Springboard DM&ML - Module Assignment (40%)

Due : Fri. 6th December.

Choose data sets from the UCI Machine Learning Repository or elsewhere. http://archive.ics.uci.edu/ml/. You will probably need two data sets, one for regression and one for classification. You will carry out three analyses.

Write a report. The report should document your analyses for

- Regression
- Decision Trees
- kNN

For each technique:

1. Give an overview of the problem.
2. Data exploration (tables and graphs)
3. Definition of training and testing sets
4. Model generation & information on the model.
5. Predictions for the test data
6. Evaluation of the model(s) and conclusion.

Section headings in the report should be numbered from 1.1 to 3.6

Submit a single zipped R project (folder). Include

- data sets in a "data" folder
- R scripts
- doc folder containing
  - report as a PDF file
  - scan of signed student plagiarism disclaimer form

Note

- R project means there should be .Rproj folder
- Don't include a tutorial/explanation on the methods in the report.
- Don't use data sets we have used in the notes or labs.
- Use relative path names in your R code.
- Comment on graphs and summaries.
- Use more than one independent variable in linear regression.
- How is model to be used?