Data Mining | Fall 2021
Tu, Th 2:00 PM – 3:15 PM | HW6 Programming Optional

1. [3pts] Given the following data set

x1	x2	Class
3	2	Y
2.5	2.25	Y
0.5	3.25	Y
2.75	2.125	Y
2.25	2.375	Y
1	3	Y
3.5	1.75	N
5	1	N
6	0.5	N
4.5	1.25	N

This data set in csv format is attached (data.csv).

- [1pt] Construct svm model for this set using e1071- package svm procedure
- [1pt] Plot the results, use 2 dim plot of SVM features (see example in http://ugrad.stat.ubc.ca/R/library/e1071/html/plot.svm.html)
- [1pt] What is the accuracy of prediction? How do you explain the classification? By visual inspection of points on the plain, should this set be separable in $\mathbb{R}^2$?

2.[3pts]Use this model to

- [1pt] Predict the class of the following data set.

x1	x2	Class
1	8	Y
2.5	3	Y
3	6	Y
1.5	1	Y
2	4	Y
3.5	5	Y
6	3	N
3	1	N
4	3	N
3.5	2	N

The data set in csv format is attached (data1.csv)

- [1pt] Find its accuracy is on the test set?
- [1pt] Plot the classified data. By visual inspection: can this set be separated by the 'correct' linear hypothesis?

3. [4pts] Compare SVM and Bayesian classification on ionosphere data (data("Ionosphere") from mlbench library. Read data description to see what it is (Ionosphere)

- [1pt] Clean the data from datapoints with undefined values. Prepare testing and training sets. Use 2/3 of all records for training and 1/3 for testing.
- [1pt] Remove columns 1 and 2 from the data. These columns are nominal - SVM can only deal with real data (in fact ordinal that can be imbedded into real will work too). Construct SVM model and perform the prediction on testing data using constructed model.
- [1pt] Construct Naïve Bayes model and perform prediction using constructed model
- [1pt] Compare accuracies of the models. Does one model perform better on ets set than another? If so why?