

# Lloyd's algorithm

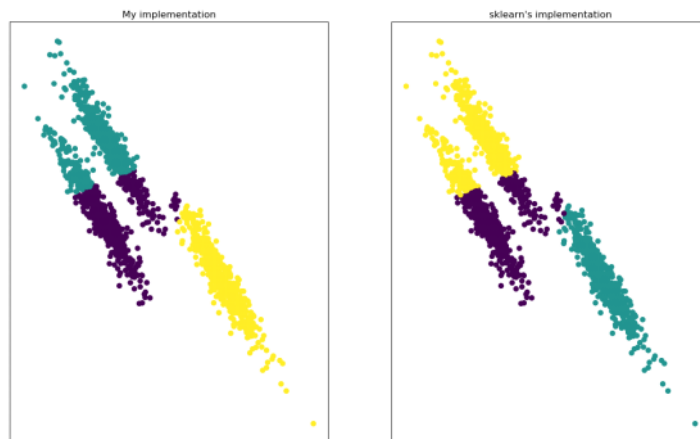
1.1

Implementation from

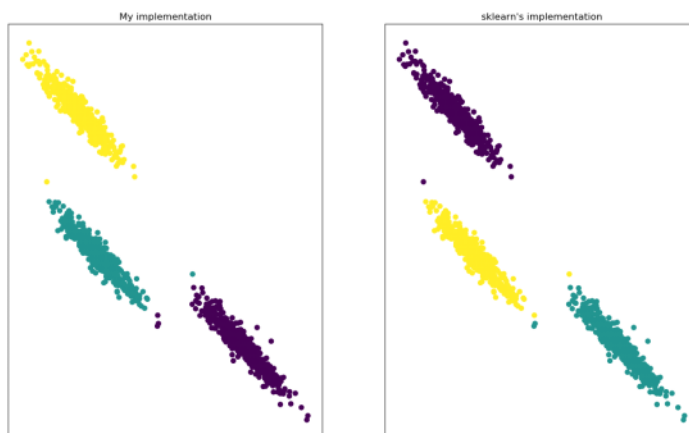
<https://codereview.stackexchange.com/questions/80050/k-means-clustering-algorithm-in-python>

1.2

random\_state = 170



random\_state = 100



Both implementation are quite similar. Only noticeable difference is that some points in the middle are classified differently.

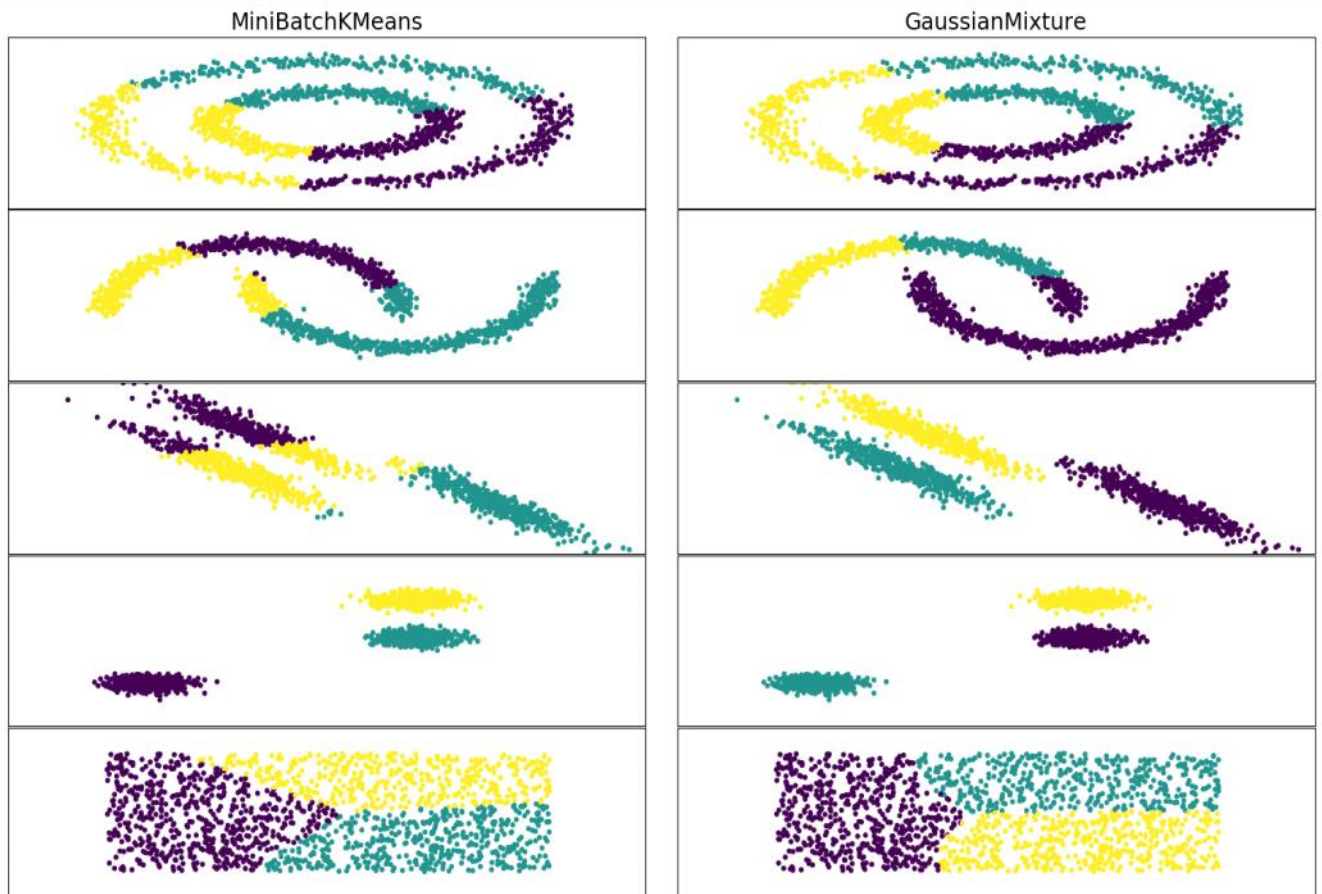
1.3

Yes, it will work with random initialization of centroids.

Worst case is if all identical points. K-means itself cannot adjust k if such things happens, but usually not a problem. One centroid will usually change after the first iteration, and move into different directions.

# Mixture of Gaussians

## 2.1



GaussianMixture performs better than KMeans in plot two and three. GaussianMixture works better with non-linear geometric distributions than KMeans.

# Hierarchical Clustering

## 3.1

Changes in the code:

```
data = ["ATGTAAA", "ATGAAAA", "ACGTGAA", "ACGAGGG", "ACGAGGA", "ACGAGTC",  
"ACGAGCC"]
```

```
labels = ["shark", "ray-finned fish", "amphibians", "primates", "rodents",  
"crocodiles", "dinosaurs"]
```

```
z = hier.linkage(mat)
```

```
dn = hier.dendrogram(z, labels = labels)
```

Result:

