M435 Project I

Classification and Feature Selection

Due Wednesday, October 15, 2003

This project concerns the classification problem and the related feature selection problem. The primary application is to the Human Computer Interface problem, one long-term goal of which is to be able to classify modes of mental thought via recording EEG signals. In this project you will consider two different thought classes, i.e., mental arithmetic versus an imagined writing task. These tasks were selected given that they should, at least in theory, activate different cerebral lobes. The hope is that a good classification algorithm should be able to separate these classes. Your project should apply the classification approaches developed in class, as described below, as well as the singular value decomposition for feature selection.

The **assessment criteria** for all projects include

- Quality of project report with an emphasis on organization and appearance.
- Correctness and completeness of report.
- Creativity of project.
- Demonstrated added value, summary of weaknesses of approach, potential future work.

Step 1

To establish the correctness of your code and your theoretical mastery of this material please include the results of the singular value decomposition worksheet in your description of the method. Include the results of applying k-means and k-nearest neighbors classification to the image of Saturn.

Step 2

Download the data for two math trials and two letter trials. Use the same subject in each case. This data is available directly from the computer science

department's web page or from the class web page.[1] Construct a new data matrix from the raw data that treats each point as a *window* of data that includes 128×7 values. This corresponds to 1/2 second of recording time. These windows can be staggered by 1/4 second intervals.

Step 3

Apply the singular value decomposition to reduce the data to a manageable size. Your cleverness here may greatly impact your classification rates.

Step 4

Now employ k-means and k nearest neighbors for classification. Compute sensitivity versus specificity plot as described in class to assist in determining an appropriate value for k in each case. Also compute the accuracy of the classifier. One data set for each task should be used to build the classifier and the second data set should be used to test the classifier (by computing sensitivity, specificity and accuracy).

Step 5

Write-up your results using Latex and following the sample on the class web site. The final write-up is extremely important in the assessment of the project and should not be left to the last minute. In fact, the process of constructing the document should begin even before you start writing code. Here added value should include how well the classification works on raw data.

Step 6

One student in each group should prepare a 20 minute presentation for October 8. (All students will be required to do this.) This presentation should summarize the group project but may emphasize the students contribution.

[1] Note that you may select other tasks and trials if you desire for this EEG classification project.