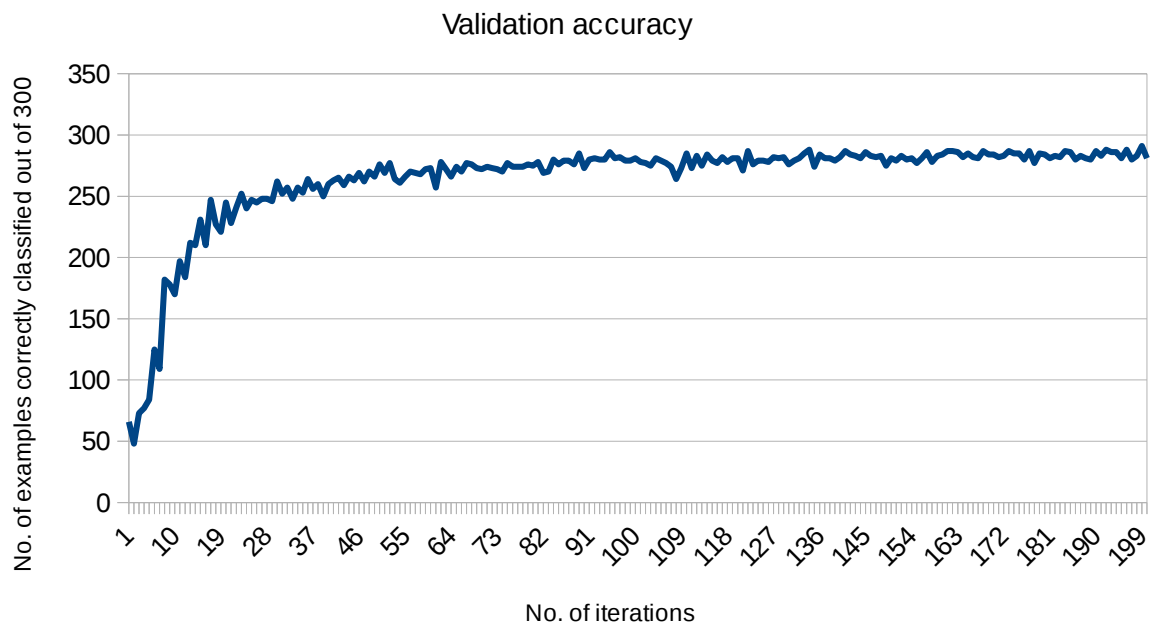
2. Using 6400 examples with a mini batch of 32 using a learning rate of 0.1 and validation on 300 examples. Accuracy $\geq 93\%$



3. Using 6400 examples with a mini batch of 64 using a learning rate of 0.1 and validation on 100 examples. Accuracy $\geq 95\%$

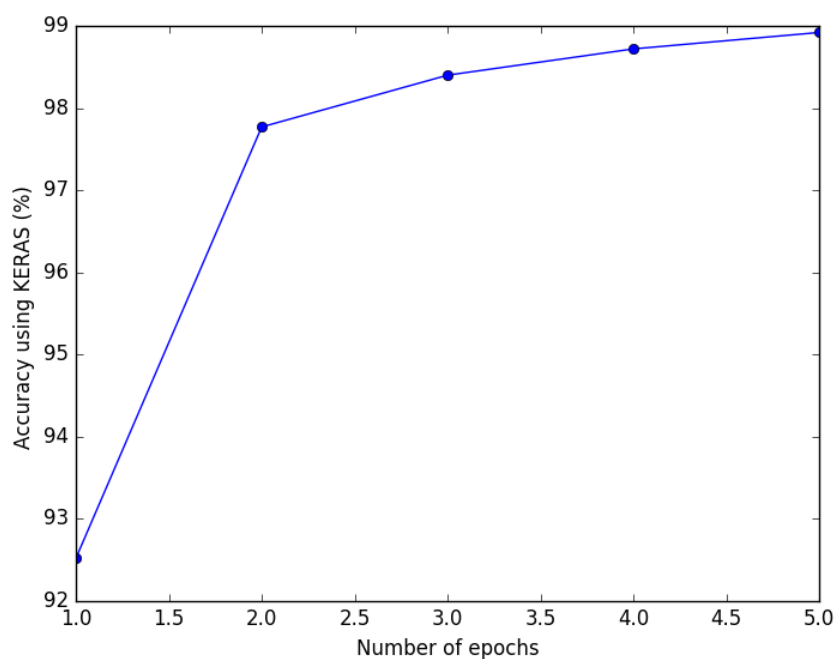
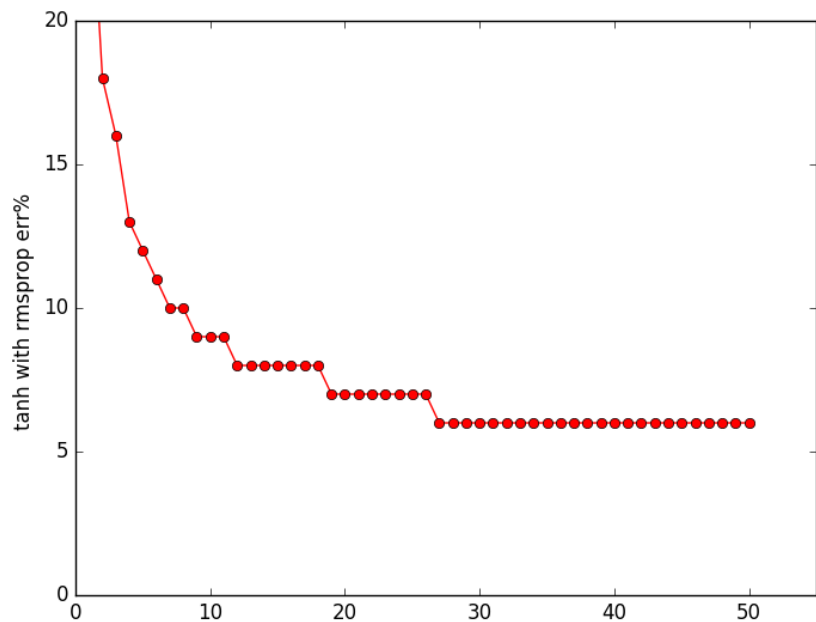


4. Using 12800 examples with a mini batch of 128 using a learning rate of 0.2 and validation on 300 examples. Accuracy $\geq 96\%$



This is in accordance to what we know about stochastic gradient descent. Larger mini batches allow us to reduce the variance of the stochastic updates and converges quicker

Using MLP [784,30,10] on the entire training data and many epochs(50) gives us the same accuracy as LeNet-5 with very small number of training examples and total iterations. We can clearly see the superiority of LeNet-5 over the small MLP.

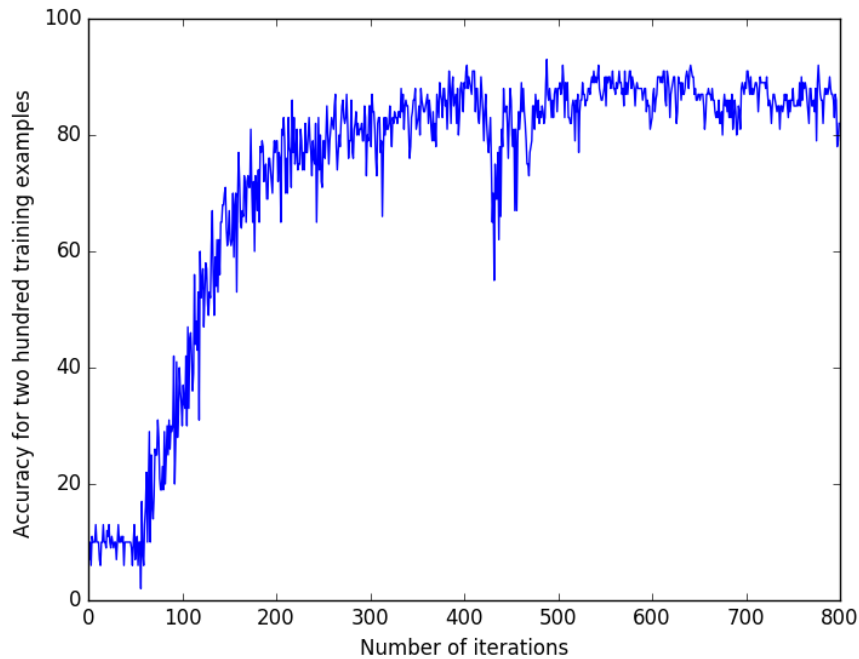


Accuracy using the Keras model.

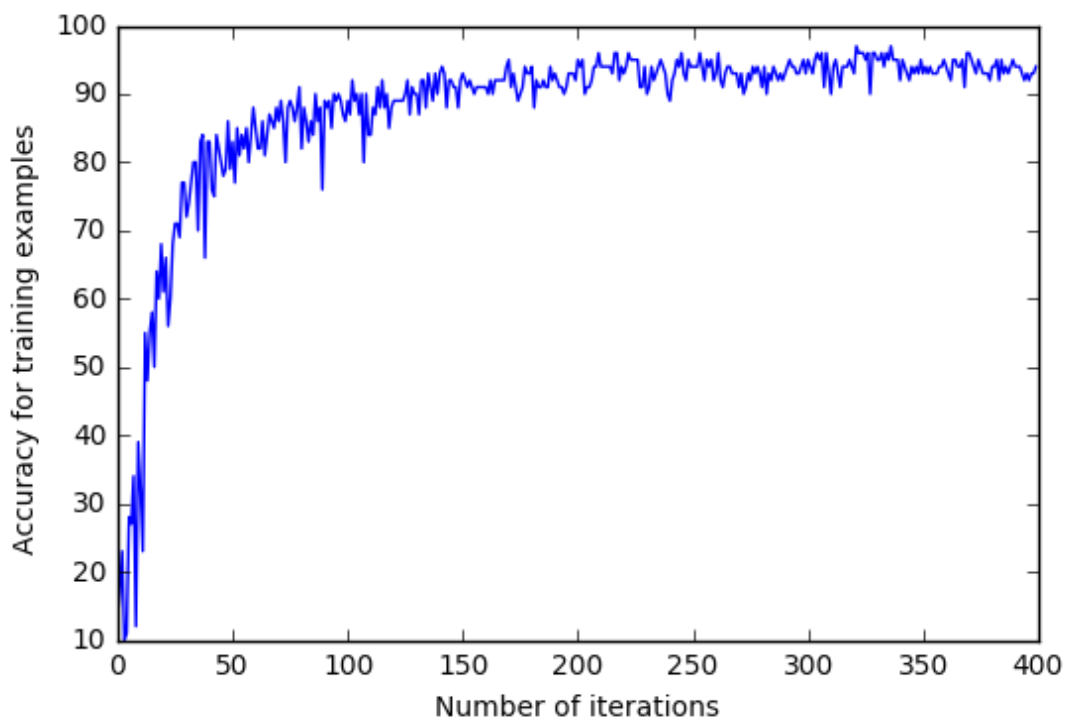
Note: It's trained on all the training examples and accuracy is on the test set.

Training Accuracy

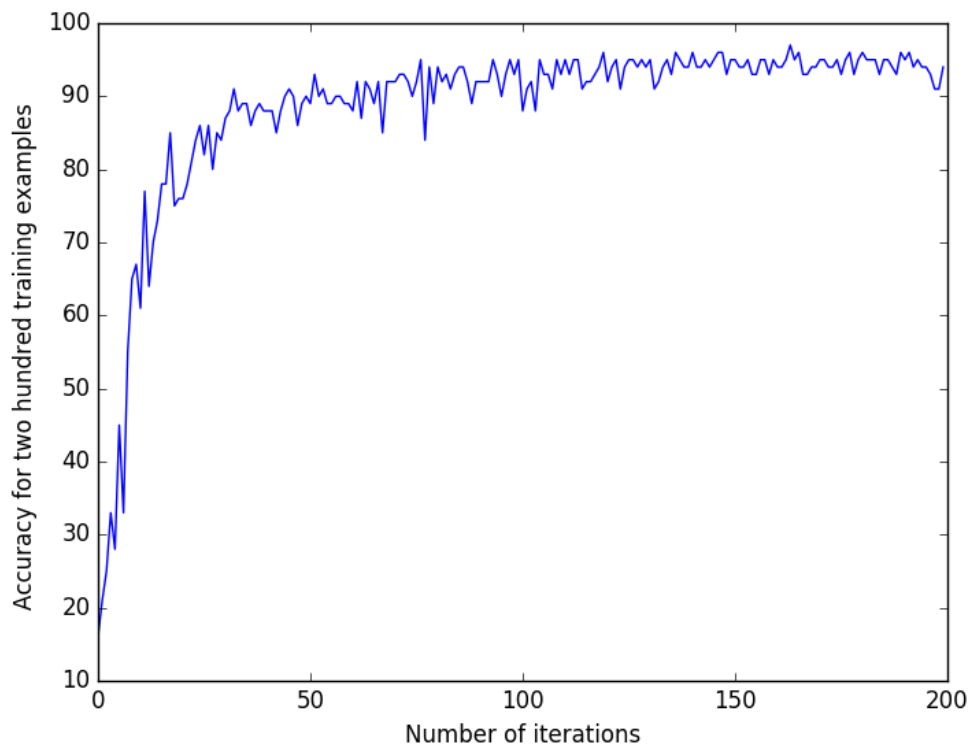
1. Using 6400 examples with a mini batch of 16 using a learning rate of 0.1 and validation on 200 examples. Accuracy \geq



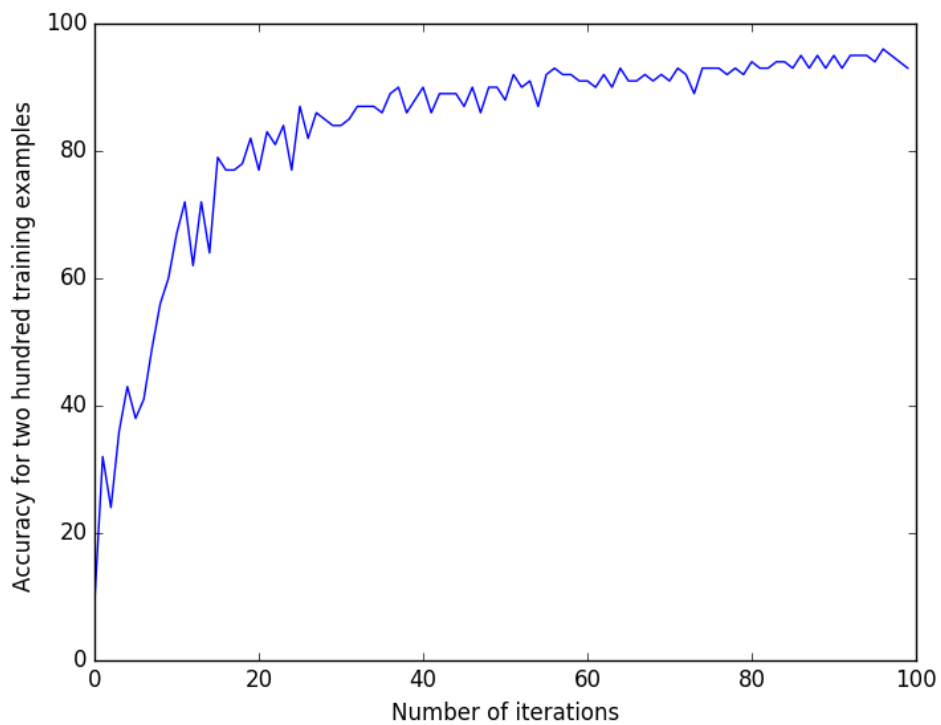
2. Using 6400 examples with a mini batch of 32 using a learning rate of 0.1 and validation on 200 examples. Accuracy \geq 95%



3. Using 6400 examples with a mini batch of 64 using a learning rate of 0.1 and validation on 200 examples. Accuracy $\geq 95\%$



4. Using 6400 examples with a mini batch of 128 using a learning rate of 0.1 and validation on 200 examples. Accuracy $\geq 97\%$



Number of Parameters in the Convolution and Fully Connected Layers

Params in CONV : $6*1*5*5 + 6 + 16*6*5*5 + 16 = 2572$

Params in FC : $400*120 + 120*84 + 84*10 = 58920$

Time taken by Convolution and FC layers

Forward Pass(2048 IMAGES) :

Convolution layers : 0.0707539905561s

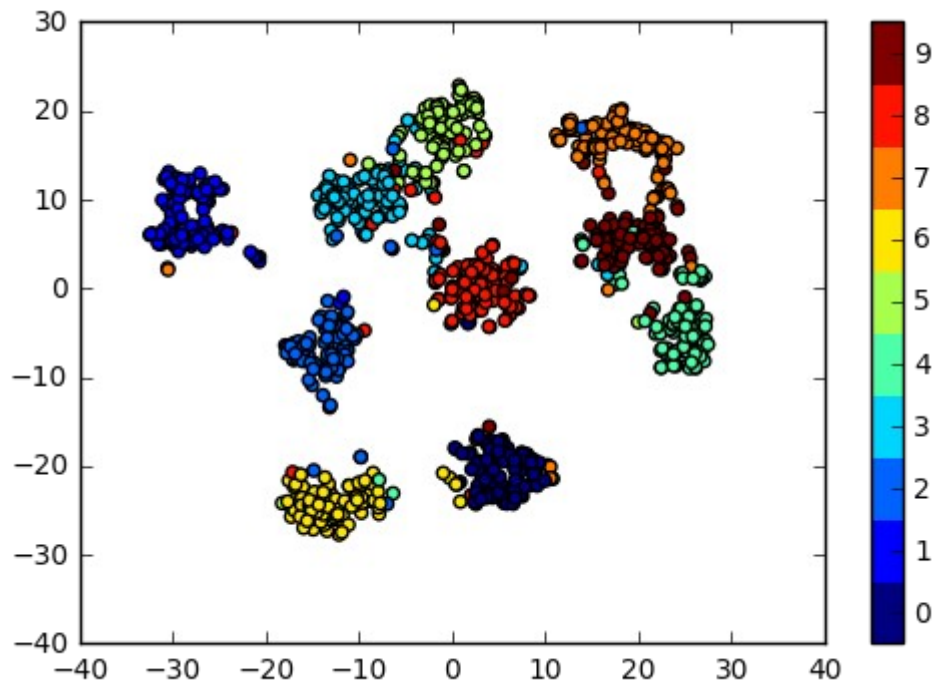
FC layers : 0.000273373210803s

Backward Pass(2048 IMAGES) :

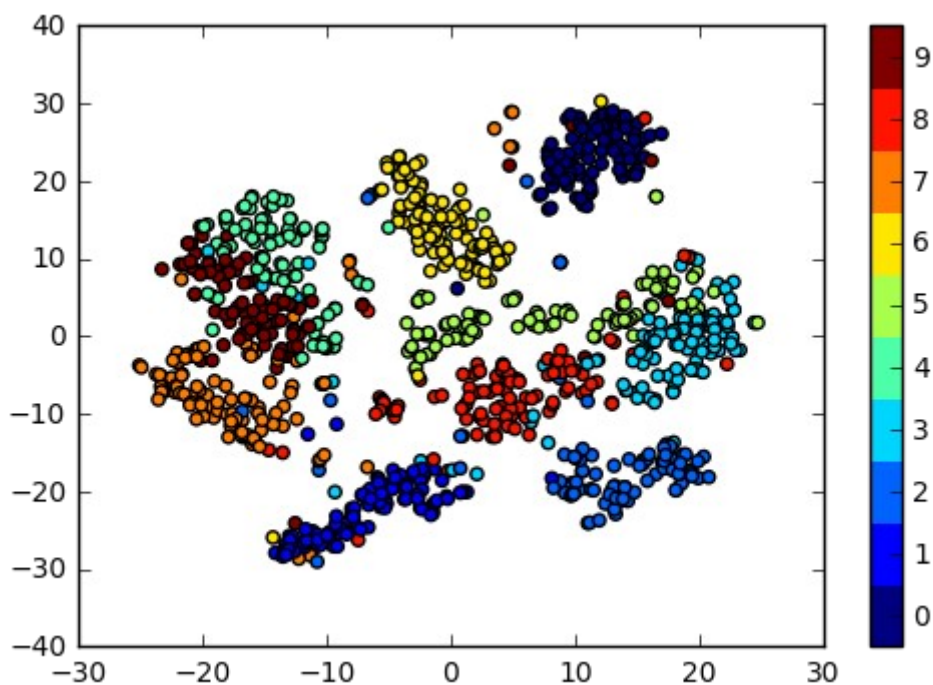
Convolution layers : 0.0229447189486s

FC layer : 0.000417098170146s

t-SNE plots :



t-SNE plot of features taken from the last fc layer. 1000 validation examples are used for this plot. It gives a very good clustering with very less training(1 epoch with batch size of 128 and number of training examples used is 6400)



t-SNE plot of features taken from the second(last) conv layer. 1000 validation examples are used for this plot. It gives a convincing clustering with very less training(1 epoch with batch size of 128 and number of training examples used is 6400)