Project II:

Building classifiers using a function estimator.

The objective of all these projects is to construct a classifier. The classifier should be a good class predictor for future observations.

In order to build the classifier, you need to define a feature space, and a partition of the feature space.
These are the steps:

1. Define the set of features for prediction.
   The data = features + response-s.

2. Split the data into training and testing sets.

3. Use a function estimator (SVM, Random Forest, Ada Boosting, Logistic Regression, Deep Learning, GLM Net) in order to compute the probability distribution of the labels for any observation, as a function of the feature values. Select the best set of features or best model -- apply cross-validation when possible.
4. Write a function with
   Inputs: x=Data table with predictors, y=response labels, percent testing(default 20%), list of methods among the six above that you want to apply, extra parameters)
   Output:
   1. Table with summarizing the performance of the methods.
      Produce a table of performance. The rows of the table are the 6 methods and the columns are measures of performance: 1-2.

Misclassification rate on training and testing sets. 3. Area under the ROC curve on the testing set.

2. List of fitted objects.

5. Data example: Plastic explosive detection dataset.

Data Set: Pex23

The data comes from a study for the detection of plastic explosives in suitcases using X-ray signals.

The 23 variables are the discrete xcomponents of the xray absorption spectrum.

The response is the last variable in the dataset. It takes two values:

0: There is explosive

1: There is not.

The objective is to detect the suitcases with explosives. (See textbook)

Run the function on the Pex23 data. Produce the output table and interpret the results. Can you say anything about the partition induced by the classifiers on your feature space? How easy is to interpret the results?