**Assignment 4**

Machine Learning, SS2021

|  |  |  |
| --- | --- | --- |
| **Team members** | | |
| Last name | First name | Matriculation Number |
| Rodiga | Luca | 1620972 |
| Binford | Bob | 1600002 |

1.1)

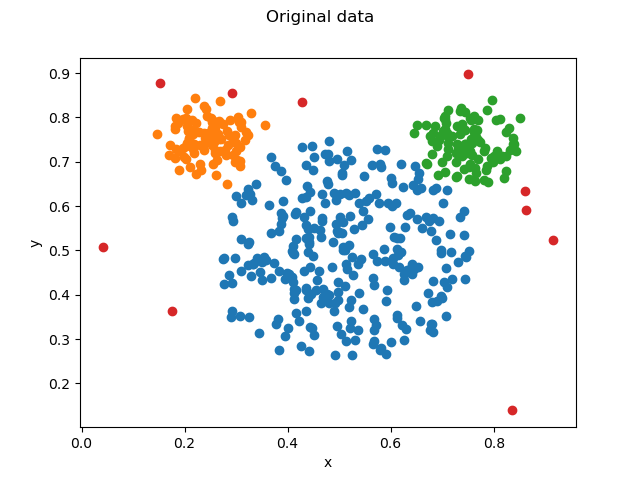
￼
￼￼￼
￼￼￼
￼￼￼￼￼￼￼￼￼
￼
Ink Drawings


￼
￼￼￼
￼￼￼
￼￼￼￼￼￼￼￼￼
￼
Ink Drawings

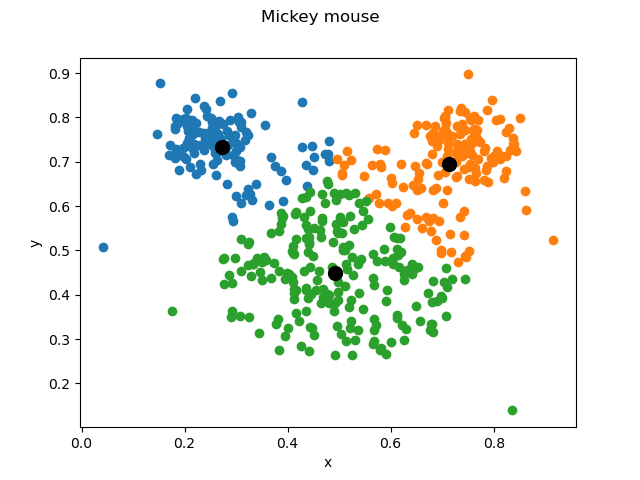

￼
￼￼￼￼￼
￼￼￼
￼￼￼￼
￼￼￼￼￼
￼￼
￼￼￼
￼
￼

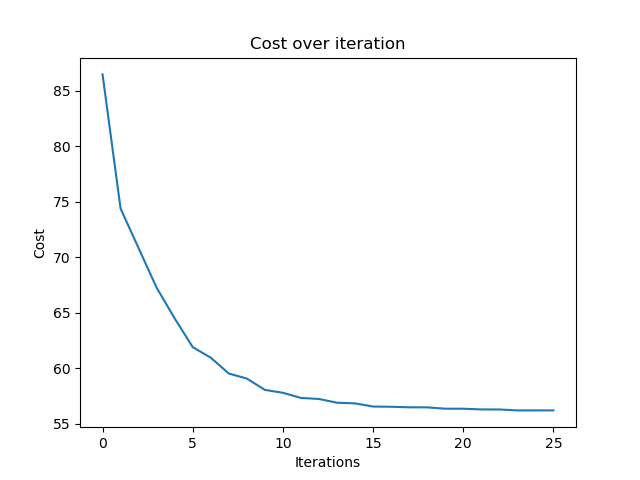

￼￼￼￼￼￼￼
￼


2.1)



K = 3



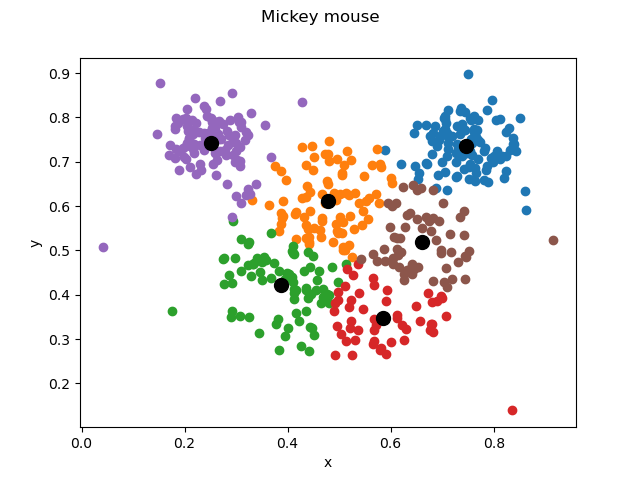


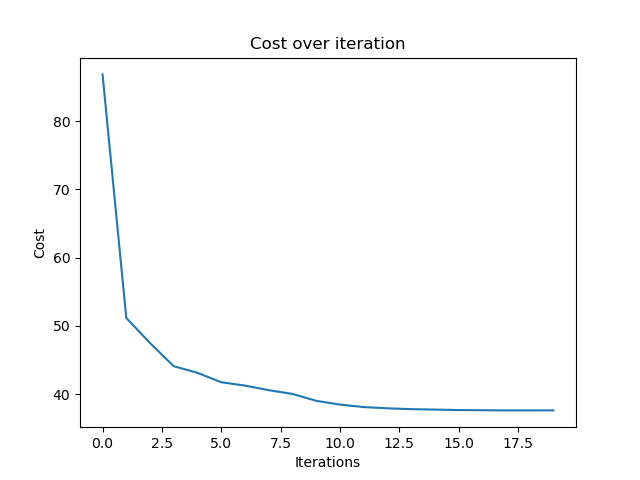
K-Means with K= 3 offered the most accurate clustering, compared to the original data plot, as we still have one cluster for each ear and one cluster for the head.

The algorithm converged in iteration 12.

Compared to the original data, it can be seen, that it was not possible for K-Means and K = 3 to achieve the sharp edges between the head and the ears, because the outliers distorted the centroids.

K= 6

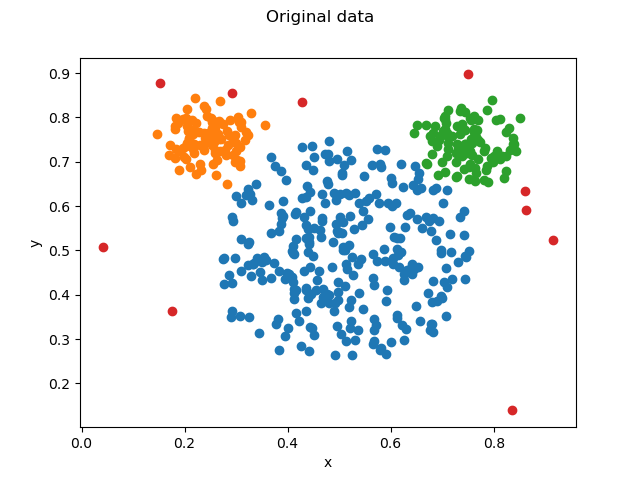




K-Means with K = 6 allows for a better separation of the head and the ears compared to K = 3, but we do have 4 clusters for actually only one head in the original data.

2.3)

2.3.2) Yes, as you can see the red dots in the original data are outliers.



K-Means is **not** robust against outliers, because centroids are always calculated with the mean of all assigned datapoints of the cluster and therefore, also always take the outliers into account, which can hinder performance, if there are too many outliers.