

Exercise 22 - Clustering

Please implement the clustering algorithms discussed in the lecture.

- **(4 Points)** Complete the function `kmeans_diy` for the Lloyd's algorithm for k-means clustering.
- **(4 Points)** Complete the function `spectralclustering_diy` for the spectral clustering algorithm with unnormalized graph Laplacian:

$$L = D - W$$

where D is the **degree matrix** and W is the **weighted adjacency matrix** of the graph. The similarity graph here is set to be the fully connected type, i.e., we construct all points with positive similarity with each other and we weight all edges by

$$s(x_i, x_j) = \exp(-\|x_i - x_j\|^2/\gamma)$$

where x_i, x_j denote the i -th and j -th data points in the dataset, γ is fixed to be 1 in this exercise.

- **(3 Points)** Now apply your clustering code (k-means and spectral clustering) to the provided dataset `ex9_data` and visualize your results by running the script `ex9_main.m`. The data file contains two datasets `data1` and `data2`, where each includes two groups of data points X_1, X_2 in R^2 , you need to apply the clustering on `data1` and `data2` separately. What's your observation and conclusion? (Written solution on paper.)

Hints:

- for spectral clustering, you can use `eig` function to compute the eigen vector and eigen value of a matrix. You are encouraged to read the tutorial from Ulrike Von Luxburg to get a better understanding of spectral clustering. The tutorial can be downloaded from http://www.informatik.uni-hamburg.de/ML/contents/people/luxburg/publications/Luxburg07_tutorial.pdf
- you are recommended to implement a function `compute_pairwise_distance` to compute pairwise distance between two datasets which can be shared by both of your clustering algorithms' code.

Submission:

- Create **one** zip-file containing the m-files (`kmeans_diy.m`, `spectralclustering_diy.m`) and send the file to your tutor. The filename has to follow the following convention: `[group:A,B,C]_[matrikel numbers separated by underscore]_ex[nr].[extension]` e.g. if you are in group B and your team members have matrikelnumbers 3503239, 3028258 and the current exercise number is 14 then the filename reads: `B_3503239_3028258_ex14.zip`.