ASSIGNMENT 4

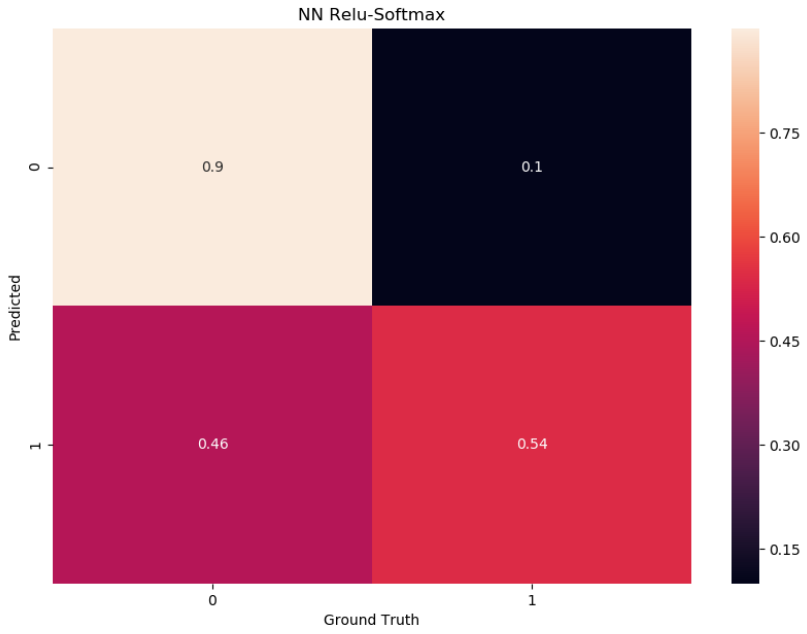
Statistical Machine Learning

Question 1

Dataset Normalization

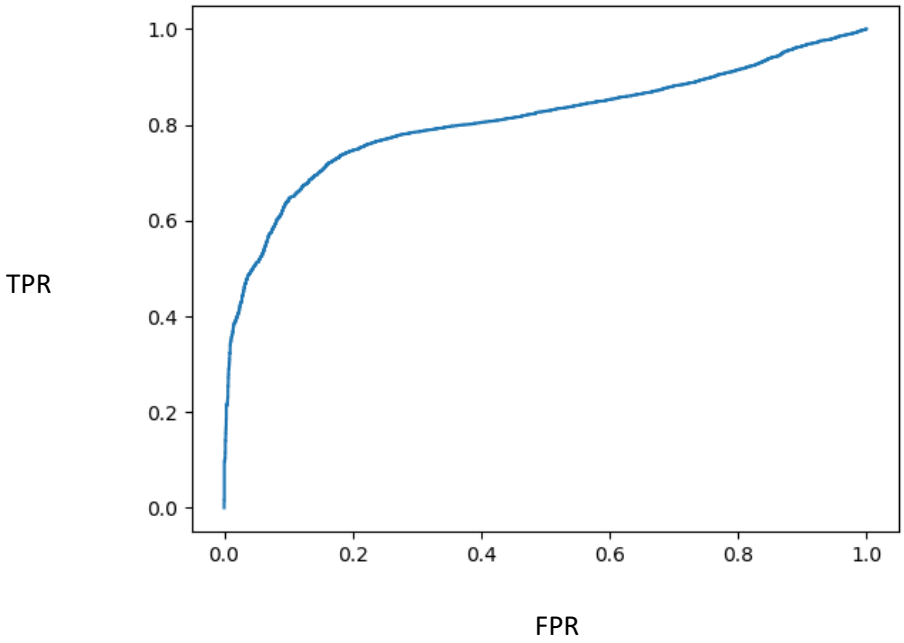
- Each of the categorical features were mapped to a natural number and replace in the dataset.
- Each column was the normalized to a value between 0 and 1 according to values in the columns.

Confusion Matrix

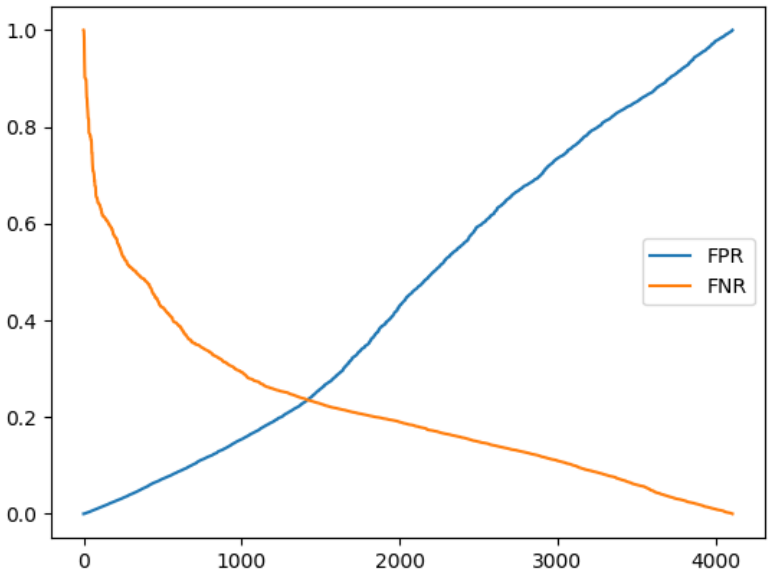


- **Overall Accuracy**: 82.082%
- **Class wise Accuracy**:
  - **0: <50k**, 90%
  - **1: >=50k**, 54%

ROC Curve



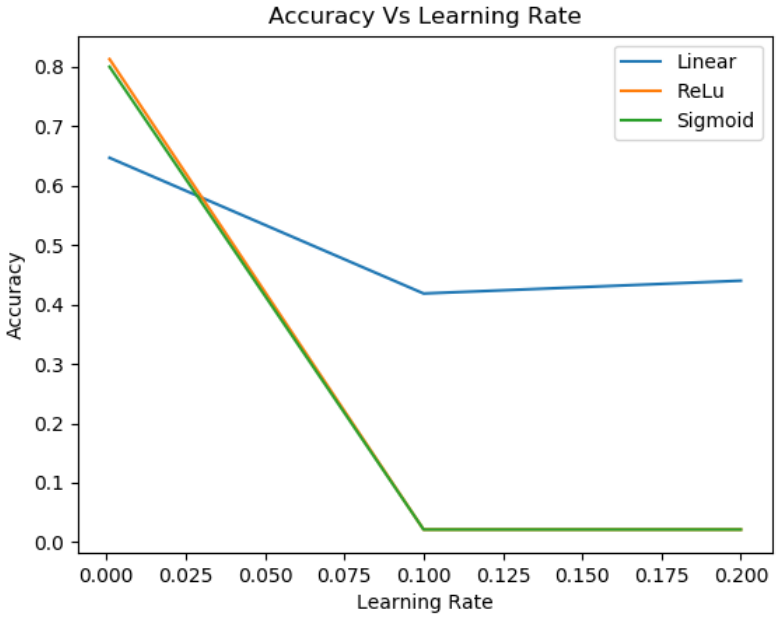
Equal Error Rate Plot



Equal Error rate can be calculated by finding the intersection point of Graph of FNN and FPR.

Question 2

a.

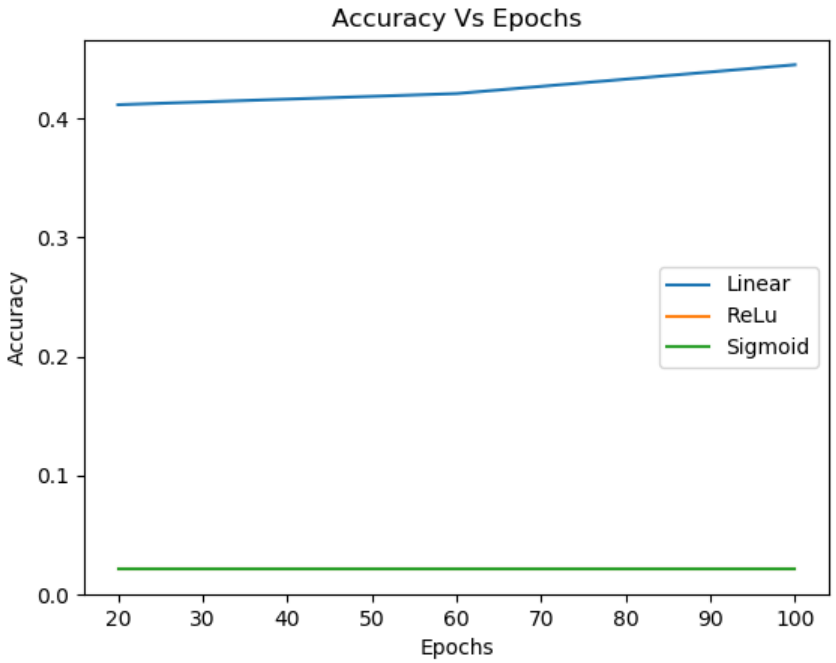


Activation/Learning Rate	0.2	0.1	0.001
Linear	0.44005	0.418510	0.646861
Relu	0.021276	0.0212765	0.8128723
Logistic	0.0212765	0.021276	0.800053

Observation and Inferences

Relu and Logistic gives poor performance on high learning rate comparative to 0.001, where as Linear or Identity gives much better performance relative to other activation. General trend is found to increase upon decrease in learning rate. Results provided by both Relu and Logistic match giving an impression that they are classifying each data point in same class, through this maybe different each time.

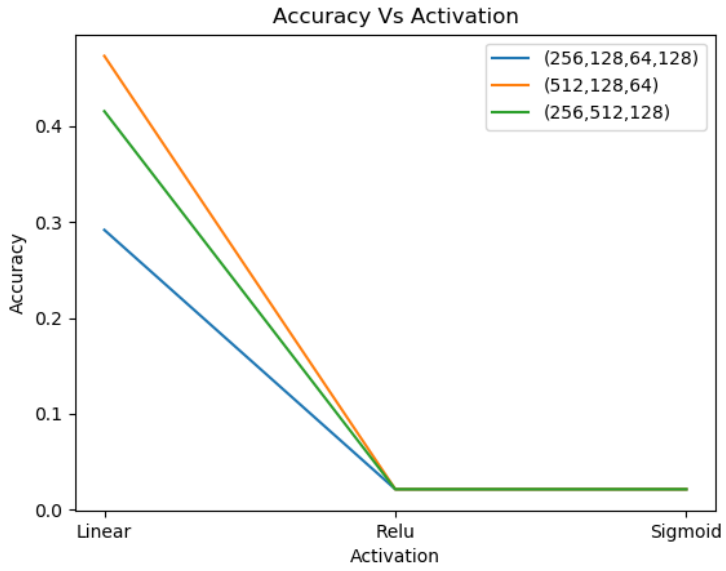
d.



Activation/Epochs	20	60	100
Linear	0.411542	0.4209042	0.4451063
Relu	0.0212765	0.0212765	0.021276
Logistic	0.0212765	0.021276	0.0212765

As stated above Relu and Logistic are performing bad on the given dataset for the deault learning rate of 0.1, and Accuracy remains constant even with increase in number of Epochs. Through for Identity/Linear activation the accuracy gradually increases with increase in Epochs.

b,c.



Layer/Activation	Linear	Relu	Sigmoid
(256,128,64,128)	0.2918085	0.0212765	0.0212765
(512,128,64)	0.473297	0.0212765	0.0212765
(256,512,128)	0.415644	0.0212765	0.0212765

f.

Solver, Alpha	Linear	Relu	Sigmoid
Sgd,0.0001	0.0212765	0.0212765	0.54930
Sgd,0.01	0.0212765	0.0212765	0.573617

Question 3

Bagging Classifier

Estimators	5	10(default)	15	NN
Accuracy	0.276548	0.295487	0.334516	0.248793

It can be observed that the accuracy increases with increase in number of estimators both with respect to original estimator and Bagging Classifier with less number of similar estimators.

Boosting Classifier

Estimators	5	10(default)	15	NN
Accuracy				