

# COMP 462 ASSIGNMENT3

## k-Means Clustering

Mert Dogan

MEF University  
Department of Computer Engineering  
doganme@mef.edu.tr  
041701041

**Abstract.** In this assignment, I design k-Means clustering algorithm for using unsupervised learning method from scratch. In the given a three datasets there are two features and their corresponding labels. In this implementation we do not need class labels for the learning method, because it is an unsupervised learning problem. I do not use an existing classifier from any other software library.

### 1 Dataset

In the given dataset, there are three different .txt files. The dataset files contain features (in 2D) and class labels. In this assignment, I discard the class labels because k-Means is an unsupervised algorithm and does not need class labels.

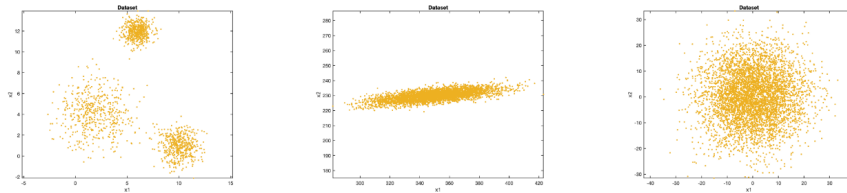


Fig. 1: Three (unsupervised) datasets.

### 2 Data Plotting and Results

In this section, the given figure shows a scatter plot which represents the final results of centroids. I represent the final plot in one drawing. \* symbol corresponding initial random points and X corresponding final cluster centres. The graphs plotted with using matplotlib library.

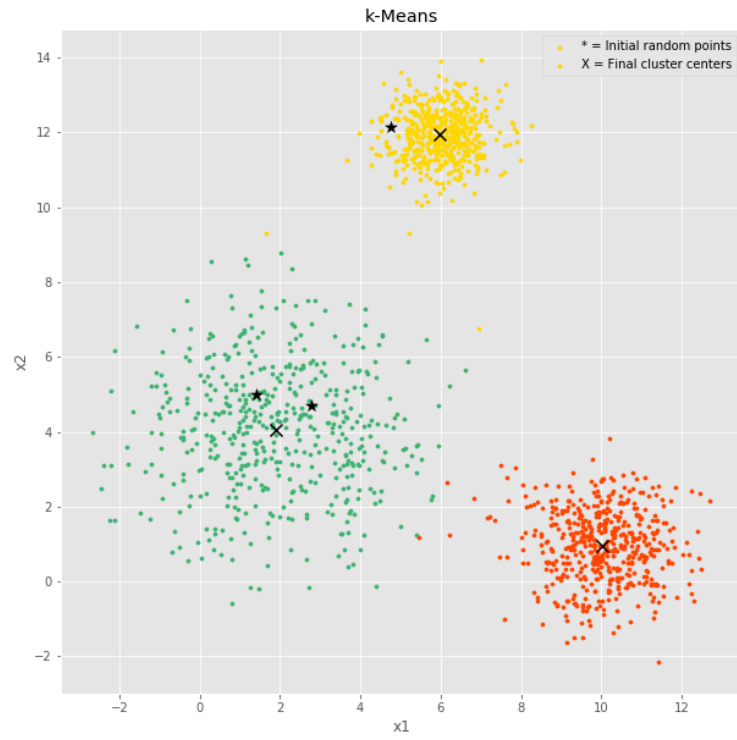


Fig. 2: Dataset = 1,  $k = 3$  - Final Cluster Centers, Iteration Count = 5.

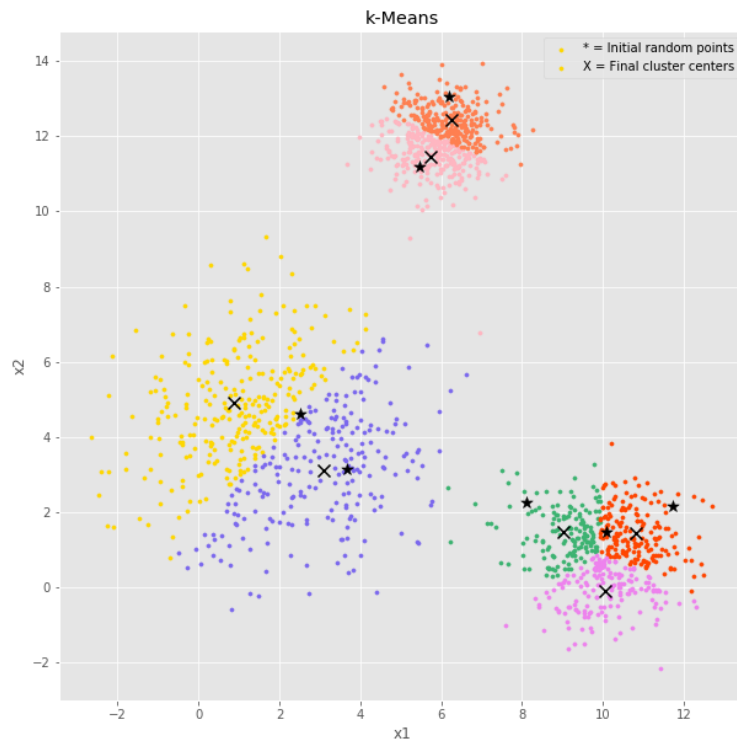


Fig. 3: Dataset = 1,  $k = 7$  - Final Cluster Centers, Iteration Count = 14.

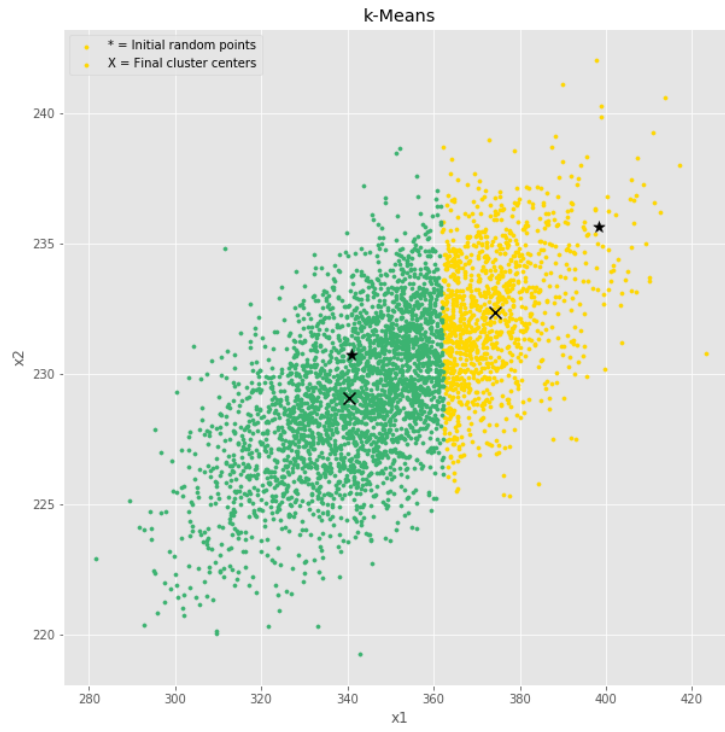


Fig. 4: Dataset = 2,  $k = 2$  - Final Cluster Centers, Iteration Count = 7.

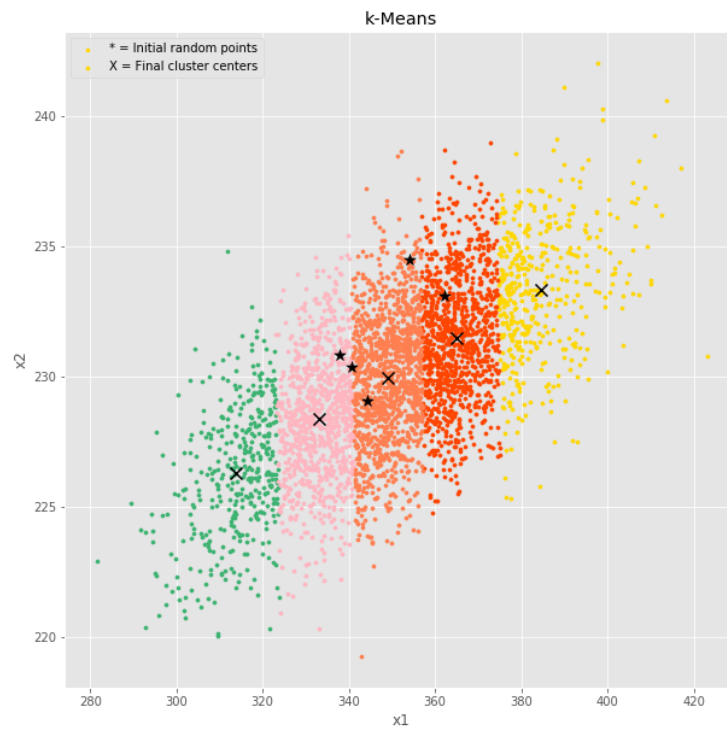


Fig. 5: Dataset = 2,  $k = 5$  - Final Cluster Centers, Iteration Count = 34.

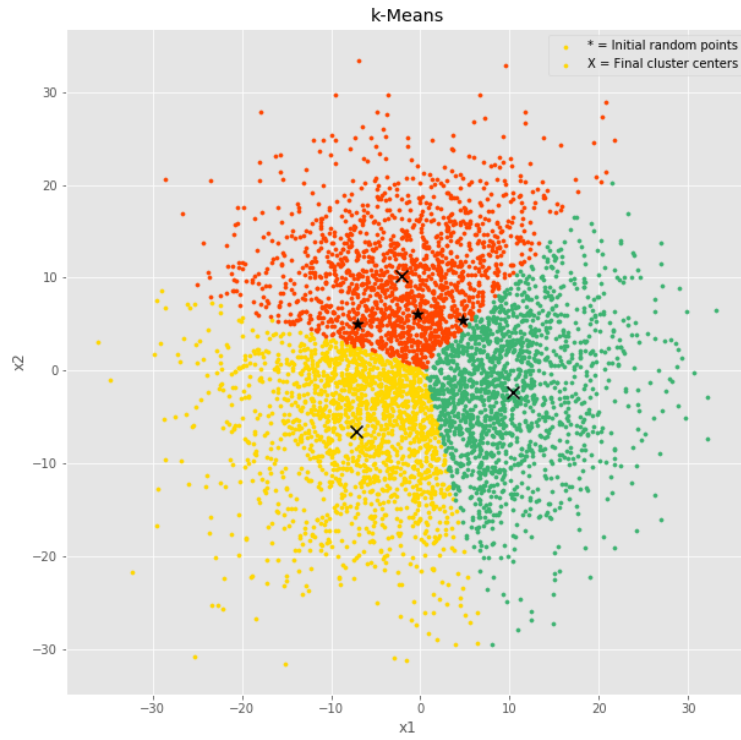


Fig. 6: Dataset = 3,  $k = 3$  - Final Cluster Centers, Iteration Count = 32.

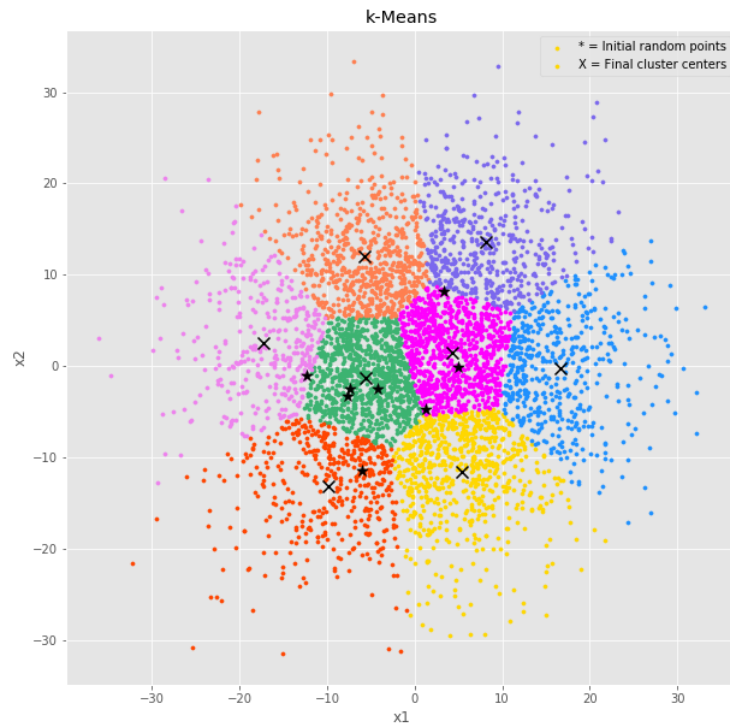


Fig. 7: Dataset = 3,  $k = 8$  - Final Cluster Centers, Iteration Count = 54.