Clarification Questions

Please ask any question related to this assignment in the class or in the Moodle forum. To be fair to everyone, it is best to ask assignment question openly rather than in email. Do not share your code with anyone. Do not post your code online. Do not ask for help in any public discussion forum (e.g., Stackoverflow or developer forums). You can search over the internet and read books. Let me know if you find any error in the assignment question as soon as possible. DO NOT SHARE OR DESCRIBE YOUR OUTPUT TO ANY OTHER STUDENT. DO NOT ASK DIRECT QUESTIONS FOR VERIFYING YOUR VISUALISATION: 'DOES MY OUTPUT LOOK CORRECT?' However, you can always ask questions about the methods, definitions or approaches to solve a question. Any case of plagiarism will be reported to the Department of Computer Science.

Assignment 2

Q1. Cluster Visualization by MDS (50 Marks)

You are given a data file that contains a sample survey result (regarding whether/how smartphone use influence family life) in form of a csv file (a3.csv). The qs.pdf contains the detailed information of the survey questions.

Your task is to apply MDS to find a scatterplot of a 5 dimensional dataset [Q64, Q44,Q37,Q35,Q34]. You must edit A3MDS.html to complete this assignment. You are already given the mds.js and numeric.js that compute the MDS-coordinates for you. You dont have to worry about them at all. If you look at the top part of the html file, then you will find out that these files are already included in your project. So keep them along with your html (in the same directory).

You will first need to produce the squared distance matrix (see the lecture slide example). Represent it as a 2d array. Each entry i,j of the distance matrix needs to be filled with a distance between the response of person i and person j.

If everything is correct, you can refresh your browser to see a scatterplot, where the colors of each point is assigned based on the Q9age. You can see that the older age group forming small clusters, whereas younger groups have high variance in their responses.

Submission Instruction

Submit the solution as Q1.html and a screenshot of the plot generated.

Q2. Cluster Visualization by PCA (20 Marks)

Your task is to apply PCA to find a scatterplot of a 5 dimensional dataset [Q64, Q44,Q37,Q35,Q34]. You must edit A3PCA.html to complete this assignment.

If everything is correct, you can refresh your browser to see a scatterplot, where the colors of each point is assigned based on the age. This will be slightly different than the plot of Q2, but I think the clusters will look only slightly better. You can see that the older age group forming small clusters, whereas younger people have high variability in their responses.

Submission Instruction

Submit the solution as Q2.html and a screenshot of the plot generated.