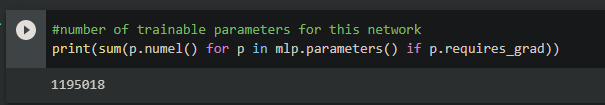
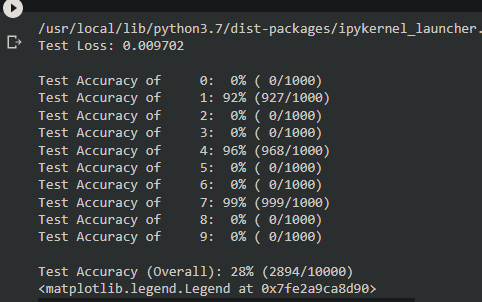
**1) Use a 5-layer multi-layer perceptron (MLP) using ReLu activation to train on the MNIST dataset using the softmax activation at the output layer. (15 marks)**

**A. Note the number of trainable parameters for this network.**

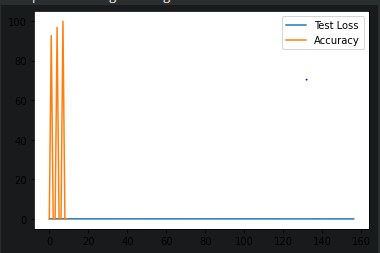
No. of parameters are: 1195018



**B. Describe the accuracy you obtain and show the loss/ accuracy curves.**



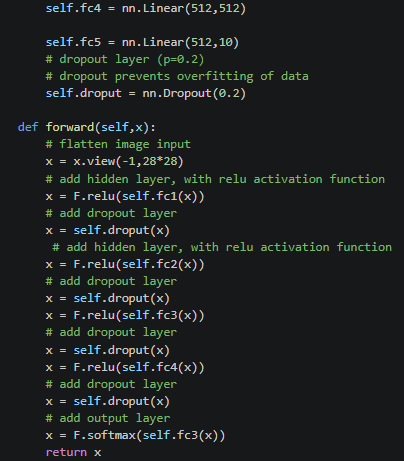
Accuracy obtain is 28% and the loss/ accuracy curves are below:



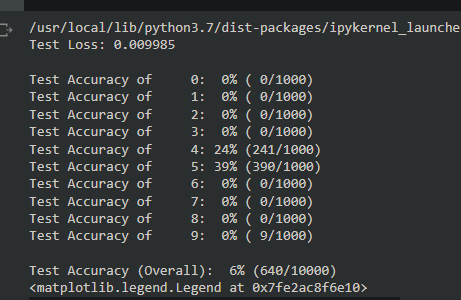
**C. Keep on training the network, do you observe any signs of overfitting — describe your observations? Implement the use of regularization using dropout on the network above—change the dropout parameter to see its effect? Document and explain the results that you obtain.**

Yes, I observe overfitting in the network because our model has remembered the images and can’t make decision when a random image came, in short it has no ability to classify any image which it has not seen before.

**Implementing the dropout parameter:**



Here is the Accuracy and it can be seen, it has dropped drastically from 28% to 6%.



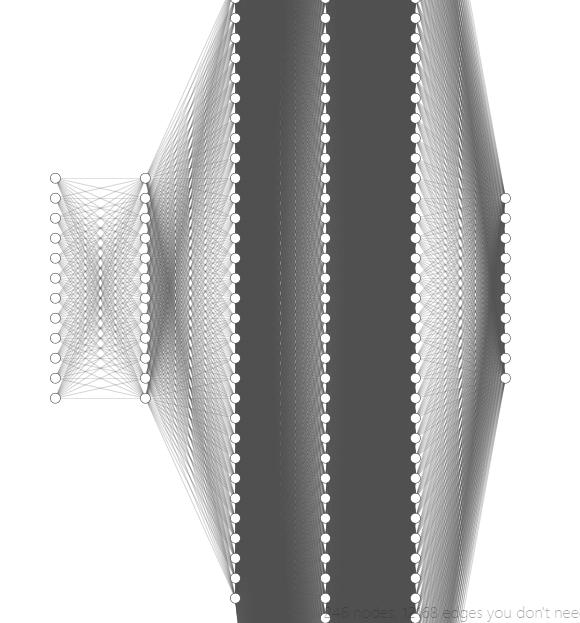
**D. Visualize the performance of the model learned by putting the model in evaluation mode and giving it input and see what outputs are produced. Show 5 sample images and the corresponding outputs from the neural network for those input images.**



Hence in the study it is proved as well that MLP doesn’t perform well as image classifier.

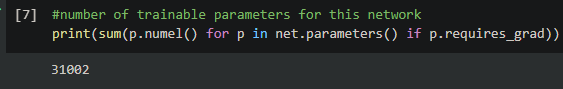
**2) Apply a convolutional neural network (CNN) on the dataset. You should apply three convolutional layer. Use ReLu activations in all these three layers. For the first convolutional layer, use a kernel size of 3x3 and use 12 filters. For the second convolutional layer, use a kernel size of 6x6 with 12 filters and a stride size of 2. For the third convolutional layer, use a kernel size of 6x6 with 32 filters and a stride size of 2. Thereafter flatten the output of the third convolutional layer into a vector and have a fully connected output layer with softmax activation. (20 marks)**

**A. Show the network diagram.**



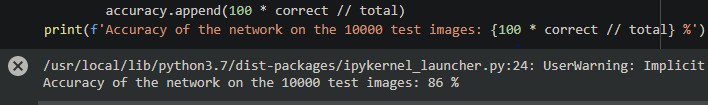
**B. Note the number of trainable parameters for this network.**

Number of trainable parameters are: 31002

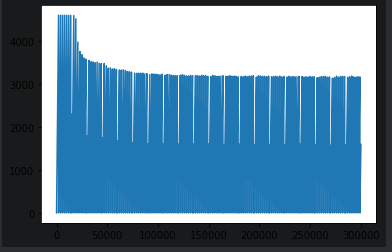


**C. Describe the accuracy you obtain and show the loss/ accuracy curves.**

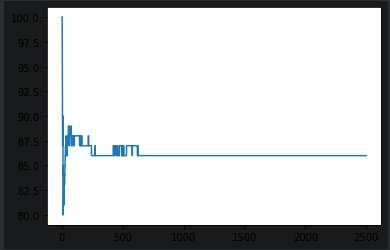
Accuracy is 86%.



**Loss curve:**



**Accuracy Curve:**

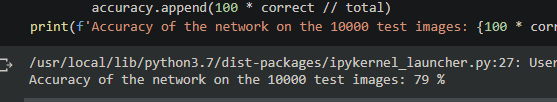


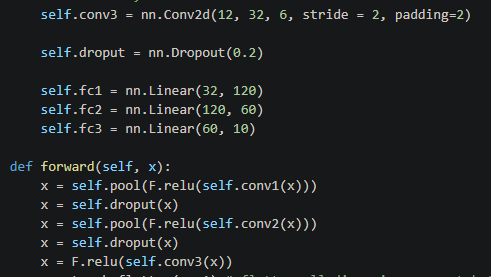
**D. Do you observe any overfitting? Apply Dropout with CNN to see the effect of regularization? Report your results.**

No I don’t observe any overfitting because it gave accuracy on test data of about 86% and seems it hasn’t remembered the images and more over it can make decision on random and unseen data, after validating it over few inputs it gave almost 100% correct result.

**Apply Dropout:**

After applying the dropout with CNN it can be seen that the accuracy of the model is dropped from 86% to 79%, but the data which I used to check the accuracy, this model gave 100% correct results.





**E. Try to visualize the filter outputs of the first layer of your CNN model and reflect on your findings? [for hint, watch the lecture** [**https://youtu.be/G1hGwHVykDU?t=300**](https://youtu.be/G1hGwHVykDU?t=300)**]**



The first convolution layer converts all the pixels in its receptive field into a single value.

**F. Compare and contrast the performance of the CNN vs the 5-Layer MLP (Q1) DNN in Q1. Compare and contrast. Summarize five major insights that you have learnt in this assignment in a video.**

1. MLP takes vector as input and CNN takes tensor as input so CNN can understand spatial relation (relation between nearby pixels of image) between pixels of images better thus for complicated images CNN will perform better than MLP.
2. DNN (MLP) uses many fully-connected layers, CNN contains mostly convolutional layers. CNN is a network with a set of layers that transform an image to a set of class probabilities.
3. CNN converges faster than the MLP model in terms of epochs but each epoch in CNN model takes more time compared to MLP model as the number of parameters is more in CNN model than in MLP model, in our model the MLP had too many parameters than CNN but the training time of CNN was more than MLP.
4. The weights are smaller and shared in the network which means less wasteful, easier to train than MLP and way more effective.
5. In CNN, the number of parameters for the network to learn is significantly lower than the multilayer neural networks since the number of units in the network decreases, therefore reducing the chance of overfitting.