Week 7
Activity

The idea of this activity is to compare two different classifiers (naive Bayes and Decision Trees (DT)). Open your preferred Data Mining Software (Weka, R, SAS, SPSS, etc), load the breast cancer data set. For this follow:

library(mlbench)

data(BreastCancer)

Answer the following questions using plots and paragraphs, you must include all trained models too.

1. Randomly split the data in training and test set, explain your separation (you can use cross validation too).

   Decision tree

2. Learn a naive Bayes and a DT, and compare them based on the training set, using the confusion matrix to compute two evaluation metrics. Explain your conclusions.

3. Test both models in the test set, and compare them using the confusion matrix to compute two evaluation metrics (ROC curve gives 0.5 extra points). Explain your conclusions.

4. Send your document with your classmates names, and also send the models.