Project 3: Classification Algorithms

Due: Dec. 13 at 11:00am.

The training and test data sets can be found at:
http://www.cse.buffalo.edu/~jing/cse601/fa12/docs/training.txt and
http://www.cse.buffalo.edu/~jing/cse601/fa12/docs/test.txt). A short description of the two datasets can be found at http://www.cse.buffalo.edu/~jing/cse601/fa12/docs/README3.txt

Complete the following tasks:

- Implement one classification algorithm to learn a classification model from the training data. The algorithm must be one of the following: Decision Tree, Naïve Bayes, Logistic Regression, Rule-based Classification.
- Apply the classification model you developed and some other classification methods to the test data. Combine the results of multiple classification models by majority voting or your own combination approach. You can use existing classification package for the other models. The places to download possible packages can be found at: http://www.cse.buffalo.edu/~jing/cse601/fa12/docs/README3.txt
- Evaluate the classification performance (accuracy) of all the base models (including your own algorithm) and the combined model on the training data. Apply the combined model (the model that combines your algorithm and several other algorithms) on the test data and submit the prediction results. We will evaluate the accuracy of your predictions. The performance obtained by all the teams in the class will be ranked according to the accuracy on test data.

Your final submission should include the following:

- Code: Implementation of classification algorithm and combination approach. Together with your code submission, you should include a README file explaining how to execute your algorithms to reproduce the results.
- Result file: One single file consists of the predicted labels of the test data. Each row should only have one character (B or M) which indicates the predicted label of the corresponding data point.
- Report: Describe the flow of all the implemented algorithms and the motivation of using the chosen base models and combination scheme. Compare the performance of the base models as well as the combined model on training data. Describe your design of the scheme that can help improve performance on the test data.

Please pack code, result file and report into one zip file and submit it to the server.

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