Machine Learning: Midterm Exam

Due on 11/25/2015 in class

Announcement: Please follow the guidelines for writing your report:

- The presentation should be as formal as possible (with a title page, names and IDs, and necessary outputs and pictures);
- Describe the basic ideas of each method you use and interpret clearly the results;
- Do not attach any R code in your report;
- Writing either in English or Chinese is fine.

The "**wine.txt**" data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines. For detailed descriptions of the data please refer to "**wine_information.txt**".

Q1: Suppose our goal is to develop the decision rules for classifying wines based on the cultivars. Construct the decision rules using: (i) Classification Tree; (ii) LDA; (iii) QDA; (v) Nearest Neighbor; and (vi) SVM to classify the wines. Evaluate and compare all the obtained decision rules, which one will you recommend?

Q2: Suppose there are 6 wines from unknown cultivars with the measurements of 13 constituents shown in "**wine_test.txt**". Based the "classification tree" you obtained in Q1, perform two stable procedures "bagging" and "boosting" to classify the 6 wines. Are the classifications of the 6 wines the same as those obtained from the five decision rules in Q1 ?

Q3: Exclude the first column (class-id) of the "**wine.txt**" data, perform a cluster analysis using (i) one hierarchical tree method (with best selected linkage); (ii) one partitioning method; and (iii) the self-organizing maps.

Q4: Do the clustering results in Q3 recover the original classification of the wine based on class-id ?

(Grading Policy: A+ = 95, A = 90, A- = 85, B+ = 80, B = 75, B- = 70, C+ = 65)