

## Assignment 5 (10%): Ensemble Learning. (Due on 16:59:59, Nov. 30, 2016)

---

Please implement the AdaBoost algorithm for classification task. Before implementing AdaBoost, you should implement a base classifier first.

### Dataset

1. Data: We adopt two binary classification datasets which come from [UC Irvine Machine Learning Repository](#). (Please use editors like Sublime or Notepad++ to open them.)
  - a. [Dataset 1 Download](#)
  - b. [Dataset 2 Download](#)
2. Data Description and Format: Both of them are binary classification datasets. **The first row indicates the type of each feature, i.e., 1 representing discrete feature and 0 representing numerical feature.** Except for the first row, each row represents an example. Except for the first row, the last column represents the label of the corresponding example, and the remaining columns represent the features of the corresponding example.

### Task Description

Please implement the naive Bayes classifier first, and then implement the AdaBoost algorithm which use the naive Bayes classifier you have implemented as the base classifier. Please implement the algorithms (both the naive Bayes classifier and the AdaBoost algorithm) by yourself, and do not invoke other existing codes or tools. But, for random number generation, you may use the existing codes or tools.

**Task 1:** The Naive Bayes Classifier for Binary Classification. (Reference: Section 10.5.1 of the [textbook](#))

1. Please handle both discrete and numerical features.

**Task 2:** AdaBoost for Binary Classification. (Reference: Section 11.8.3.3 of the [textbook](#))

1. Implement the AdaBoost algorithm using the naive Bayes classifier you have implemented in the Task 1 as the base classifier.
2. Conduct [10 fold cross validation](#) on dataset 1 and 2 for AdaBoost, and then report the mean and standard deviation of accuracy.

How to present the results:

**The report** should contain a table that shows the mean and standard deviation of accuracy for AdaBoost on dataset 1 and 2, respectively. Please arrange the table in a proper way. Besides, the report should also contain a description of the method that you apply to handle both discrete and numerical features.

### Reminders about Submission

1. **The report must contain your name, student ID, and e-mail address.**
2. Submission Deadline: **2016-11-30 16:59:59.**
3. Before submitting your assignment, please read **Submission Requirement and Description** section above carefully and obey it.
4. For assignment 5, please pack your **report**, **code** and **ReadMe.txt** into a zip file named with your student ID, e.g., MG1633001.zip.

