Michael Freeman

Tim Chartier

CSC 381

October 21, 2019

Summary

In order to cluster the data that was collected, I choose two methods: Fiedler’s Method and the Singular Value Decomposition (SVD). In both methods I needed to create a matrix that contains all of the relationships between all of the courses. Before I could create these two matrices and get into the methods, I needed to have the data in the correct form. I used MATLAB to create vectors that contained the variances of yardages on Par 3’s, 4’s, and 5’s. Then I added the vectors together to have a total variance number. This vector was what I used to create the matrices described below.

In Fiedler’s method I choose to create the connections by allowing a connection if the absolute value of the difference in total variance was less than the average variance of all the courses divided by four. Otherwise the corresponding value in the matrix would be a zero. This creates a matrix with only 1’s and 0’s in the entries. For example, if the average variance is 8 then the difference between two course’s variances must be 2 or less. So if course A has a variance of 10 and course B has a variance of 11, then a 1 is entered in the matrix in position (A, B) and (B, A). If course B had a variance of 13 then those same entries would be 0. After completing this for all the course we have an adjacency matrix. The code used for class was adapted to cluster this data. The code first takes the adjacency matrix (A) that was created with the 0’s and 1’s and sums each column and places the sum along the diagonal, while all other values are zero to create a diagonal matrix (D). Then a Laplacian matrix (L) is formed by computing L = D – A. By Fiedler’s method we will find the eigenvector of L with the second smallest eigenvalue and then each entry in the column is sorted, so that the positive signs go together, and the negative signs go together. To get more clusters we used the eigenvector with the next biggest eigenvalue as well.

In the SVD method I needed entries in the matrix to represent the distance between each course. So, for example if course A has a variance of 10 and course B has a variance of 20, then the number 10 would go in the entries of the matrix (A, B) and (B, A), so the matrix is symmetric. This creates a more accurate comparison between variances of each course. To start the clustering process, we divide each entry in a column by the column’s mean. Now that we have a matrix we can use the SVD to produce a matrix M where *M* = *UV*T . The columns are U are the left singular vectors and the columns and *V*T are the right singular vectors. Next we find the singular vector which corresponds to the largest singular value. Then each entry in the column is sorted, so that the positive signs go together, and the negative signs go together. For this project we adapted code written for class. In order get the 7 clusters shown in the data more than largest singular value was used.

Overall my project utilized linear algebra in three ways. The first way was very simple in that the vectors containing individual variances of the different types of holes were added together. The next way was using Fiedler’s method to cluster the data. This method used many types of matrices including a diagonal and Laplacian matrix as well as the principles of eigenvalues and eigenvectors. The SVD a decomposition of the original matrix and found the largest singular values and corresponding singular vector. These clustering methods and the ability to do arithmetic with matrices allowed this project to produce accurate results and provide insight into which clustering algorithm works the best.