Clustering Methods

Exercises 7/7

Find the following papers on web:

M. Rezaei and P. Fränti, "K-sets and k-swaps algorithms for clustering sets", *Pattern Recognition*, 39, 109454, July 2023.

M.I. Malinen, R. Mariescu-Istodor and P. Fränti, "K-means*: clustering by gradual data transformation", *Pattern Recognition*, 47 (10), 3376-3386, October 2014.

P. Fränti, "On the usefulness of self-organizing maps for the clustering problem in vector quantization", 11th Scandinavian Conf. on Image Analysis (SCIA'99), Kangerlussuaq, Greenland, vol. 1, 415-422, 1999.

If difficulty to find them, here is link: https://cs.uef.fi/~franti/papers.html
Read the relevant parts and find answers to the questions below.

1. Gradual k-means selects k centroid locations (straight line below) and then converts all datapoints randomly into each of them so that every data point has exactly the same value as its centroid. Gradual k-means then starts to move the datapoints gradually to their original locations while processing the clustering by k-means. Show the first transformation of the algorithm for the data below (k=4). Hint: you can freely assume the original locations of the 2-d datapoints.

A B C D

2. Below are two travelers, Lisa and Albert, and the countries they have travelled to. They have been assigned to the same cluster. Show the cluster representative and the assignment costs. Do you think they belong to the same cluster?

3. Suppose two travelers, Agneta and Dara, who have traveled to the following countries: Agneta = { US, Canada, UK } and Dara = { Bulgaria, Montenegro, Malta, Switzerland }. Which one of them would be more likely to join the same cluster with Lisa and Albert?

4. What is the reason SOM cannot find good clustering results. Do you have any concrete example to demonstrate its weakness?