Assignment 5

Swapnil Asawa, swa12@pitt.edu

Question 1

Please see hw5.m for implementation steps.

a) Done in hw5.m

SVM

W =

0.0445
0.0116
-0.0050
0.0003
-0.0002
0.0264
0.3524
0.0063

b = -2.7182

b) Submitted **apply_svlm(x, w, b)**

c) Done in hw5

Training Confusion Mat=

299 40
85 115

Testing Confusion Mat =

142 19
26 42

Misclassification Error Training =0.2319
Misclassification Error Testing = 0.1965
Sensitivity on train = 0.5750
Sensitivity on test =0.6176
specificity on train= 0.8820
Specificity on test = 0.8820

For logistic Regression:

Confusion Matrix for Training data

263 76
89 111

Confusion Matrix for Testing Data
121 40
22 46

Misclassification Training Error= 0.3061
Misclassification Testing Error =0.2707
Training Sensitivity: 0.5550
Testing Sensitivity=0.6765
Training specificity=0.7758
Testing specificity=0.7516

Naive Bayes

Confusion Matrix for Training Data
289 50
79 121

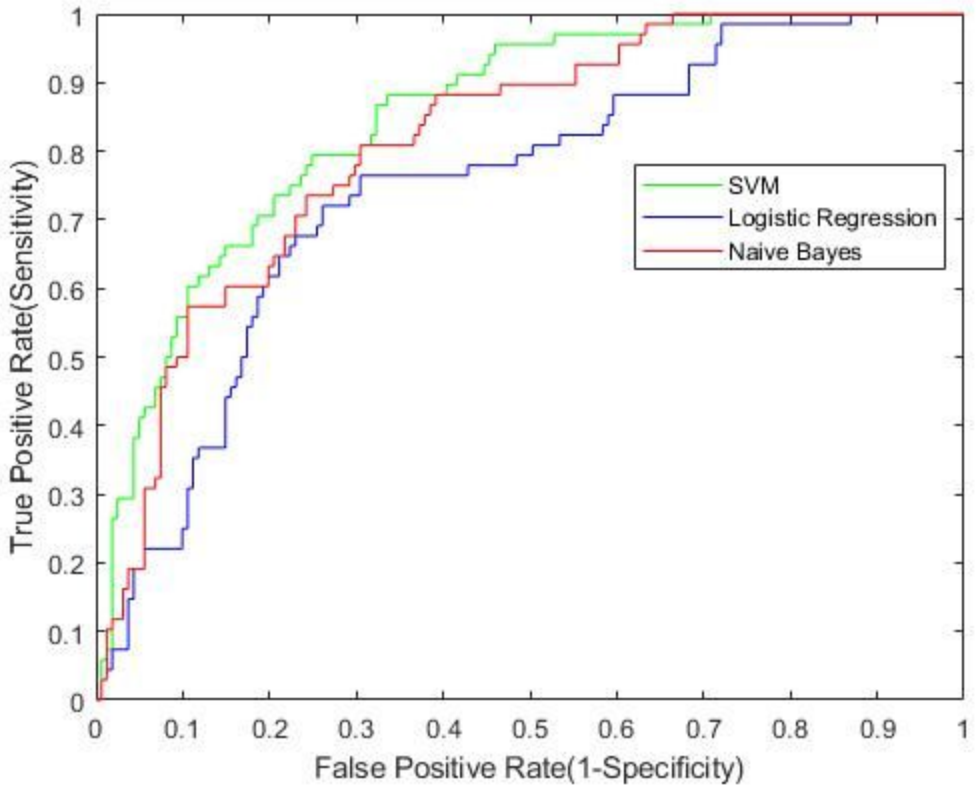
Confusion Matrix for test data
138 23
29 39

Misclassification Training error = 0.2393
Misclassification Test error=0.2271
Sensitivity train =0.605
Sensitivity test = 0.5735
Specificity train =0.8525
Specificity test =0.8571

d)Misclassification error in training as well as testing is lowest in SVM
Sensitivity on train is most in Naive Bayes, but the testing sensitivity is highest in logistic regression, followed by SVM and ast is Naive Bayes.
Specificity of the training as well as in testing is highest in SVM.
I think SVM works out to be the best, even though it has a little less testing sensitivity. It also depands on the application. If higher sensitivity is needed than SVM would not be good whereas when high specificty is needed, SVM works best. Also the misclassfication error is least in SVM.

Question 2:

a) Done
b) Done
c) See next page



AUC(Area under curve)

SVM	Logistic Regression	Naive Bayes
0.8497	0.7487	0.8150

The more the area under curve, the better.

AUC of SVM is the most, so SVM performed the best. Also, the less the overlap between predictions, True Positive rate is higher at low False Positive Rate, and the better. Highest curve is of SVM. Thus again, SVM performed best.