

TDT4173 - Task 1

Håvard Hjelmeseth

September 2021

1 K-means algorithm

For my first algorithm i have decided to implement the K-means algorithm , which is an unsupervised learning algorithm for data classification. The algorithm could be improved by running it multiple times with randomized initial centroids for the clusters, and choosing the ones which give the best results.

1.1 First dataset

The first dataset contains points in a 2D space, where the points seem to have a natural separation into two clusters. The implementation of the algorithm seems to have no issues with this dataset.

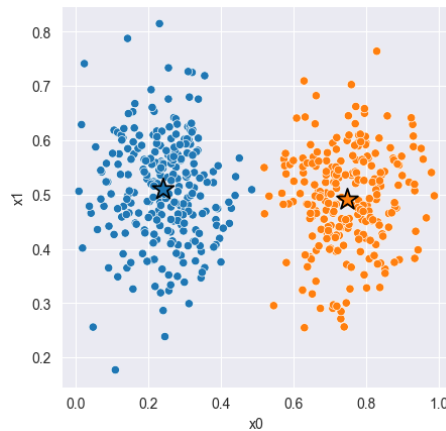


Figure 1: First dataset classification with k-means

1.2 Second dataset

The second dataset is a bit harder, where the points are still in a 2D space, but does not longer have a natural separation into two clusters. Running the

kmeans algorithm with $k=10$ clusters seems to give a good result as seen in the plots.

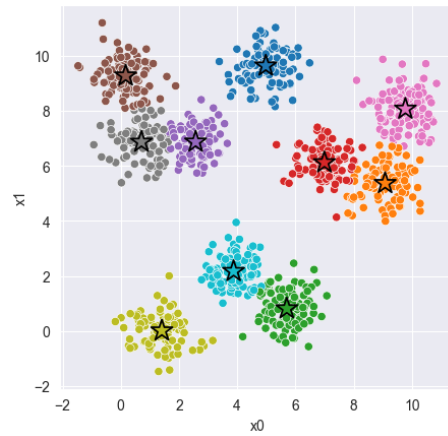


Figure 2: Second dataset classification with k-means

2 Second algorithm