A Report

On

**Deep Learning based classification of the Fashion MNIST dataset**

BY

**Abhinava M 2018A8PS0844H**

**Praneet Sai Madhu Surabhi 2019A7PS0060H**

**Rohit Kushwah 2018B5A71062H**

**D Sri Krishna Karthik 2019A7PS0189H**

Under the supervision of

**Prof. N.L. Bhanu Murthy**

**SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS OF**

**CS F425: Deep Learning**

**Assignment-1**



**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN)**

**HYDERABAD CAMPUS**

**(October 2021**)

**The Dataset:**

Fashion-MNIST is a Zalando article image dataset that includes a training set of 60,000 samples and a test set of 10,000 examples. Each sample is a 28x28 grayscale image with a label from one of ten categories. Fashion-MNIST is intended to be a drop-in replacement for the original MNIST dataset for evaluating machine learning methods, according to Zalando. The training and testing splits have the same picture size and structure.

**Preprocessing:**

Each sample in the dataset has a 28x28 pixel image with an integer associated with the pixel ranging from 0 to 255. The higher the number, the more the intensity. Each sample is assigned with one of the 10 labels:

* 0 T-shirt/top
* 1 Trouser
* 2 Pullover
* 3 Dress
* 4 Coat
* 5 Sandal
* 6 Shirt
* 7 Sneaker
* 8 Bag
* 9 Ankle boot

The data is normalized by dividing the entire set with 255.

**Comparative study of Models**

We have evaluated an MLP model with different designs and architectures by varying the Number of Hidden layers, activation functions, neurons in each layer and loss functions. The training and testing errors are as follows.

| # | No. of hidden layers | Activation functions | No. of hidden neurons | Loss | Train Accuracy | Test Accuracy |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | 3 | Tanh, tanh, tanh | 256,128,100 | Categorical cross entropy | 0.9006 | 0.11469 |
| 2 | 3 | ReLu, ReLu, ReLu | 256,128,100 | Categorical cross entropy | 0.9110 | 0.0868 |
| 3 | 2 | Tanh, tanh | 256,128 | Categorical cross entropy | 0.9042 | 0.1053 |
| 4 | 2 | ReLu, ReLu | 256,128 | Categorical cross entropy | 0.9145 | 0.1167 |
| 5 | 2 | tanh, ReLu | 256,128 | Categorical cross entropy | 0.9055 | 0.0868 |
| 6 | 9 | ReLu,tanh,tanh,tanh,tanh,tanh,tanh,tanh,tanh | 1024,512,256,128,100,64,32,24,16 | Categorical cross entropy | 0.8889 | 0.0837 |
| 7 | 9 | ReLu,ReLu,ReLu,ReLu,ReLu,ReLu,ReLu,ReLu,ReLu | 1024,512,256,128,100,64,32,24,16 | Categorical cross entropy | 0.9043 | 0.10689 |
| 8 | 6 | tanh,tanh,tanh,tanh,tanh,tanh | 512,256,128,100,64,32 | Categorical cross entropy | 0.8813 | 0.1066 |
| 9 | 6 | ReLu,ReLu,ReLu,ReLu,ReLu,ReLu | 512,256,128,100,64,32 | Categorical cross entropy | 0.9081 | 0.1031 |
| 10 | 6 | sigmoid,sigmoid,sigmoid,sigmoid,sigmoid,sigmoid | 512,256,128,100,64,32 | Categorical cross entropy | 0.9059 | 0.10769 |
| 11 | 4 | tanh,tanh,tanh,tanh | 256,128,100,64 | Categorical cross entropy | 0.8968 | 0.08569 |
| 12 | 4 | ReLu,ReLu,ReLu,ReLu | 256,128,100,64 | Categorical cross entropy | 0.9075 | 0.0965 |
| 13 | 4 | Tanh, tanh,ReLu,ReLu | 64,32,16,12 | Categorical cross entropy | 0.9014 | 0.0869 |
| 14 | 3 | Tanh, sigmoid, ReLu | 512,256,128 | Categorical cross entropy | 0.9075 | 0.0886 |
| 15 | 3 | Tanh, sigmoid, ReLu | 64,32,16 | Categorical cross entropy | 0.9040 | 0.0982 |
| 16 | 2 | ReLu,ReLu | 256,128 | KL Divergence | 0.1135 | 0.1131 |
| 17 | 2 | Tanh,Tanh | 256,128 | KL Divergence | 0.08 | 0.0786 |
| 18 | 3 | ReLu,ReLu,ReLu | 512,256,128 | KL Divergence | 0.133 | 0.131 |
| 19 | 3 | Tanh,Tanh,Tanh | 256,128,64 | KL Divergence | 0.1035 | 0.105 |
| 20 | 4 | Tanh,Tanh,Tanh,Relu | 256,128,64,  32 | KL Divergence | 0.0908 | 0.092 |
| 21 | 4 | Relu,Relu,Relu,Relu | 512,256,128,64 | KL Divergence | 0.1065 | 0.1067 |
| 22 | 4 | Tanh,Tanh,Tanh,Tanh | 512,256,128,64 | KL Divergence | 0.1014 | 0.1013 |
| 23 | 5 | Tanh,Tanh,Tanh,Tanh,Tanh | 512,256,128,64,32 | KL Divergence | 0.1074 | 0.1068 |
| 24 | 5 | Tanh,Tanh,Tanh,Tanh,Tanh | 1024,512,256,128,64 | KL Divergence | 0.1049 | 0.1035 |
| 25 | 5 | Relu,Relu,Relu,Tanh,Tanh | 512,256,128,64,32 | KL Divergence | 0.1106 | 0.1124 |
| 26 | 4 | Relu,Relu,Relu,Tanh | 512,256,128,64 | KL-Divergence | 0.101 | 0.1 |

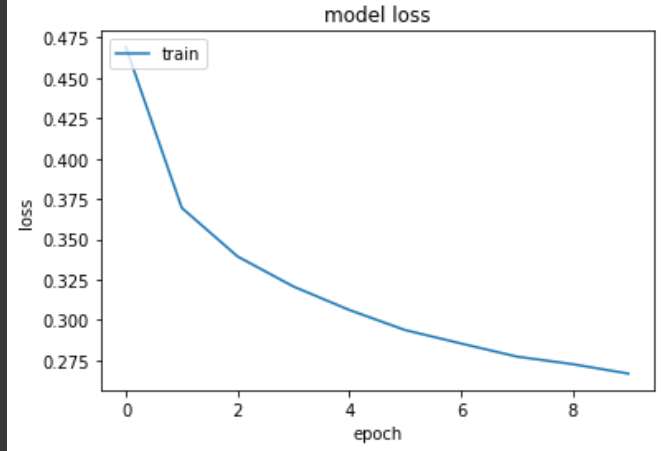
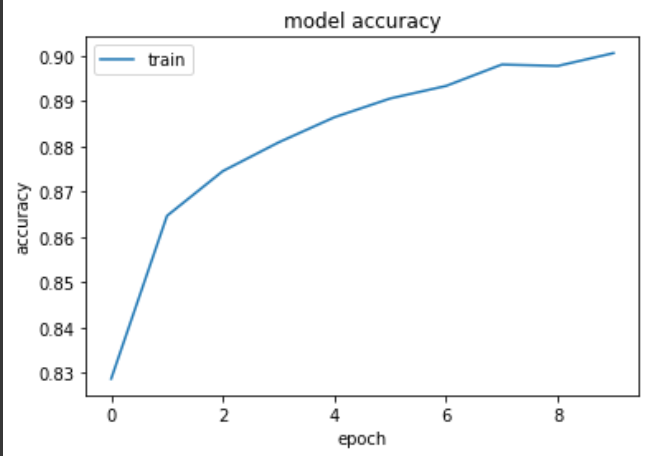
Best model:

The best model is model no. 4 with two hidden layers(256,128) with ReLu activation function on both the layers. The loss function used is categorical cross entropy. It has a training accuracy of 0.9145 and a test accuracy of 0.1167.

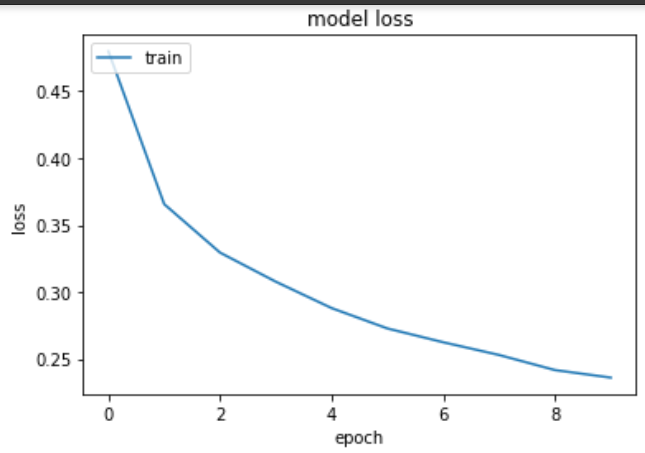
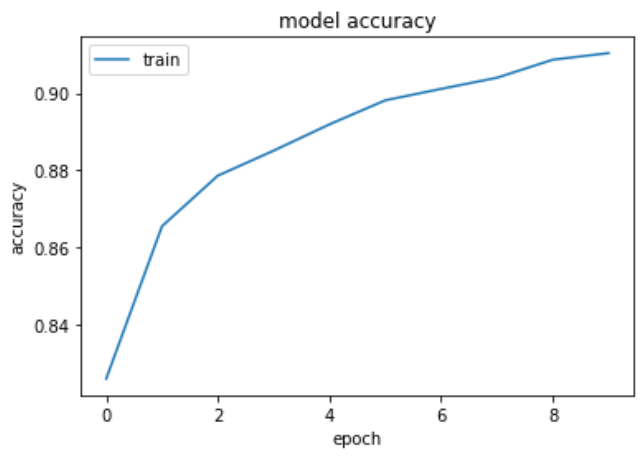
Inferior model:

The inferior model is model no.17 with two hidden layers(256,128) with Tanh activation function on both the layers. The loss function used is KL Divergence. It has a low training and testing accuracy of 0.08 and 0.0786

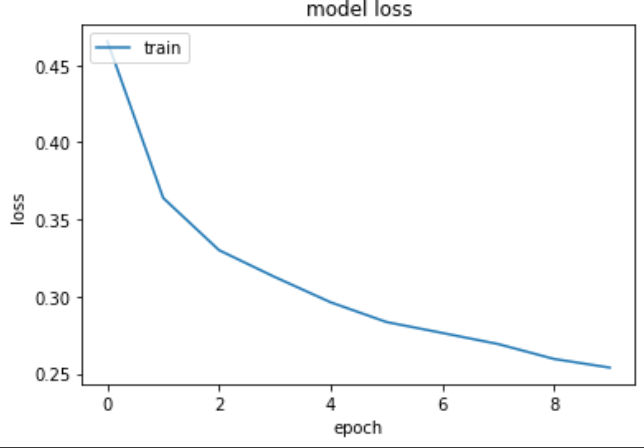
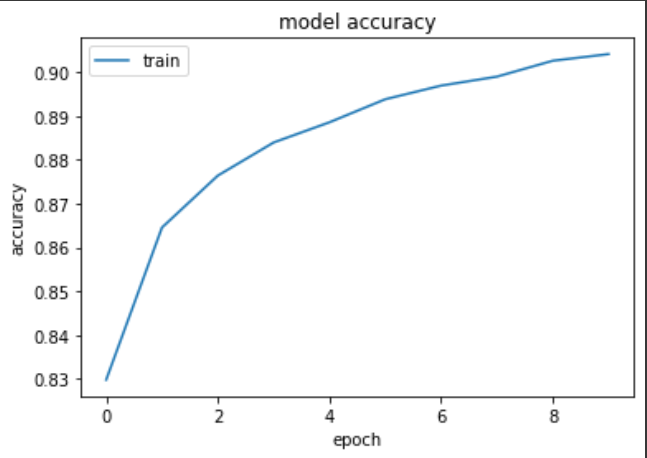
Model 1:



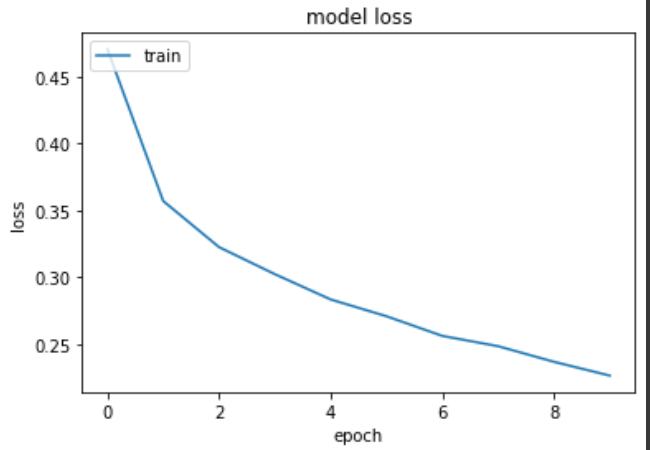
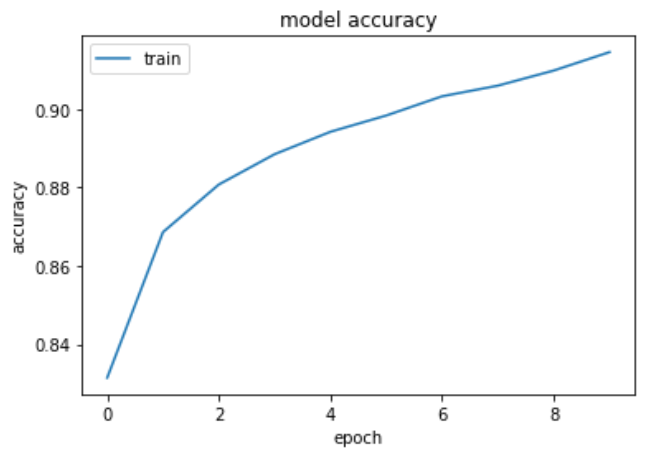
Model 2:



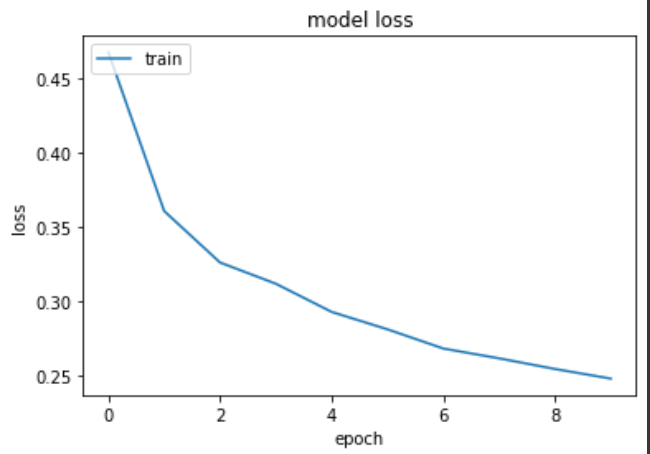
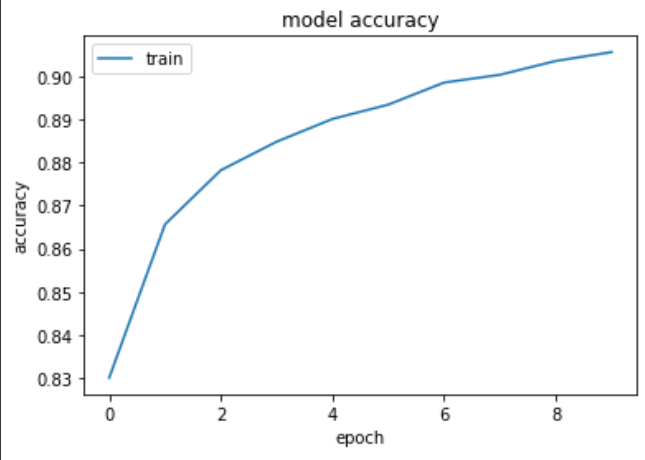
Model 3:



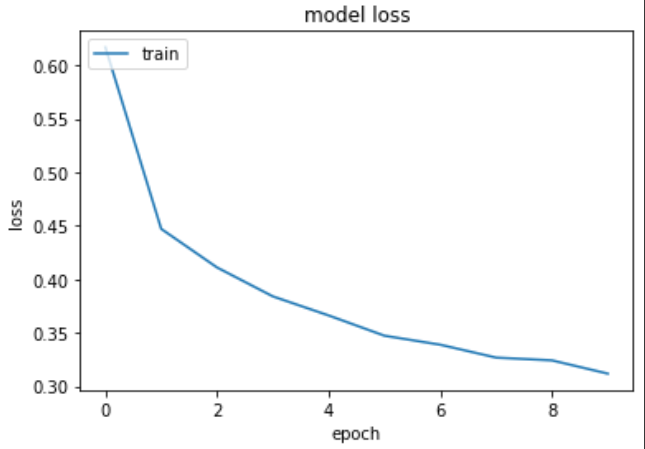
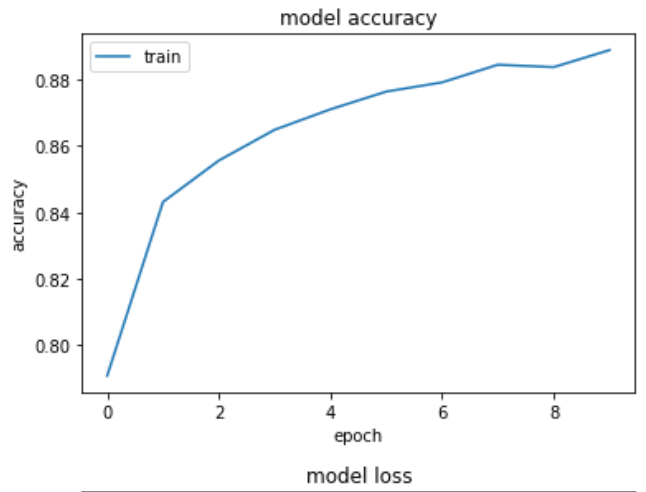
Model 4:



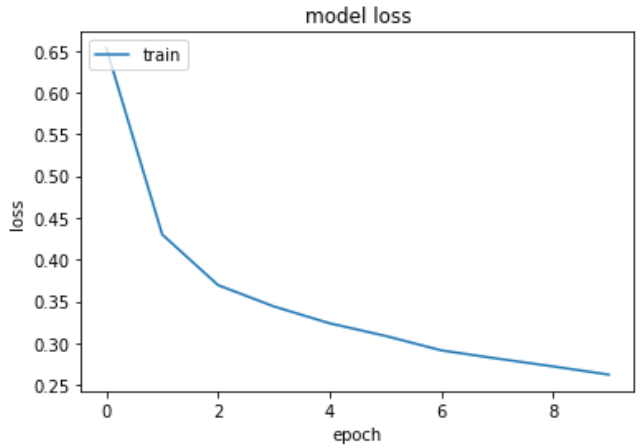
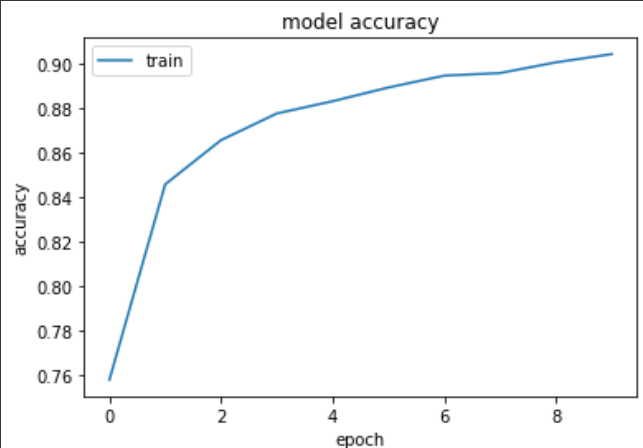
Model 5:



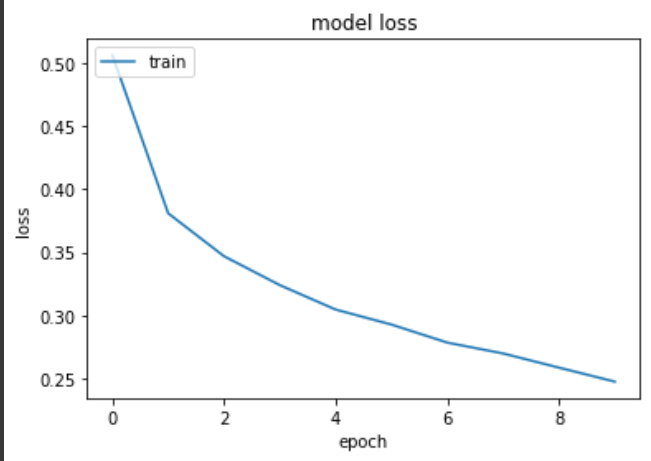
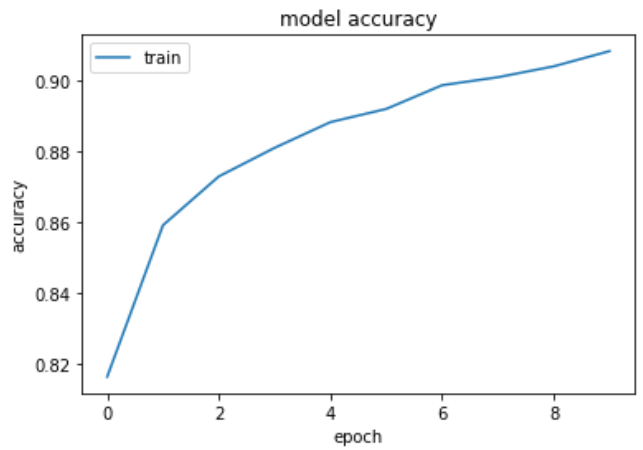
Model 6:



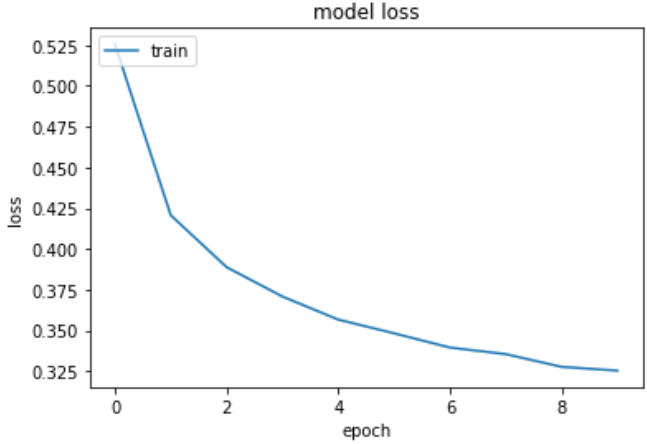
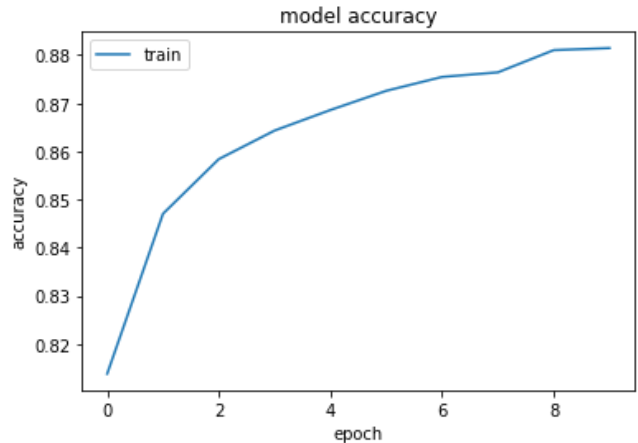
Model 7:



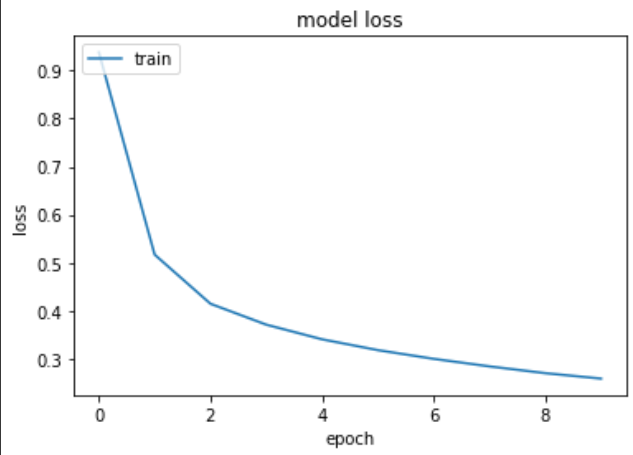
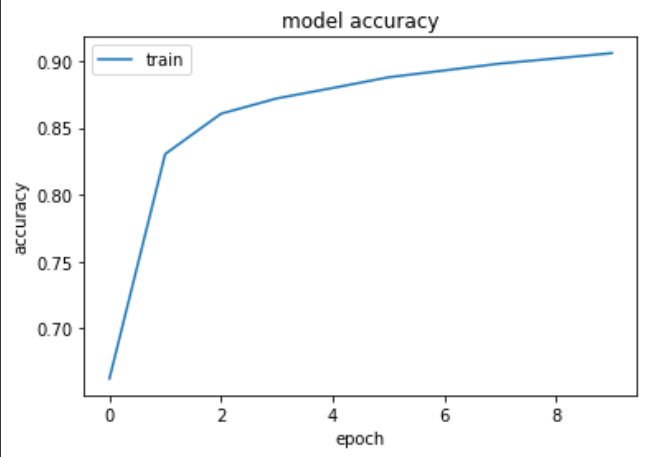
Model 8:



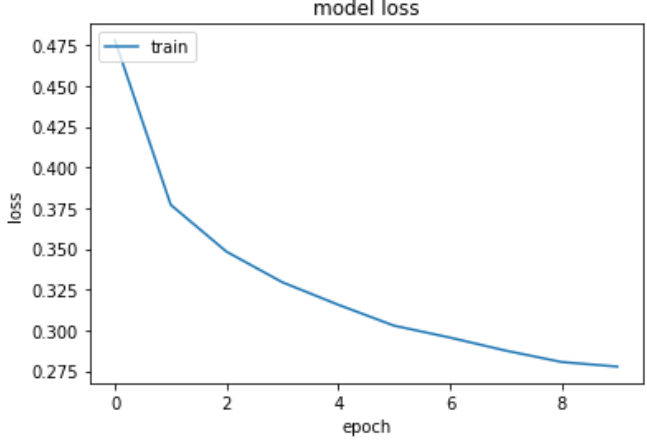
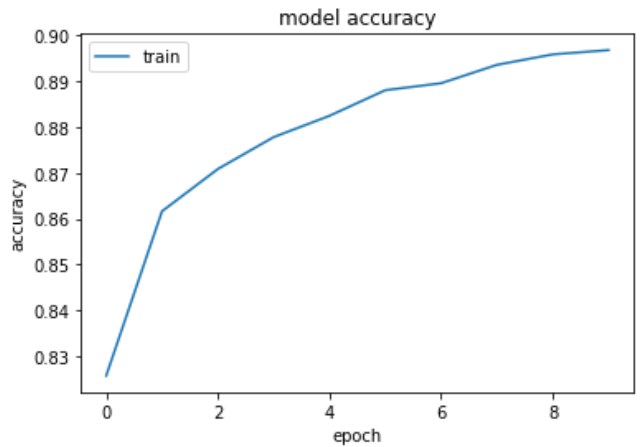
Model 9:



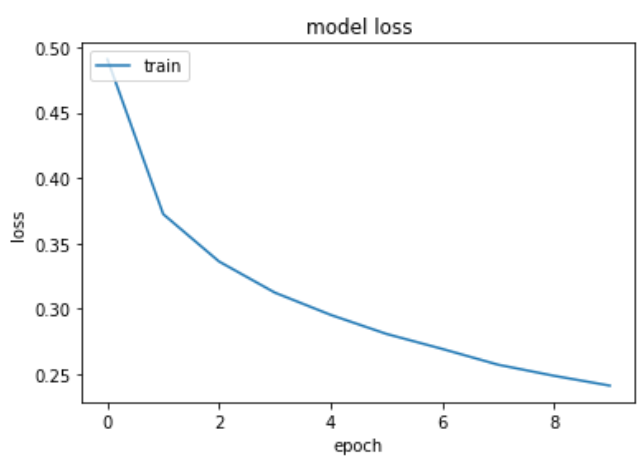
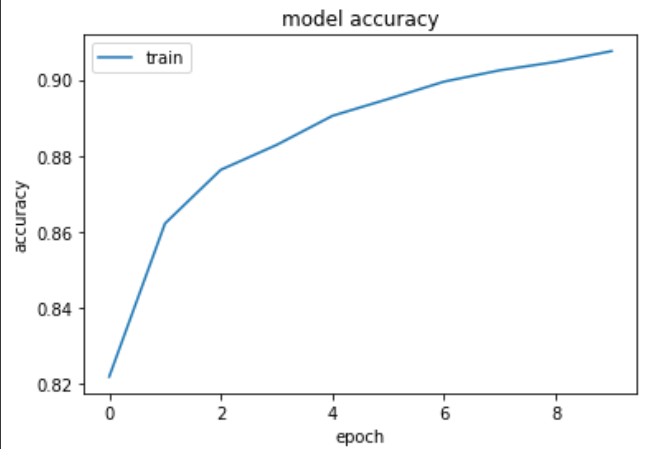
Model 10:



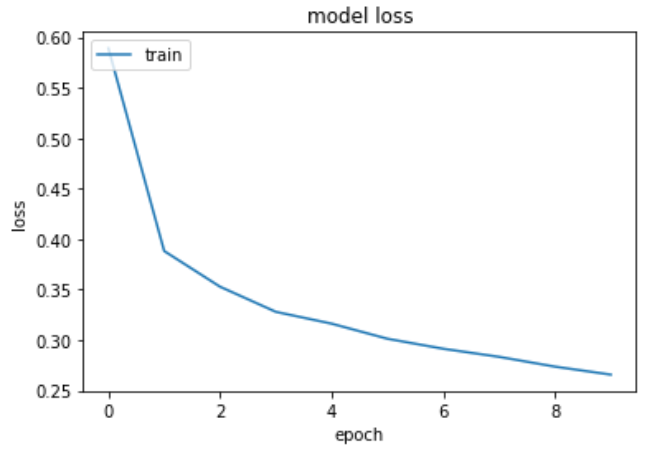
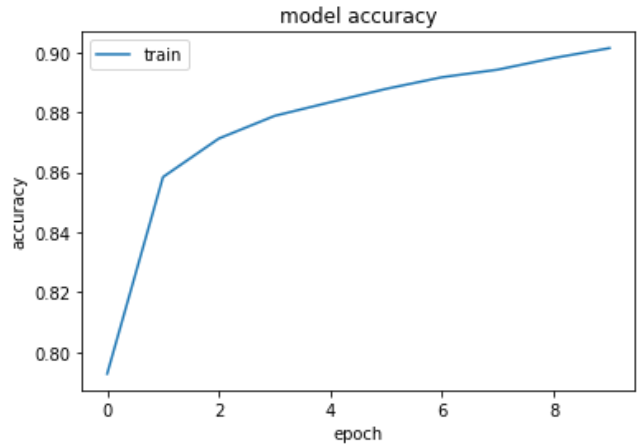
Model 11:



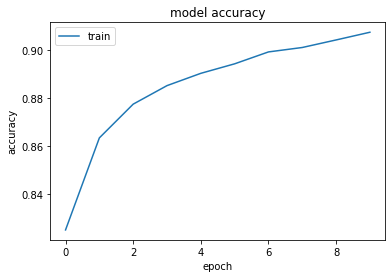
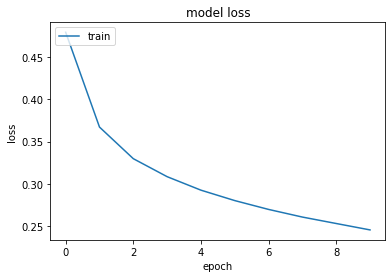
Model 12:



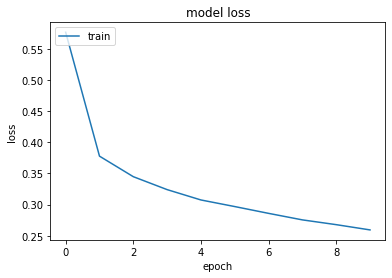
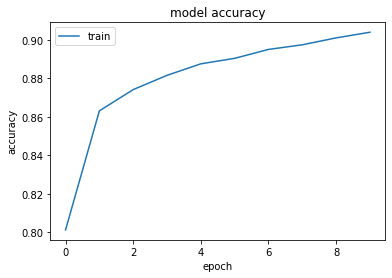
Model 13:



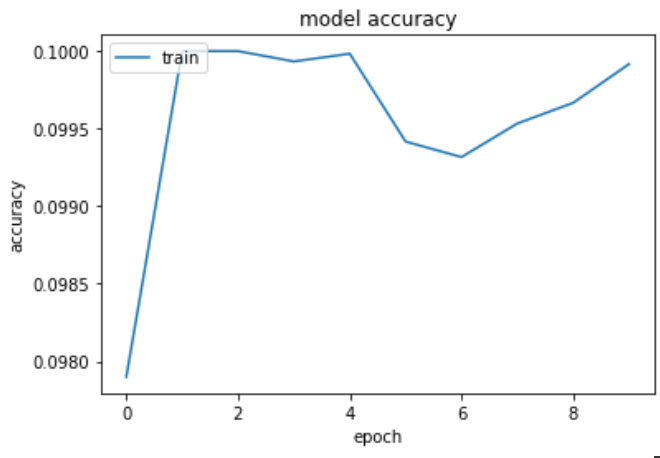
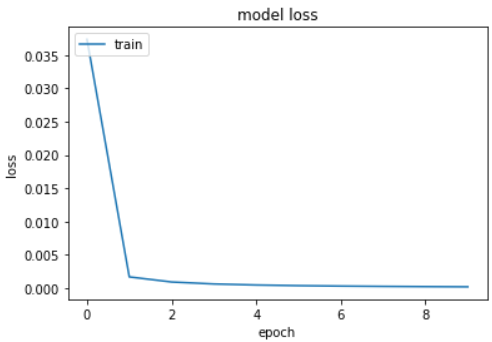
Model 14:

****

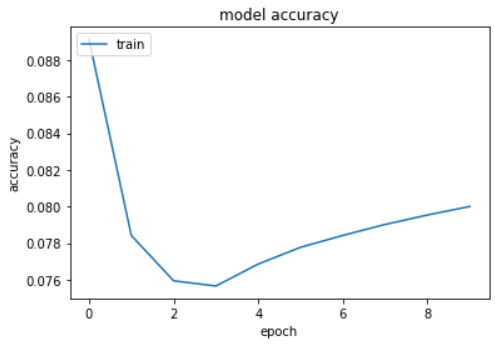
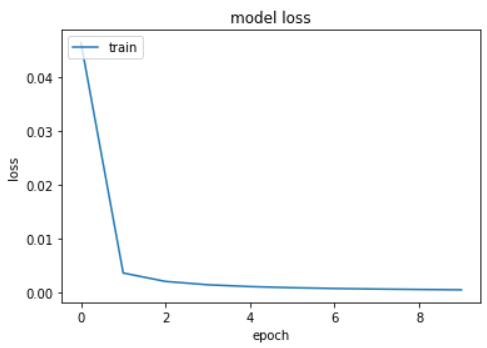
Model 15:



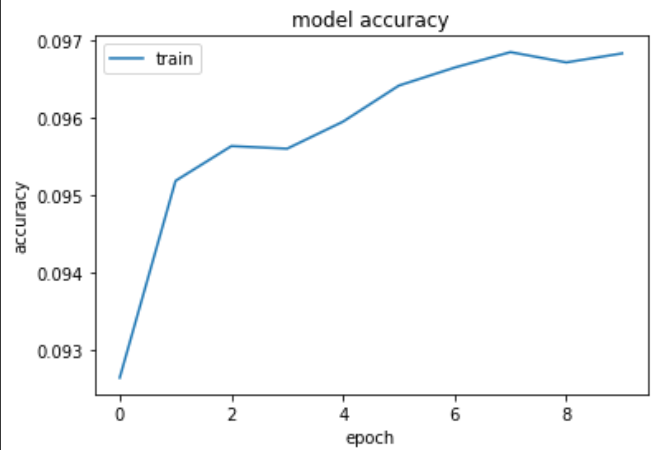
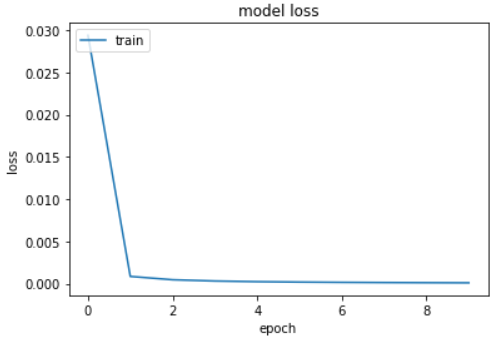
Model 16:

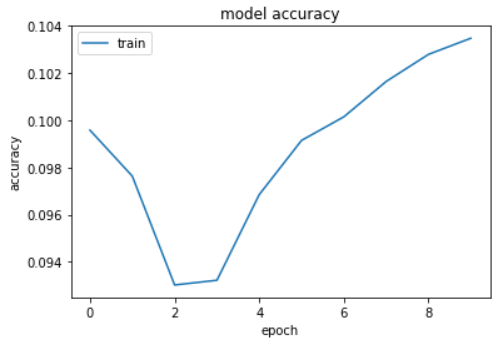
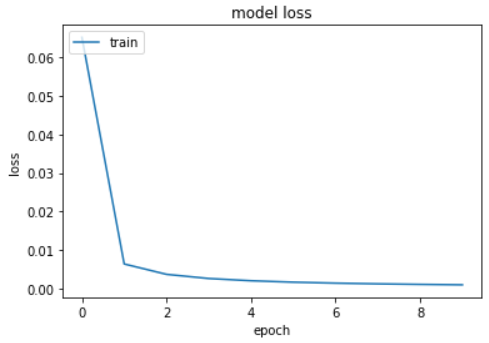
Model 17:

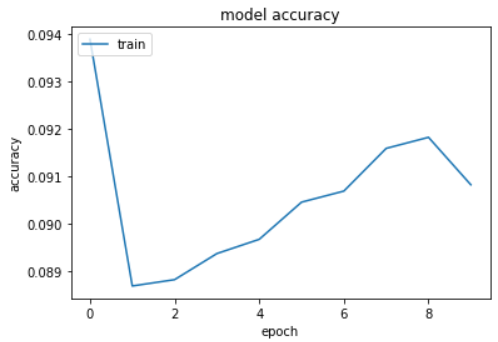
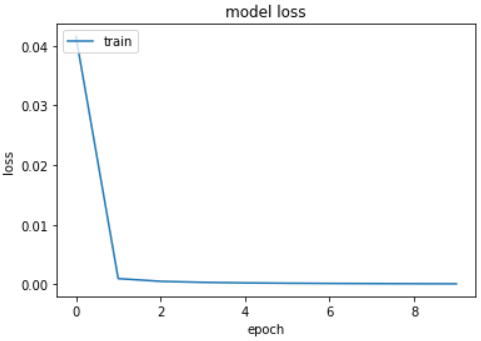
Model 18:

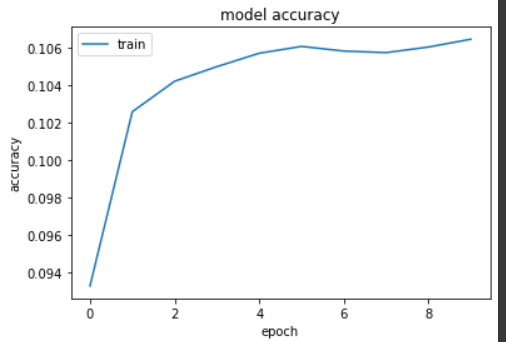
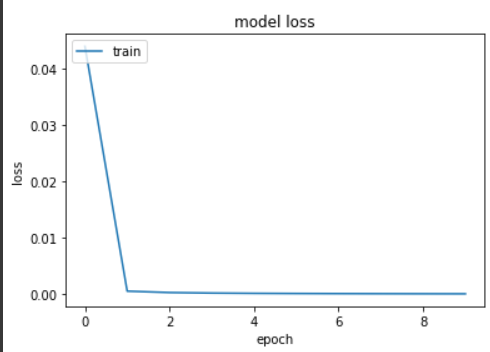
Model 19:

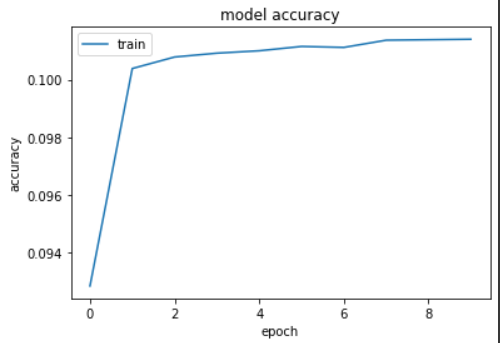
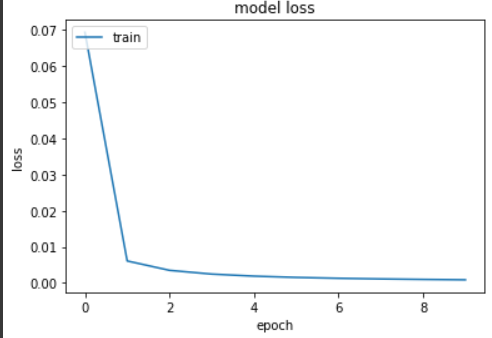
Model 20:

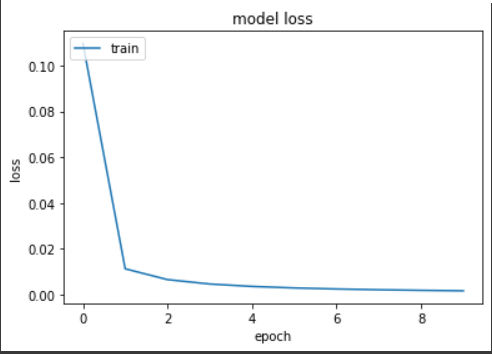
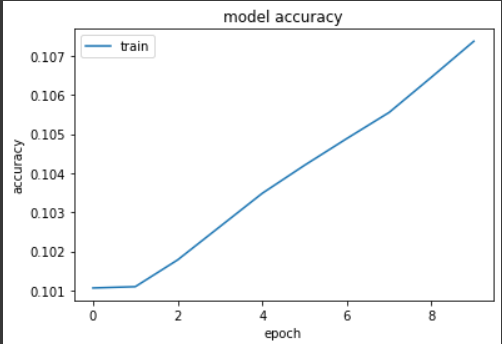
Model 21:

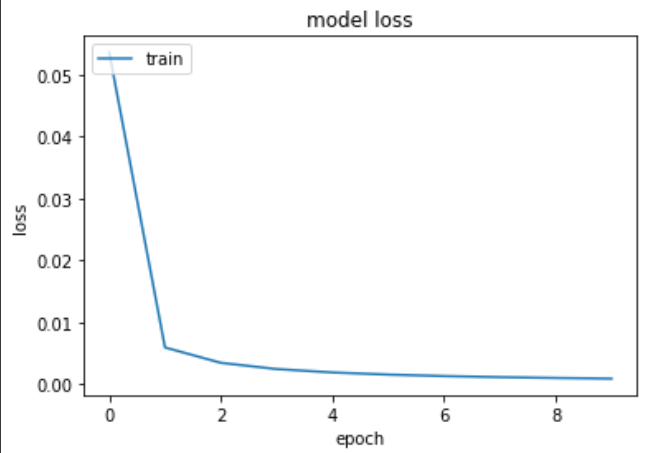
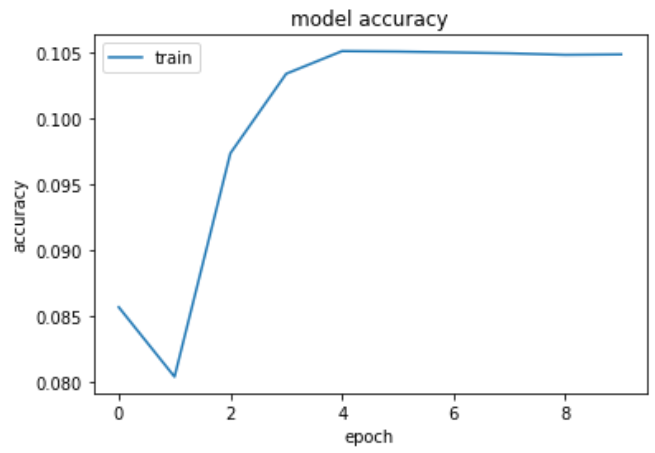
Model 22:

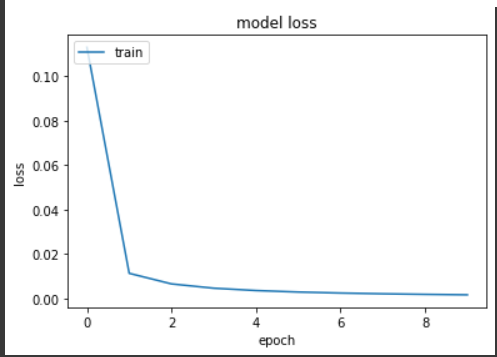
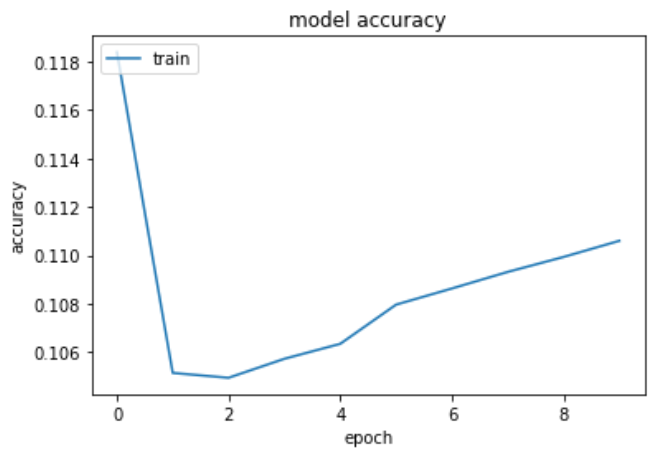
Model 23:

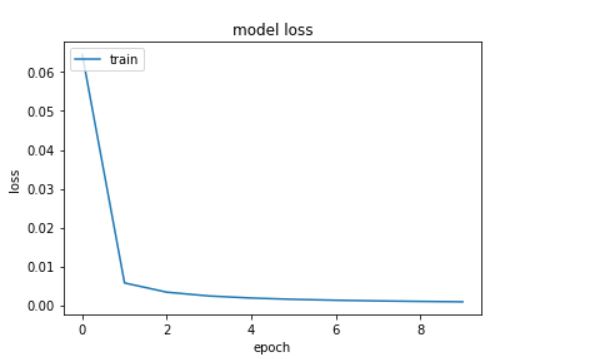
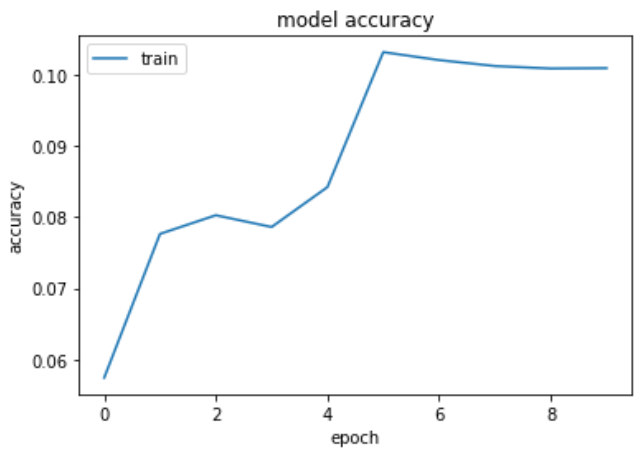


Model 24:



Model 25:

Model 26:



**Observations and comparisons:**

As we can see from the above plots, in Categorical Loss Function we are getting accuray of around 0.87 ~0.9. One more factor we can observe is, as the number of tanh functions increase in hidden layers ,the accuracy drops to around 0.88, but on the other hand if the number of relu functions are more(for the same number of layers) , then the observed accuracy is around 0.91

We also noticed that, as the number of hidden layers increases, model training time also increases significantly because the number of computations increases. For 10 epochs, a 3 hidden layer categorical cross entropy model takes roughly 1.231 sec whereas for 8 hidden layers the runtime is 4 min 30 s.

Similarly for KLDivergence the runtimes are 1.46 sec and 5 min 44 sec respectively.

In the case of K-L , train(accuracy) is around 0.1 and test(accuracy ) is around 0.11 . Which is comparatively very low as compared to Categorical Loss function. Clearly, Fashion MNIST shows better results when Categorical cross entropy loss is considered.

Also, it is observed that types of activation function used also affects the accuracy of both(test and train) as Relu function shows good accuracy as compared to tanh with the same number of hidden layers.

Graphs related to categorical cross entropy show a pattern where accuracy increases linearly and then saturates.Similarly, the loss starts to saturate around 0.55 and then exponentially tends to 0.With the KL divergence, increase in the number of hidden layers doesn't seem to have an improvement in the overall performance, the accuracy tends to saturate around a certain value each time.

It turns out that the test accuracy is very less compared to the training accuracy. This is due to overfitting. Overfitting happens when a machine learning model "memorizes" or learns the training features so well that it cannot generalize on new test data. An overfitted model learns the noise and details in the training data exactly that it assumes the same for new data.