Homework 3

This is your first homework assignment working in teams. It is meant as much an exercise in fitting algorithms as in working together.

The following data set comes from UCI Machine Learning Repository:

https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data

The file https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.names contains description of the data, variable names and possible values.

Our goal is to predict whether a person makes over 50K a year. Use the first half of the data as a training data. Explore application of various classification methods, models and algorithms that we had studied in class. Use the second half of the data for validation. Comment on the results.

You can make it as complex or as simple as you want it to be but keep in mind that your final report needs to make a coherent story. You are not obligated to explore every model or algorithm we had studied but do choose a justifiable subset.

Brainstorm the plan of the attack with your team first. Assign parts to group members and crosscheck work later. We will be using github for version control and as soon as we have all your aliases, we will give you permission for the appropriate repository. We will later check your usage of it, we suggest you keep all your work in it, including all intermediate steps.

Please refer to the following homework guidelines for the format of report:

- **no more than 8 pages (including code and graphs) -if it doesn't fit, use smaller graphs or par mfrow**
- **only code essential in showing how you achieved your solution should be included in the report**
- **graphs should have title, meaningful axis labels, and captions - but try to make the graphs as self explanatory as possible**
- **points will be deducted for poor formatting and convoluted explanations (it's a report, try to demonstrate the concepts you are working with as clearly as possible)**