Project2-K-means Report

Nourhan ElNaggar

1221732434

1. Introduction

The main objective of this project is to implement the k-means algorithm on a dataset of 2-D points. We are given initial cluster centers and we need to apply two strategies. The first one is randomly picking the initial centers from the given samples and the second one is more optimized by choosing the first samples randomly and then iterating over all other sample to choose another centroid where the distance between the new centroid and all previous chosen centroids is maximized. Cleary based on the result below, second strategy has showed better results than the first randomized clusters strategy.

2. Strategy 1 Results

2.1 K=3 centroids

```
First centroids: [[1.72614408, 6.81819407],
                  [2.70699582, 1.64002569],
                  [6.11106851, 6.23497555]]
```

Final centroids:

x	y
2.56146449	6.08861338
5.47740039	2.25498103
6.49724962	7.52297293

The loss: 1294.2094505393381

2.2 K=5 centroids

```
First centroids: [[5.32508246 7.68399917]
                  [2.87448907 2.657599  ]
                  [5.48121965 6.55171777]
                  [2.16641743 2.99414637]
                  [7.59763505 7.93924   ]]
```

Final centroids:

x	y
3.22202355	7.15937996
7.55616782	2.23516796
5.37514379	4.53101654
2.68198633	2.09461587
7.49365367	8.52417952

The loss: 592.0694342732747

3. Strategy 2 Results

3.1 First centroids

K= 4

x	y
3.33995748	2.59215224
6.60345839	7.57042104
7.38076264	2.33245532
2.85859235	6.93136525

The loss: 788.2693490065564

3.2 Second centroids

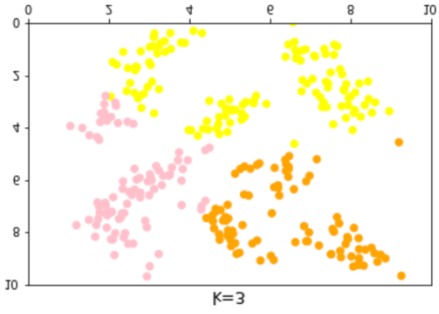
K = 6

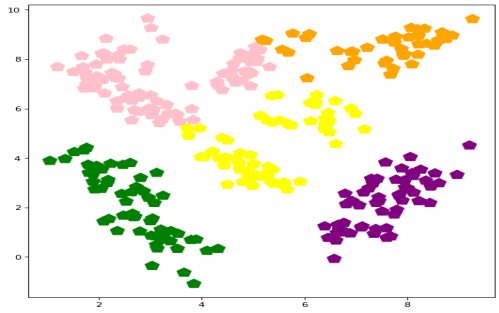
x	y
7.41419243	2.32169114
2.52382885	7.02897469
3.14506148	0.90770655
7.75648325	8.55668928
3.502455	3.62870476
5.46427736	6.83771354

The loss: 592.0694342732747

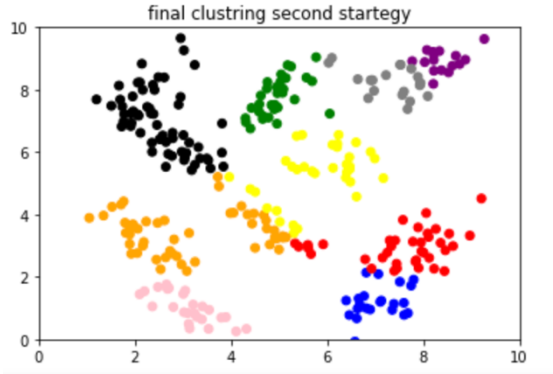
4. Analysis

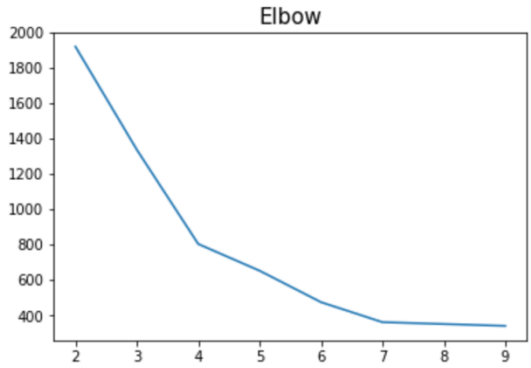
- Below are graphs for strategy 1 for k = 3 and k = 5:





- **Second strategy final clustering and elbow curve:**





-