STAT244 **ASSIGNMENT 1** **SPRING, 2011**

due 2/11/11

1 Power Method for Eigenvalues and Eigenvectors

Use the power method to iteratively find the eigenvalues and eigenvectors of a real, symmetric 5×5 matrix of your choice. If you don't have an orthogonalization routine, you can use the Gram-Schmidt program in the file `~s244/samples/gs.c` . Also extract the eigenvalues and eigenvectors using some standard routine (EISPACK, LAPACK, etc.) and compare the results to those from using the power method, both numerically and with regard to execution time.

2 Cluster Analysis using the Leader algorithm

Write a program to implement the leader algorithm. Your program should try a variety of threshold values, and calculate the within cluster sum of squares for each threshold used, as well as listing the number of clusters found, and the members of each cluster. Examine the relationship between the threshold you use and the sum of squares to see if some "natural" number of clusters emerges. Then compare your results to the results of a cluster analysis method of your choice using `R`, `sas`, `matlab` or some other program. You can use the dataset crime in the `~s244/samples` directory or class web page to test your program, or you can use a data set of your choice. (The file `~s244/samples/crime.des` describes the contents of the crime dataset).

For extra credit, use the clustering from the leader algorithm as the starting point for a k-means algorithm. What kind of improvement in within cluster sum of squares does the k-means algorithm provide compared to the leader algorithm?