UiT
THE ARCTIC UNIVERSITY OF NORWAY

6. Unsupervised Learning and PageRank

FYS-2021 Exercises
Department of Physics and Technology
Faculty of Science and Technology

K-means

Problem 1

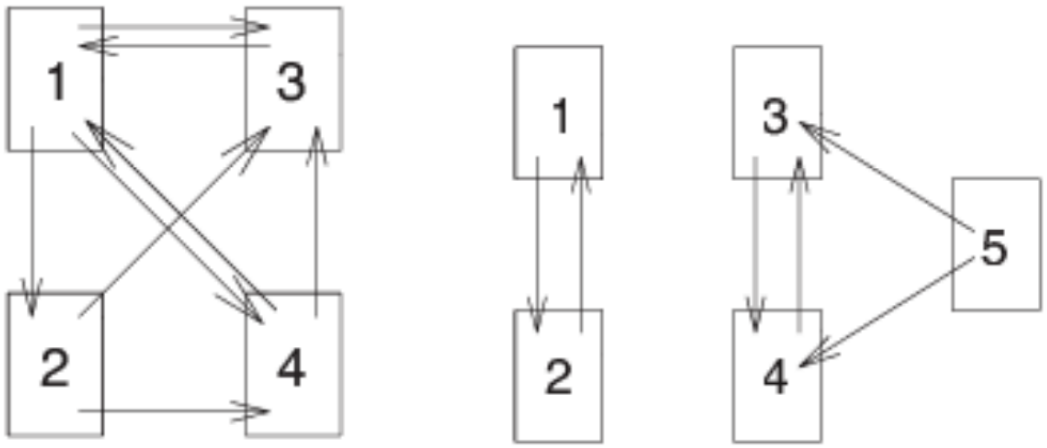
(1a) Implement your own version of the k-means algorithm.

(1b) Test your implementation on the datasets provided in `blobs.csv` and `flame.csv`. Plot and comment on the results.

(1c) Test your implementation on the optdigits dataset provided in `optdigits.csv`. Plot the centroids of each cluster and use these to determine which digits they represent. Plot some of the wrongly assigned digits, and explain why they were misclassified.

PageRank

Problem 2



(a) **(b)**

Figure 1: Webs

(2a) Compute the adjacency matrix for each web in Figure 1.

(2b) Compute the transition matrix and the Google matrix for each web in Figure 1.

(2c) Use the power method to rank the two webs in Figure 1. In both cases, run the ranking algorithm using both the raw transition matrix $\mathbf{H}$, and the Google matrix $\mathbf{G}$ with $\alpha = 0.85$. Try different initial values for $\boldsymbol{\pi}$ in the power method. What do you observe?

(2d) Let us consider the scenario where the owners of Page 3 in the network depicted in Figure 1a are displeased due to Page 3's lower importance score, as determined by the Google matrix, in comparison to Page 1's score. To enhance the importance score of Page 3, they establish Page 5, which links to Page 3, and reciprocally, Page 3 links to Page 5. The question arises: Does this strategy elevate the importance score of Page 3 above that of Page 1?