DISCUSSIONS

Question 1:

Objective: To perform 2 class classifications between 1 and 7 on the MNIST dataset.

Task: Classification is performed using SVM and 5 layers Neural Network. A 5 layer neural network with architecture: [128 -- 128 -- 128 -- 64 -- 1)]

Performance Metrics:

1. Means and Standard Deviation,

2. ROC and EER(Equal Error Rate)

3. Precision-Recall Curve

Inferences from the observed results:

1. ROC Curves summarize the trade-off between the true positive rate and false-positive rate for a predictive model using different probability thresholds.
2. Precision-Recall curves summarize the trade-off between the true positive rate and the positive predictive value for a predictive model using different probability thresholds.
3. ROC curves are appropriate when the observations are balanced between each class, whereas precision-recall curves are appropriate for imbalanced datasets.

Assume you have a "positive" class called 1 and a "negative" class called 0. Y^ is your estimate of the true class label Y. Then:

- Precision=P(Y=1|Y^=1)
- Recall=Sensitivity=P(Y^=1|Y=1)
- Specificity=P(Y^=0|Y=0)

The key thing to note is that sensitivity/recall and specificity, which make up the ROC curve, are probabilities conditioned on the true class label. Therefore, they will be the same regardless of what P(Y=1) is. Precision is a probability conditioned on our estimate of the class label and will thus vary if we try our classifier in different populations with different baseline P(Y=1). However, it may be more useful in practice if we only care about one population with a known background probability and the "positive" class is much more interesting than the "negative" class. This is because it directly answers the question, "What is the probability that this is a real hit given my classifier says it is?"

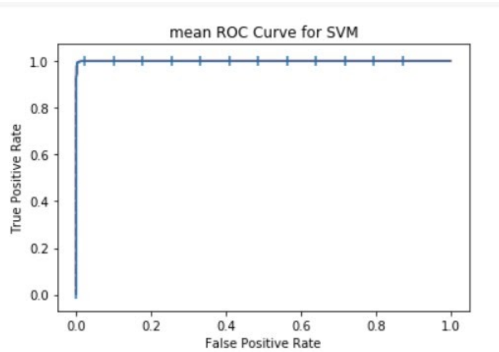
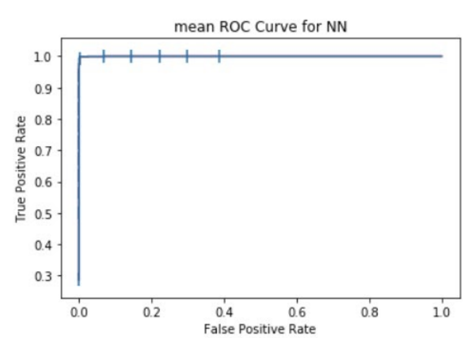
So, if our question is: "How meaningful is a positive result from my classifier given the baseline probabilities of my problem?", we shall use a PR curve and If our question is, "How well can this classifier be expected to perform in general, at a variety of different baseline probabilities?", we shall go with a ROC curve.

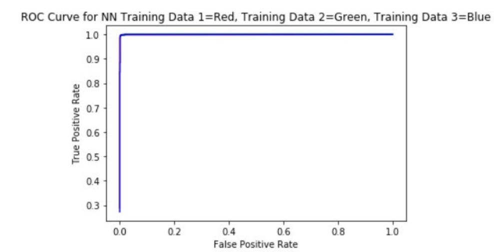
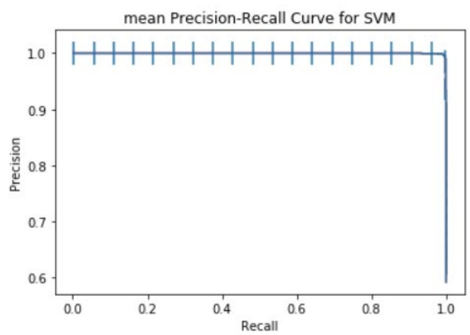
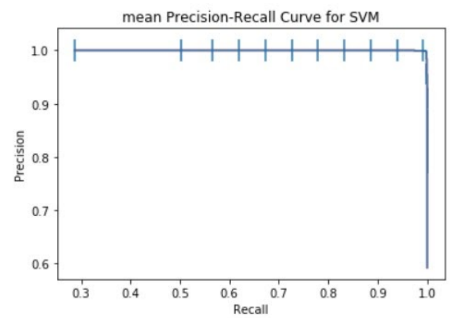
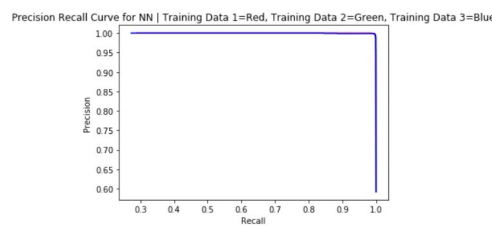
RESULTS

QUESTION 1:

Sr. No.	SVM	Neural Network

Confusion Matrix	Train Data 1 — `([[1877,    0],`
	`[   35, 1258]])`
	Train Data 2 — `([[1876,    1],`
	`[   38, 255]])`
	Train Data 3 — `([[1877,    0],`
	`[   30, 1263]])`
		Train Data 1 — `([[1876, 1],`
		`[   16, 1277]])`
		Train Data 2 — `([[1876,    1],`
		`[   22, 1271]])`
		Train Data 3 — `([[1876,    1],`
		`[   17, 1276]])`

Accuracy and Deviation	Accuracy and Deviation for Training Data 1)98.9+ -0.006666666666674814 2)98.77+ -0.13666666666668448 3)99.05 + 0.14333333333331666 ============================= Accuracy for SVM : 98.91% =============================	Accuracy and Deviation for Training Data 1)98.9+-0.48666666666665037 2)98.77 + -0.6166666666666 3)99.05 + -0.3366666666666589 ============================= Accuracy for SVM : 99.39 % =============================
Receiver Operating Characteristics (ROC)		

		
Precision-Recall	 	 
Equal Error Rate