Homework on clustering

This week's assignment is the practical application of your clustering knowledge.

Occupational health clinic has the following parameters from 106 subjects:

1. Age
2. Body mass index
3. Physical health (Occupational stress questionnaire)
4. Psychological health (Occupational stress questionnaire)
5. Index of depression

You might not need all!

Background: The clinic has the electroencephalographic (EEG) signal available from all those patients. The aim of the clinic is to find some objective features from EEG which would characterise the subjects' brain state and could be used later for brain state evaluation without the need to fill in those long subjective questionnaires.

Assignment: The aim for you is to group the subjects into clusters, so the clinic can later perform further EEG signal processing on those clusters.

The variable data is in the file "data.mat". Each row in the file represents one subject and each column represents one parameter (age, ..., index of depression).

Taking into account all you have learned about clustering, perform the k-means clustering and evaluate the results. Figure out the number of clusters and the selection of parameters most suitable for the assignment, justify your decision. Cluster the subjects as well as you can. Give a short overview of your used methodology and the results, describe the composed clusters considering their parameter values. Were there some parameters, which did not improve the clustering and were rejected? Can you assume, what were the reason for that?

MatLab functions you will need: `kmeans`, `silhouette`.

Please upload the MatLab code and your report (comments, justifications) to Moodle.

Good luck!

Maie