

Project 3: Classification Algorithms

Code and Report Submission Due: December 10 2016 1:59PM

Demo: December 10 2016 2:00PM ~ 6:00PM

Hard Copy Report: bring to your demo.

Please clearly state the UB Person numbers and UB IT names for all the group members on the cover of the report.

Two datasets (*project3_dataset1* and *project3_dataset2*) can be found on Piazza. Please check the README file first for a short description of the two datasets.

Complete the following tasks:

- Implement three classification algorithms: **Nearest Neighbor**, **Decision Tree**, and **Naïve Bayes**.
- Implement **Random Forests** based on your own implementation of Decision Tree.
- Implement **Boosting** based on your own implementation of Decision Tree.
- Adopt 10-fold **Cross Validation** to evaluate the performance of all methods on the provided two 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 findings.

The details about Demo will be released at **3pm on Dec. 8 through Piazza**. Please note:

- New datasets will be given to check your implemented classification methods and performance measures. The data format will be consistent with the README file that we already provided.
- During the demo, you will be asked to adopt specific setting and run your code.

Note that copying code/results/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>