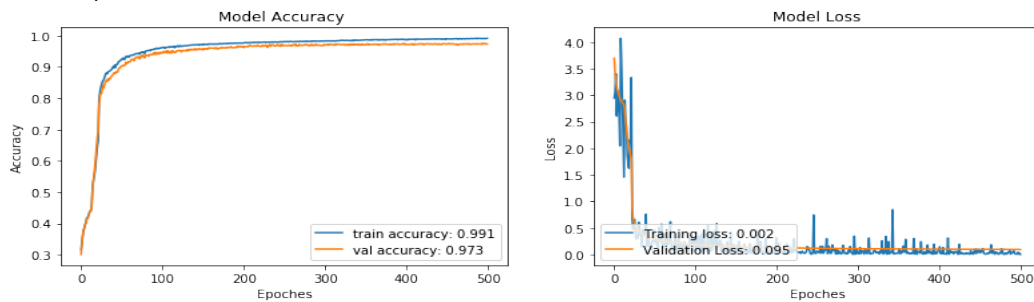i.   A DNN with was constructed and then tested for the following hyper parameter values. The Best model which we found are with the following parameters.
     Learning rate: 0.001 Number of iterations: 500 Batch size: 16 Hidden layers: 3 Number of neurons in hidden layer: 200.


We performed a grid search through parameters learning rate, number of iterations, batch size, hidden layers, number of neurons in hidden layer. Each experiment below shows set of hyper parameters.

➢   In first experiment grid search over hidden layers were performed. Model with 1, 3, and 5 hidden layers, each of which had 100 neurons was tested. The model with three hidden layers performed the best. When we increase layers from three the loss increases and it seems the model will overfit. The remining parameters were kept constant. Learning rate: 0.001 Number of iterations: 500 Batch size: 50
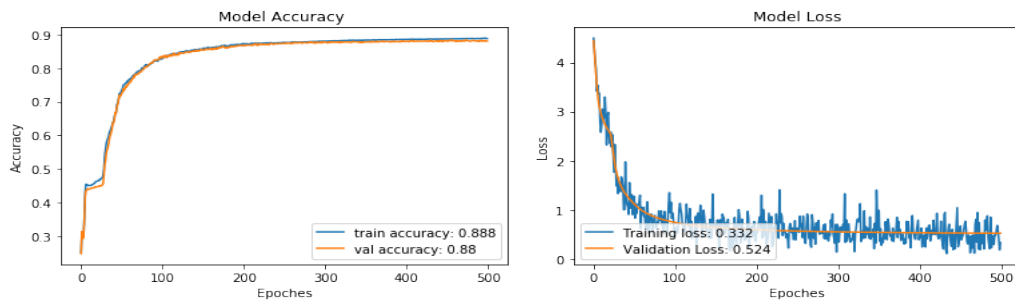
   a) Hidden layers: 1
       Number of neurons in hidden layer: 100
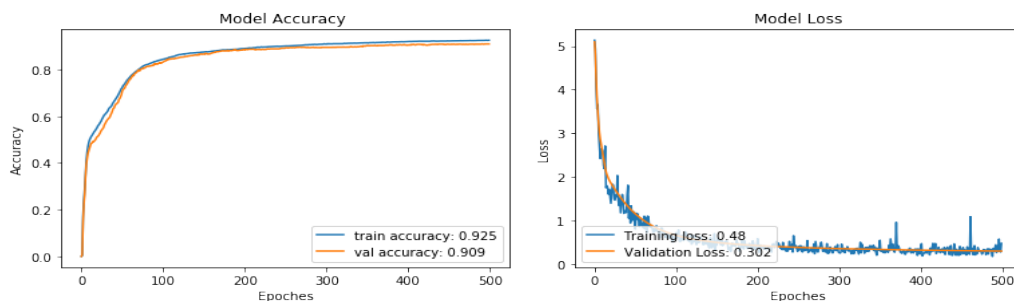
   b) Hidden layers: 3
       Number of neurons in hidden layer: 100,100, 100
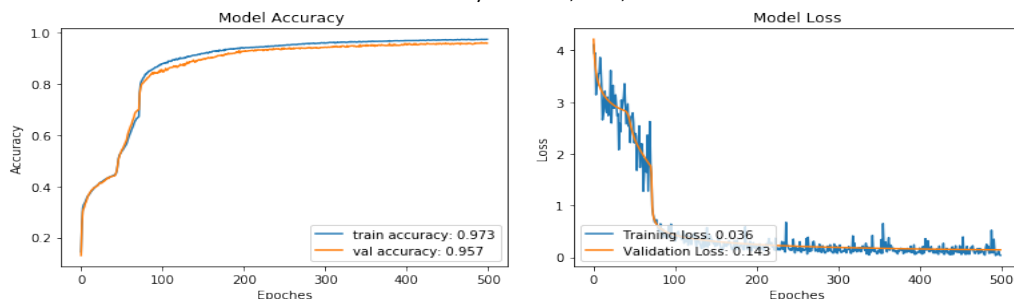
   a) Hidden layers: 5
       Number of neurons in hidden layer: 100,100, 100,100

➢ In second experiment the hidden layer was kept three and learning rate of 0.1, 0.001 and 0.0001 were tried. The model performed best with learning rate of 0.001. with learning rate of 0.1 although it reached very fast to high training accuracy but the validation loss also increased. With learning rate of 0.0001, it is taking very long to converge. The remining parameters were kept constant. Number of hidden layers: 3 Number of iterations: 500 Batch size: 50
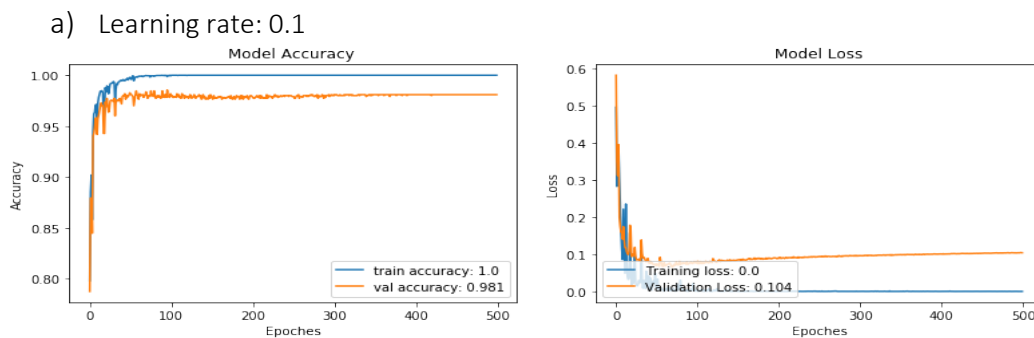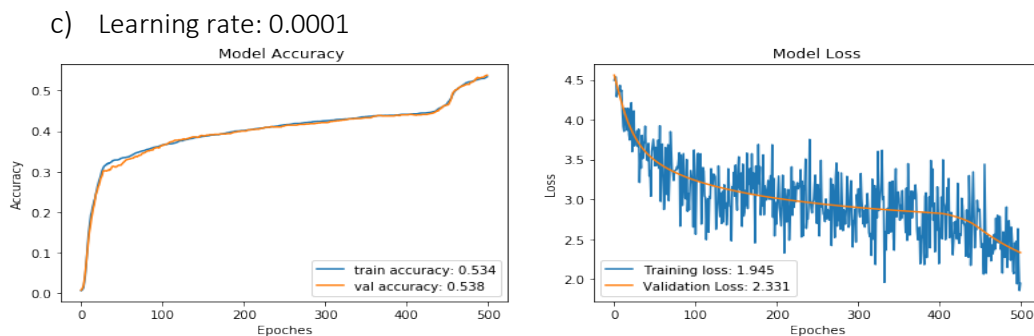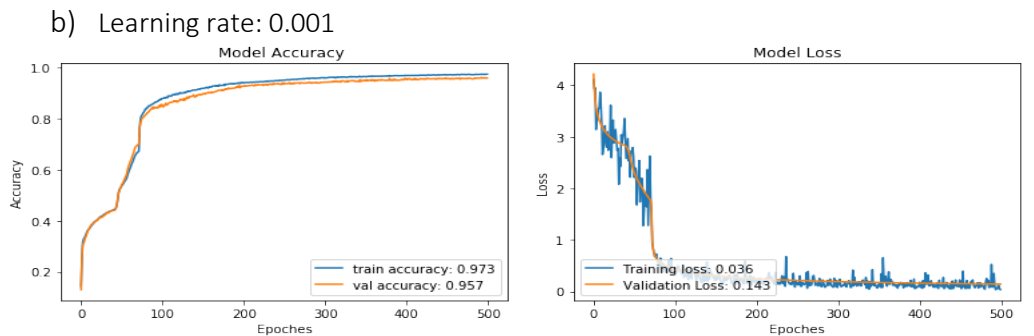
a) Learning rate: 0.1

Figure 1. Model Loss   *Figure 2. Model Accuracy*

b) Learning rate: 0.001

c) Learning rate: 0.0001
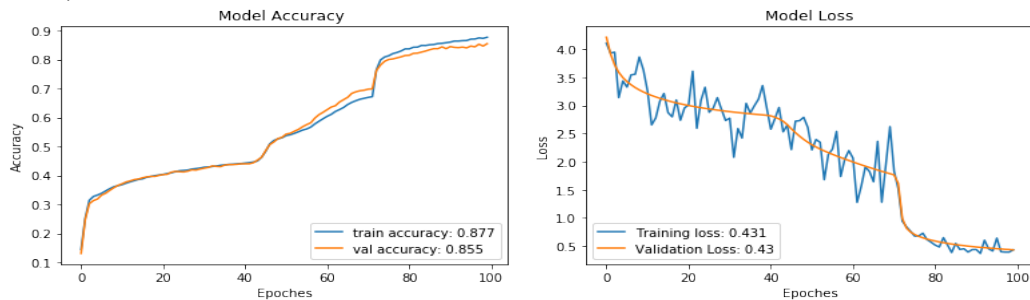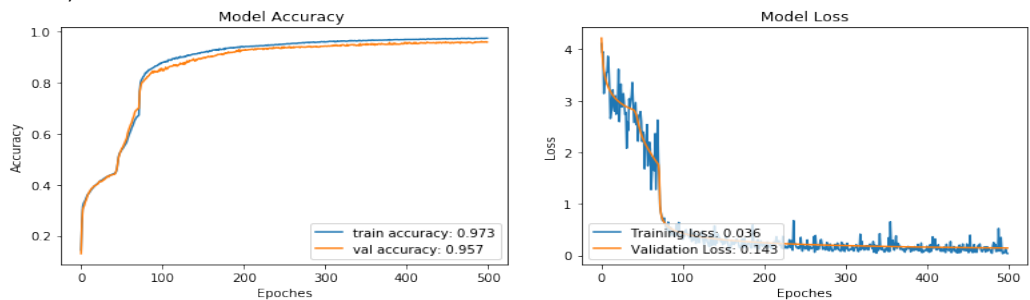

➢ In third experiment the model was trained for 100, 500, 1000 iterations. The model performed best with 500 iteration. With 100 iteration it seems the model needs more iteration to

converge. With 1000 iteration there is some negligible changes in accuracy and seems it will overfit if we increased the epochs. The remining parameters were kept constant. Number of hidden layers: Learning rate: 0.001, Number of hidden layers: 3 Batch size: 50

a) Number of iterations: 100

b) Number of iterations: 500
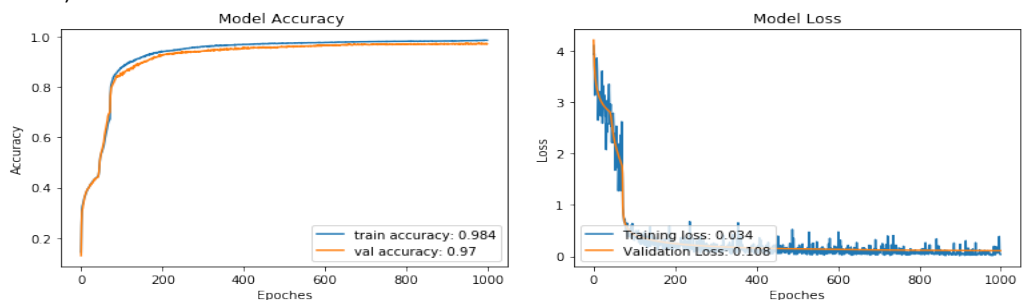

c) Number of iterations: 1000

Figure 5. Model Loss   Figure 6. Model Accuracy

➢ In fourth experiment the model was trained with batch sizes of 16, 32, 128. The model performed best with batch size of 16. Other parameters were kept constant. Learning rate: 0.001, Number of hidden layers: 3, Number of iterations: 500
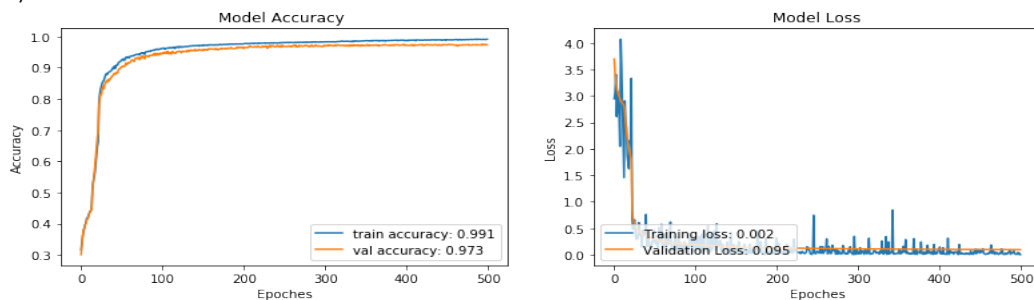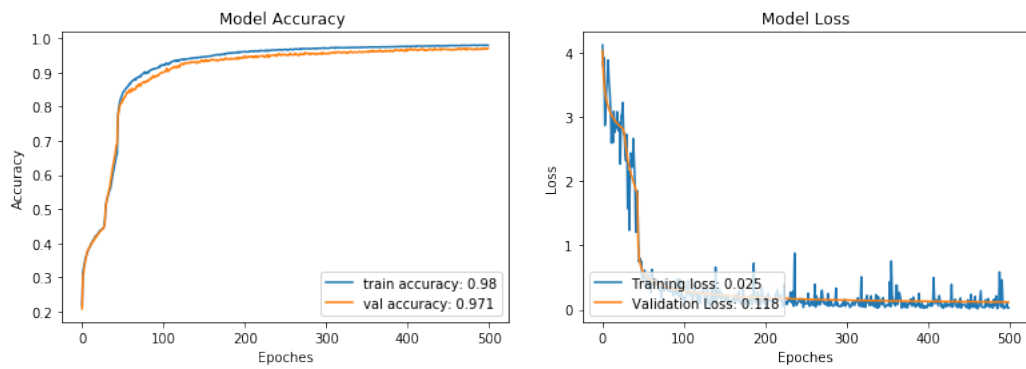
a) Batch size: 16

Figure 1. Model Loss   Figure 2. Model Accuracy
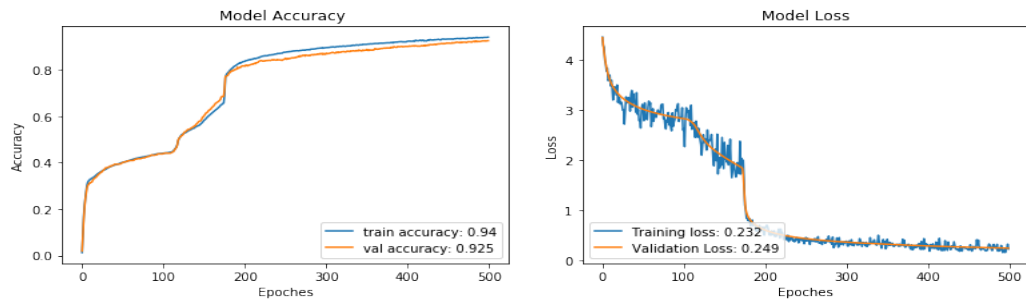
b) Batch size: 32

c) Batch size: 128

Figure 5. Model Loss

Figure 6. Model Accuracy

ii. The Scores for our best model is shown here.

Matric	Score
Accuracy score	0.9727767695099818
Precision Score (micro)	0.9727767695099818
Precision Score (macro	0.9631171986383662
F1 score (micro)	0.9727767695099818
F1 score (macro)	0.9625024545043145

iii. The comparison of the optimizers are shown in figure below.
The adagrad optimizer is given here. In comparison to our best model in gradient descent with same hyperparameters the adagrad didn't perform well.
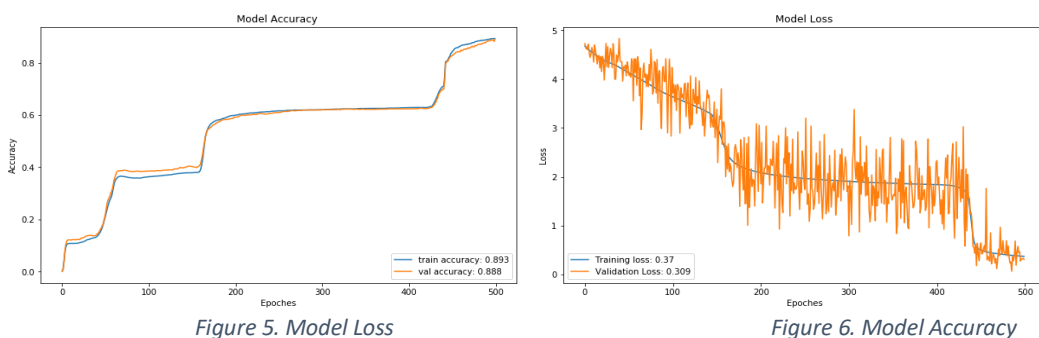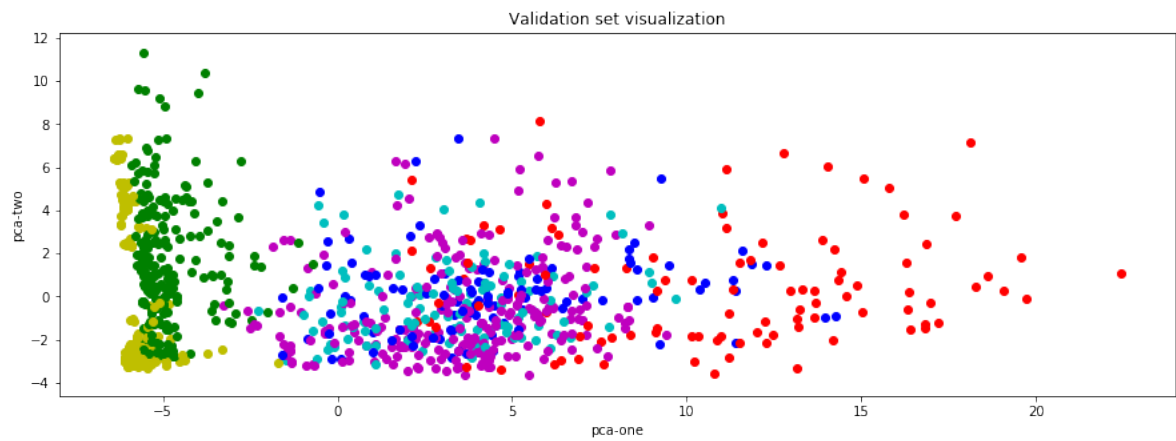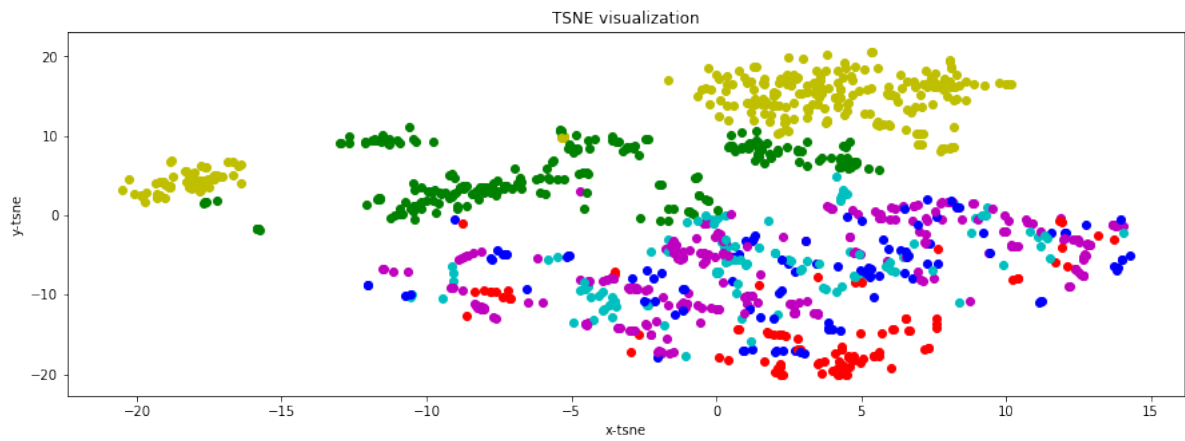

Figure 5. Model Loss

Figure 6. Model Accuracy

iv. Visualization of validation set through PCA using 2-component analysis.

Validation set visualization

v.   Projection of validation set onto a 2D space by t-SNE

TSNE visualization

Q2: Please find the file "abc_answer" in assignment folder.