REPORT - ASSIGNMENT 4
STATISTICAL MACHINE LEARNING

VAIBHAV VARSHNEY (MT17065)

Dataset information

Three datasets are given for this assignment.

(1) Adult Dataset:

Dataset consist of categorical data and some continuous variables. The categorical variables are mapped to numeric values and then normalized between the range 0 to 1.

The output variable is mapped to 0 and 1. The value '<=50K' is mapped to 0 and '>50k' is mapped to 1.

(2) Extended MNIST:

In extended MNIST, there are 47 classes.

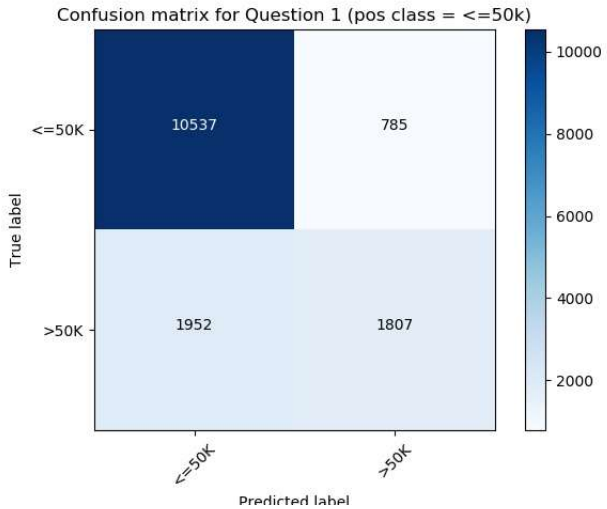
(3) CIFAR-10:

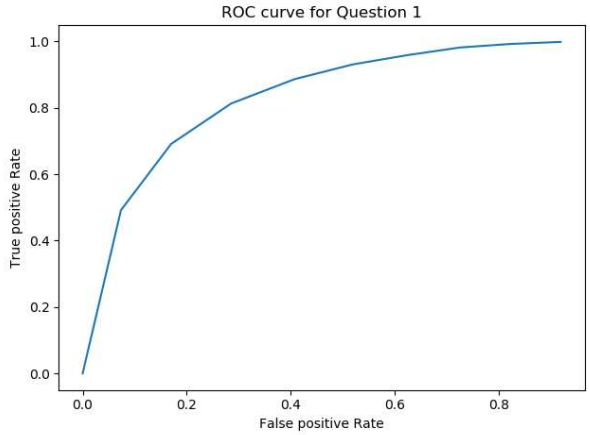
This dataset is considered for Question 3 for bagging and boosting classifier.

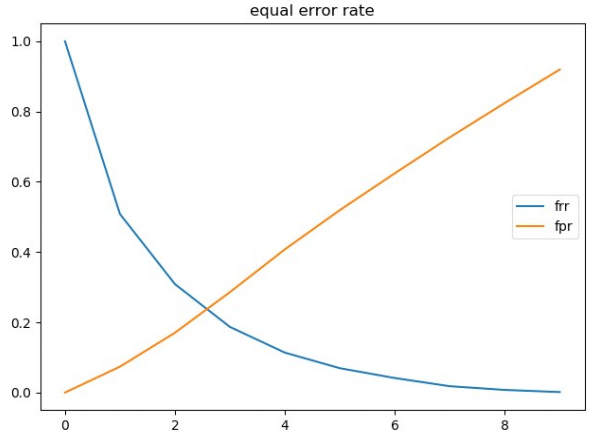
Question 1: Neural Network Implementation

Neural Network is implemented with one hidden layer having 3 nodes. **ReLu** activation is used at hidden layers. **Sigmoid function** is used at output layer.

Overall accuracy	0.82
Class 0 accuracy	0.93
Class 1 accuracy	0. 48







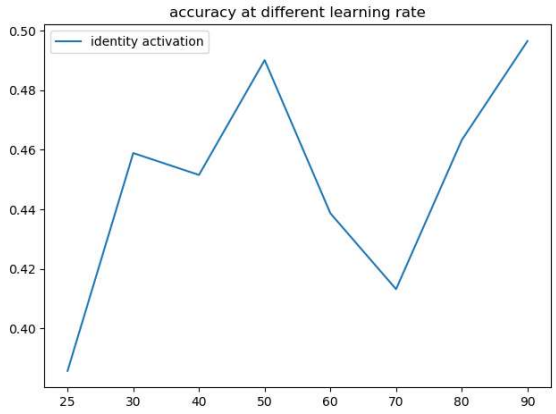
Observations and Inferences

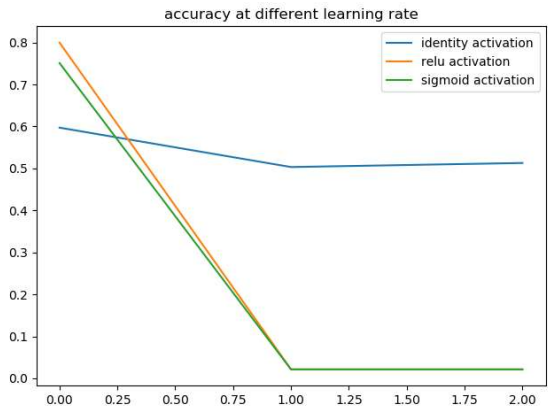
(1) Observing the dataset, the two classes have **class imbalance problem, i.e., <50k is class 0 here, and >50k is class 1 here. Class 0 is nearly 75% and class 1 is nearly 25%.**
(2) The class-wise accuracy shows that the model trained doesn't work well for one of the classes because class.
(3) Hence, accuracy is not a proper metric for evaluation of the model. It is a biased metric.
(4) Looking at the confusion matrix, nearly half of the samples of class 1 are classified as class 0 samples.
(5) The ROC curve of class 0 shows that **with increase in TPR there is also an increase in FPR rate,** indicating that with increase of correct classification of class 0 samples, class 1 samples are also classified as class 0 samples.
(6) **ERR, equal error rate, is at** nearly 0.2 threshold value. This is shown in the attached figure.

Question 2: Experiments with Neural Network

(a) Identity activation function

Network	Accuracy
(256, 128, 64)	0.503
(350, 700, 350)	0.494
(64, 256, 128, 64, 128)	0.501
(256, 128, 64, 128)	**0.526**





(b) Sigmoid function

Network	Accuracy
(256, 128, 64)	0.0212
(350, 700, 350)	0.0212
(64, 256, 128, 64, 128)	0.0212
(256, 128, 64, 128)	**0.0212**

With learning rate=0.001,

Network	Accuracy
(256, 128, 64)	**0.75**

With sigmoid activation function, with the learning rate = 0.1 (as stated in assignment), accuracy was obtained to be **0.0212.** Even after experimentation, at different layers and adding additional hidden layer, accuracy obtained to be same.

But with change in learning rate, i.e., with lowering the learning rate, accuracy tend to increase. **Accuracy at learning rate = 0.001, is 0.751.**

Even with change in iteration didn't show any progress with learning rates 0.1 and 0.2

(c) ReLu activation function

Network	Accuracy
(256, 128, 64)	0.0212
(350, 700, 350)	0.0212
(64, 256, 128, 64, 128)	0.0212
(256, 128, 64, 128)	**0.0212**

With learning rate=0.001,

Network	Accuracy
(256, 128, 64)	**0.799**

With sigmoid activation function, with the learning rate = 0.1 (as stated in assignment), accuracy was obtained to be **0.0212.** Even after experimentation, at different layers and adding additional hidden layer, accuracy obtained to be same.

But with change in learning rate, i.e., with lowering the learning rate, accuracy tend to increase. **Accuracy at learning rate = 0.001, is 0.799.**

Even with change in iteration didn't show any progress with learning rates 0.1 and 0.2

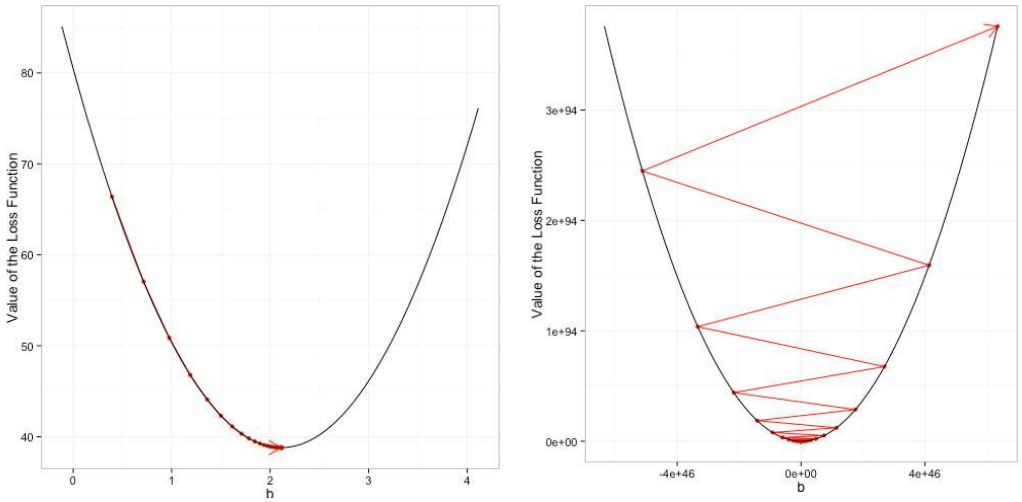
Bonus experiment

By varying over **alpha parameter,** following results were obtained:

activation	Alpha	Solver	Accuracy
Identity	0.001 0.008	sgd	0.41 0.79
ReLu	0.008	sgd	0.776
Sigmoid	0.0011 alpha : 0.008 alpha : 0.005	sgd	0.095 0.109 0.635

Observations and Inferences

- In case of **activation functions** with high learning rate (i.e., 0.1 and 0.2), reLu and sigmoid, both performed very bad even on varying the number of hidden nodes and number of hidden layers because considering the gradient descent algorithm, it is possible that the optimizing cost function may get stuck in local optima. Also the function may not converge with high learning rate, as shown in the figures below.



Hence, with lower learning rate, accuracies are improved for both the cases of **sigmoid** and **ReLu activation.**

Question 3: Bagging and Boosting

Bagging

estimator	Accuracy
N = 10	0.289
N=5	0.273
N=15	0.328
Neural Network(200, 150)	0.255

- In case of default neural network, the accuracy obtained is only 25%.
- By using the bagging concept with estimator, Neural network with a hidden layer of 50 nodes, accuracy tends to improve with increase in number of estimators in the bagging.