Project 3: Classification Algorithms

Due: Code and Report should be submitted by 5pm on Dec. 11.
Demo is on Dec. 12.

Two datasets can be found at
http://www.cse.buffalo.edu/~jing/cse601/fa13/docs/project3_dataset1.txt and
http://www.cse.buffalo.edu/~jing/cse601/fa13/docs/project3_datast2.txt. A short description
is at http://www.cse.buffalo.edu/~jing/cse601/fa13/docs/README.txt.

Complete the following tasks:

- Implement three classification algorithms: **Nearest Neighbor, Decision Tree,** and **Naïve Bayes**.
- Implement **Random Forests** based on your implementation of Decision Tree.
- Adopt 10-fold **Cross Validation** to evaluate the performance of all methods on the provided datasets in terms of **Accuracy, Precision, Recall**, and **F-1 measure**.

Your final submission should include the following:
- Code: Implementation of four methods. Together with your code submission, a README file should be included to explain how to execute your code.
- Report: Describe the flow of all the implemented methods, and describe the choice you make (such as parameter setting, pre-processing, post-processing, how to deal with over-fitting, etc.). Compare their performance, and state their pros and cons based on your finding.

The details about Demo will be released at **5pm on Dec. 11**. Please note:
- New datasets will be given to check your implemented classification methods and performance measures. The data format will be the same with the ones we already provided.
- During the demo, you will be asked to adopt specific setting and run your code.

Note that copying code/report from another group or source is not allowed and may result in an F in the grades of all the team members. Academic integrity policy can be found at
http://www.cse.buffalo.edu/shared/policies/academic.php