INF552: Programming Assignment 3

Part 1: Implementation [7 points]

PCA (2 points)

Use PCA to reduce the dimensionality of the data points in **pca-data.txt** from 3D to 2D. Each line of the data file represents the 3D coordinates of a single point. Please output the directions of the first two principal components.

FastMap (5 points)

Use FastMap to embed the objects in **fastmap-data.txt** into a 2D space. The first two columns in each line of the data file represent the IDs of the two objects; and the third column indicates the symmetric distance between them. The objects listed in **fastmap-data.txt** are actually the words in **fastmap-wordlist.txt** (n^{th} word in this list has an ID value of n). The distance between each pair of objects is the Damerau–Levenshtein distance between them. Plot the words on a 2D plane using your FastMap solution.

You can write your program in any programming language. However, you will have to implement the algorithms yourself instead of using high-level library functions - except for computing eigenvectors and eigenvalues. Provide a description of the data structures you use, any code-level optimizations you perform, any challenges you face, and of course, the requested outputs.

Part 2: Software Familiarization [Optional - No Credit]

Do your own research and find out about library functions that offer good implementations of PCA and FastMap. Learn how to use them. Compare them against your implementations and suggest some ideas for how you can improve your code. Describe all this in your report.

Part 3: Applications [Optional - No Credit]

Do your own research and describe some interesting applications of PCA and FastMap.

Submission Guidelines

In your report, please include the names of all group members and mention their individual contributions. The maximum number of the members in a team is 2. The report should be in PDF format. Your submission should include the code as well as the report. It is due before **10/08**, **11:59pm** in an archive in zip, tar.gz or tar.xz formats. Only one submission is required for each group by one of the group members. Please submit your homework on **BlackBoard** (do NOT email the homework to the instructor or the TA).