



Pattern Recognition Methods
and
Introduction to Machine Learning

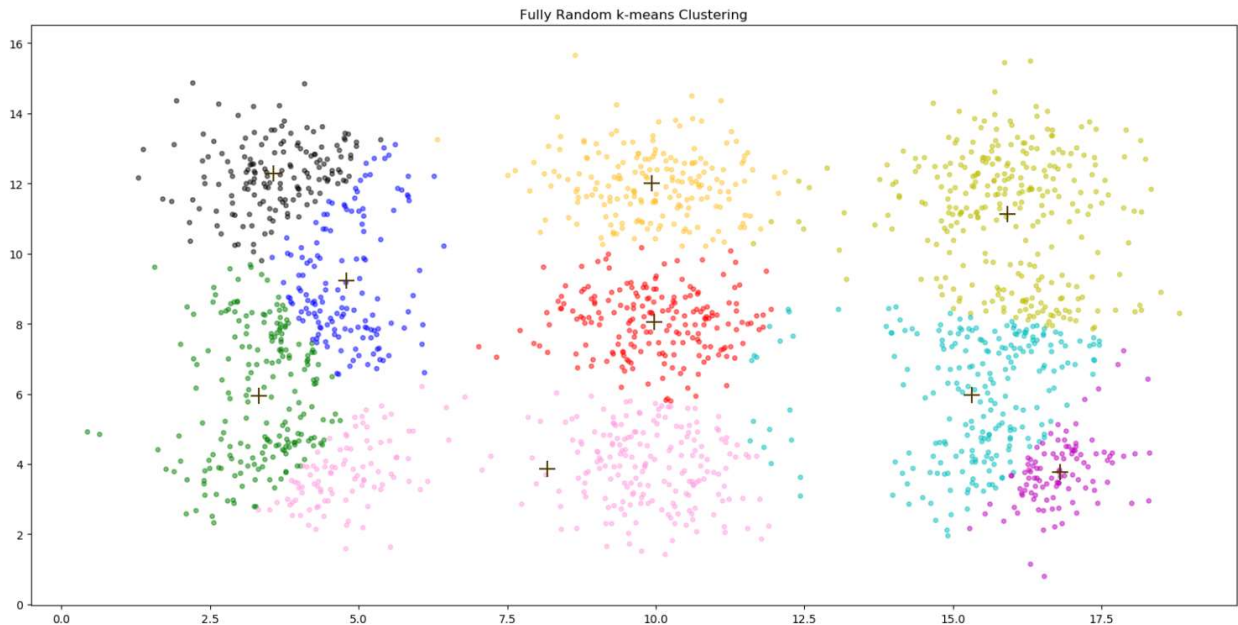
Homework 5 - Report
Benchmarking The Initialisation Methods

Prepared by: Mehmet Kapson

7.05.2019

1. k-means Clustering with Fully Random Initialisation

Dataset are created similar to the given dataset. Cluster centers are chosen randomly and updated iteratively. The result of this k-means is shown below:



P.S: Black plus signs are the centers of each cluster for all initialisation settings in this report.

2. k-means Clustering with Random Partition Initialisation

Scikit-learn is used to determine the clusters. It is run for 5 times. The result of this k-means is shown below:

```
Anaconda Prompt
(base) C:\Users\Mehmet>cd Desktop
(base) C:\Users\Mehmet\Desktop>random_partition.py
The Davies-Bouldin score is : 0.6431120315177721
The Silhouette score is : 0.5273424559377415

(base) C:\Users\Mehmet\Desktop>random_partition.py
The Davies-Bouldin score is : 0.6199829505429216
The Silhouette score is : 0.5361342915595941

(base) C:\Users\Mehmet\Desktop>random_partition.py
The Davies-Bouldin score is : 0.6173394480445622
The Silhouette score is : 0.5377324272779415

(base) C:\Users\Mehmet\Desktop>random_partition.py
The Davies-Bouldin score is : 0.6102921888061947
The Silhouette score is : 0.5434021069046533

(base) C:\Users\Mehmet\Desktop>random_partition.py
The Davies-Bouldin score is : 0.6072543770476985
The Silhouette score is : 0.5392468049797844

(base) C:\Users\Mehmet\Desktop>random_partition.py
The Standard Deviation of Davies-Bouldin scores is 0.014116675028343895
The Standard Deviation of Silhouette scores is 0.0059230138096279875
```

Low Davies-Bouldin score shows good clustering. Besides, high Silhouette score shows good clustering. In these 5 runs, the results are okay.



3. k-means++ Clustering

Scikit-learn is used to determine the clusters. It is run for 5 times. The result of this k-means++ is shown below:

```
Anaconda Prompt

(base) C:\Users\Mehmet>cd Desktop

(base) C:\Users\Mehmet\Desktop>k-means++.py
The Davies-Bouldin score is : 0.6112876306106566
The Silhouette score is : 0.5460715188674168

(base) C:\Users\Mehmet\Desktop>k-means++.py
The Davies-Bouldin score is : 0.6144865083780929
The Silhouette score is : 0.5423765220527778

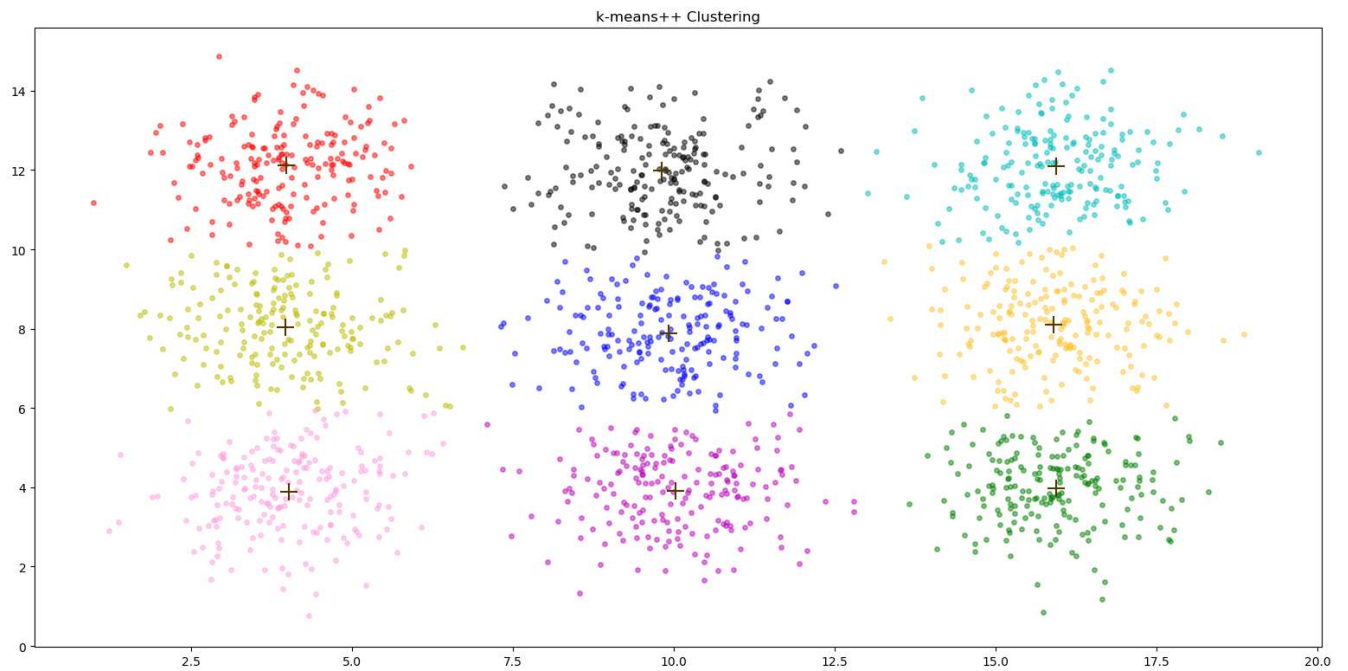
(base) C:\Users\Mehmet\Desktop>k-means++.py
The Davies-Bouldin score is : 0.6210693206648323
The Silhouette score is : 0.5362437321272338

(base) C:\Users\Mehmet\Desktop>k-means++.py
The Davies-Bouldin score is : 0.6095993011151435
The Silhouette score is : 0.544117256180449

(base) C:\Users\Mehmet\Desktop>k-means++.py
The Davies-Bouldin score is : 0.6127567806946416
The Silhouette score is : 0.5440847926039414

(base) C:\Users\Mehmet\Desktop>k-means++.py
The Standard Deviation of Davies-Bouldin scores is 0.004425848249518014
The Standard Deviation of Silhouette scores is 0.0037751803779169247
```

Again, results are close to the previous ones. The same situation is valid in these cases.



If the initial centers are chosen correctly, k-means clustering algorithm works good and it gives decent clusters. This conclusion can clearly be seen by looking the results.