Programming #3

CS 545: Machine Learning

Kirtan Patel ID: 973582669

# Algorithm #1: K-Means clustering

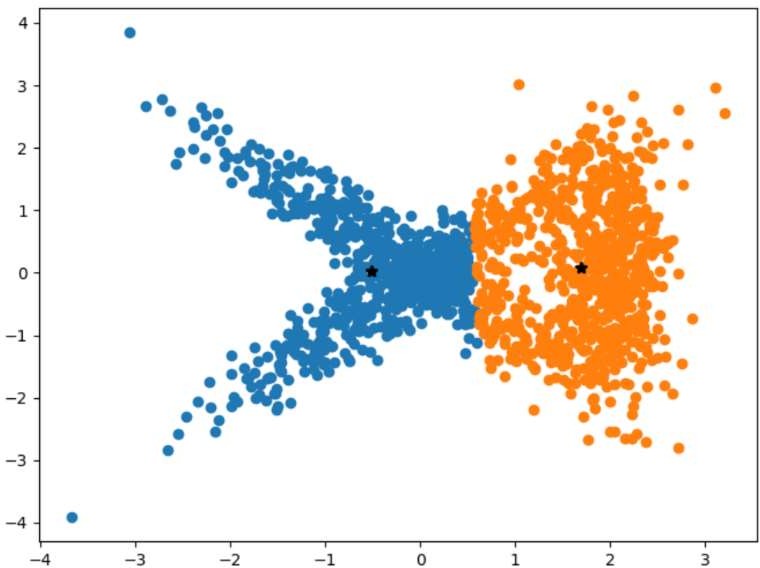
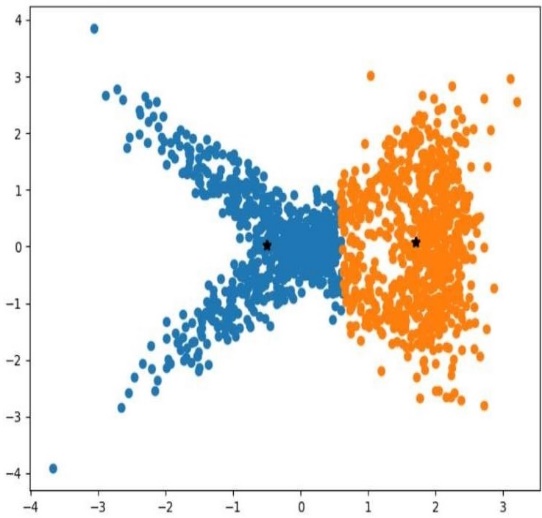
In this experiment, standard version of k-means algorithm is implemented. The initial centers(centroids) of ‘k’ clusters are randomly chosen from the dataset.

For a fixed ‘k’ value, This algorithm runs for ‘r’ number of times. For each of this ‘r’ we randomly chosen ‘k’ number of points (initial cluster centers). The sum-square-error is calculated for each ‘r’ and finally we select one model that gives minimum sum-square-error, for that ‘k’. (here r value is taken as 10)

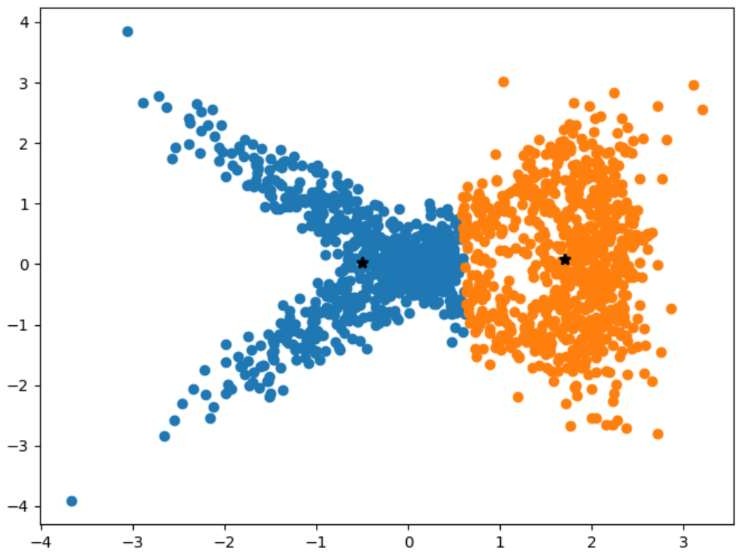
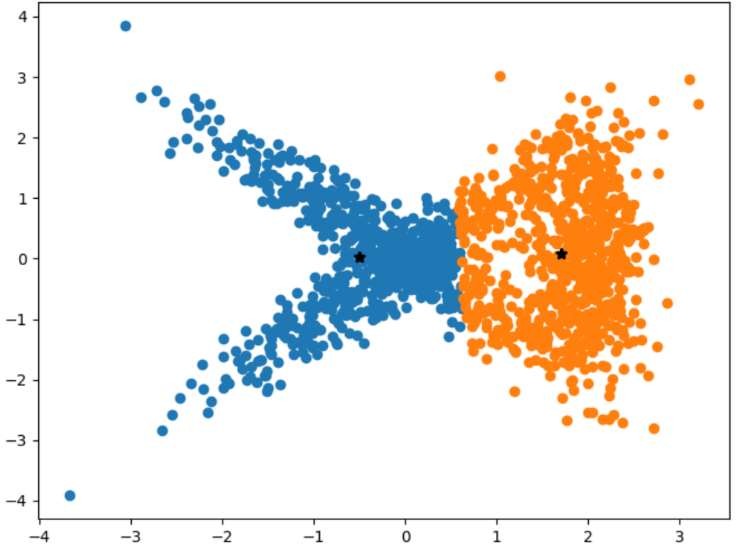
# Observations:

1. **Number of Clusters (k) = 2:**

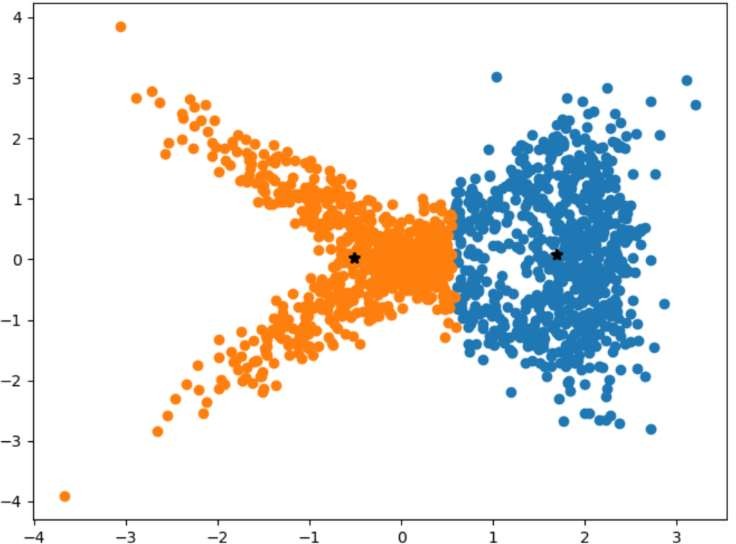
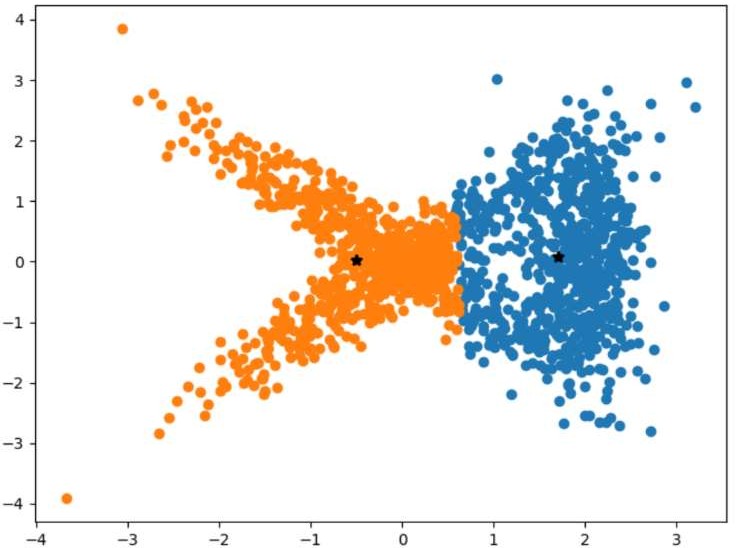
**r=1 r=2**



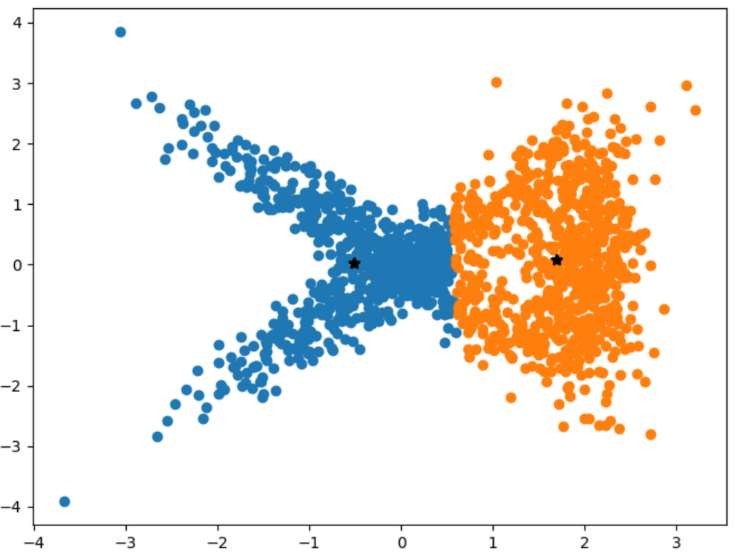
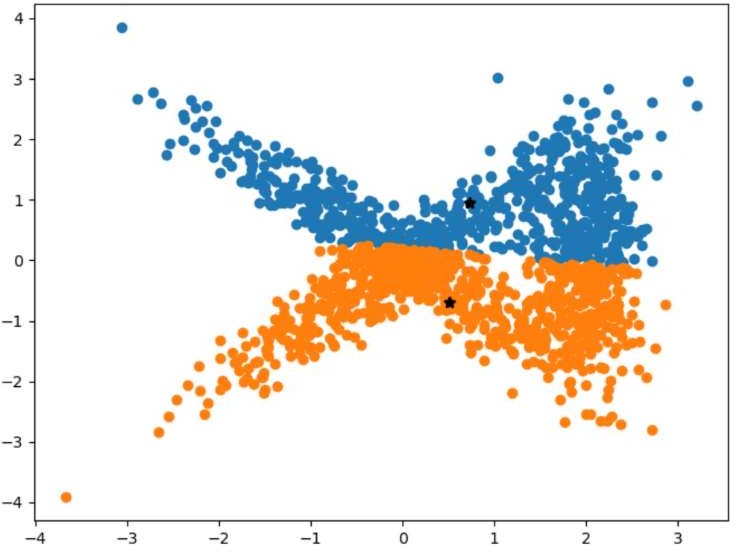
**r = 3 r = 4**



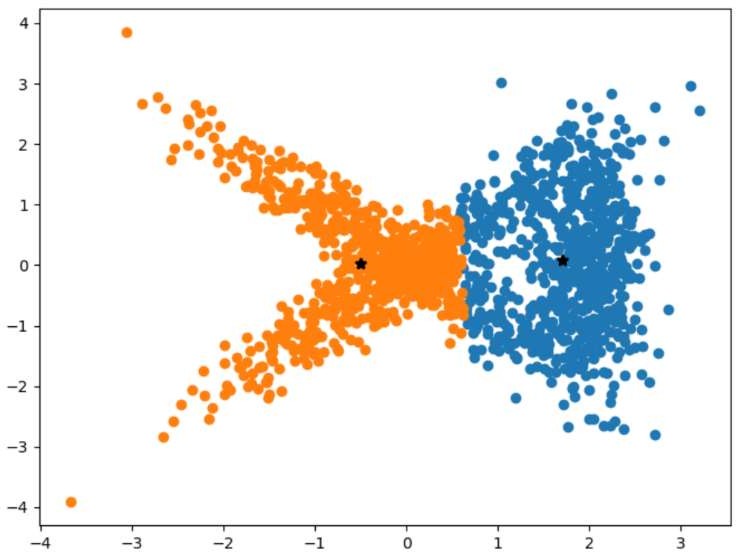
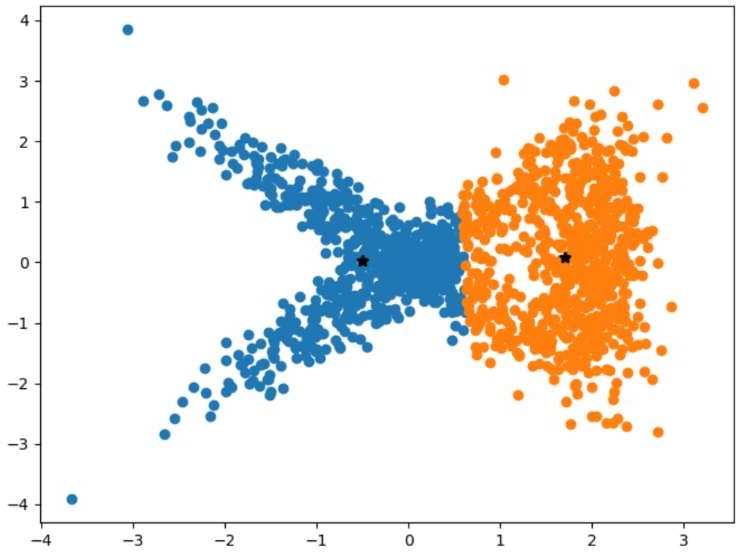
# r = 5 r = 6



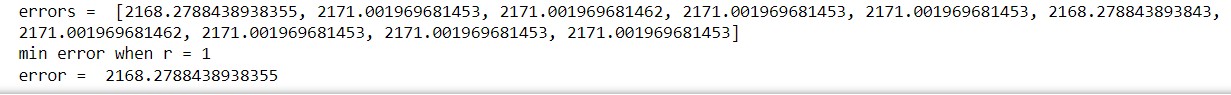
**r = 7 r = 8**



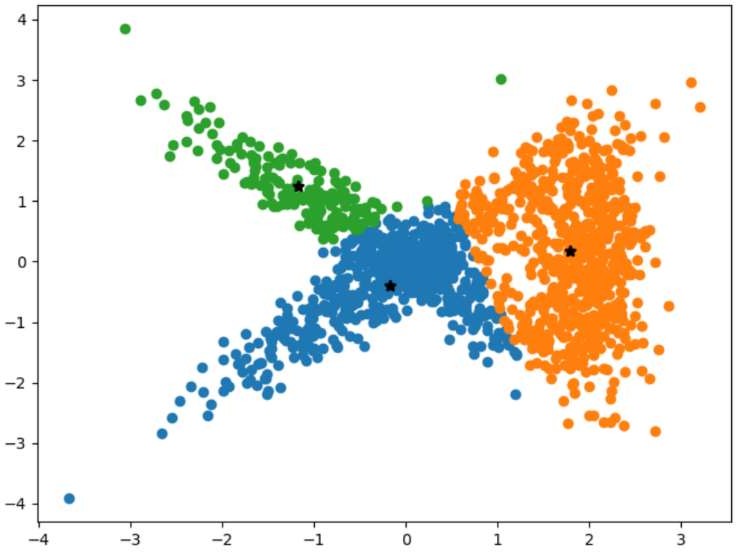
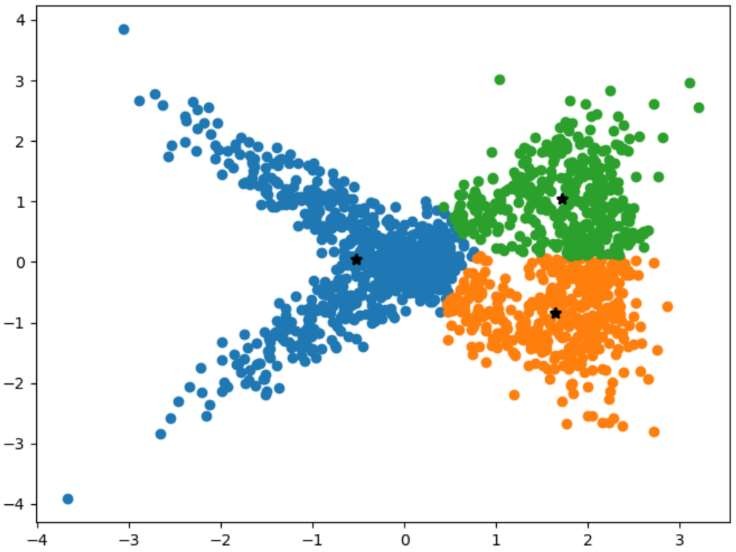
# r = 9 r = 10



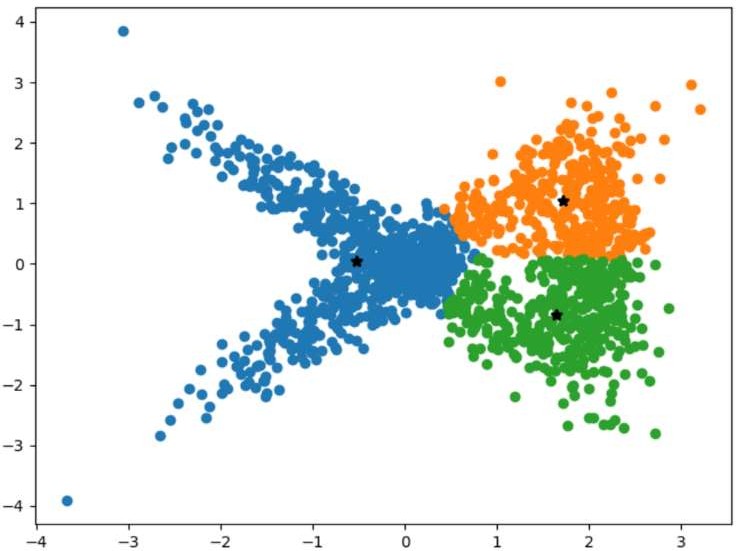
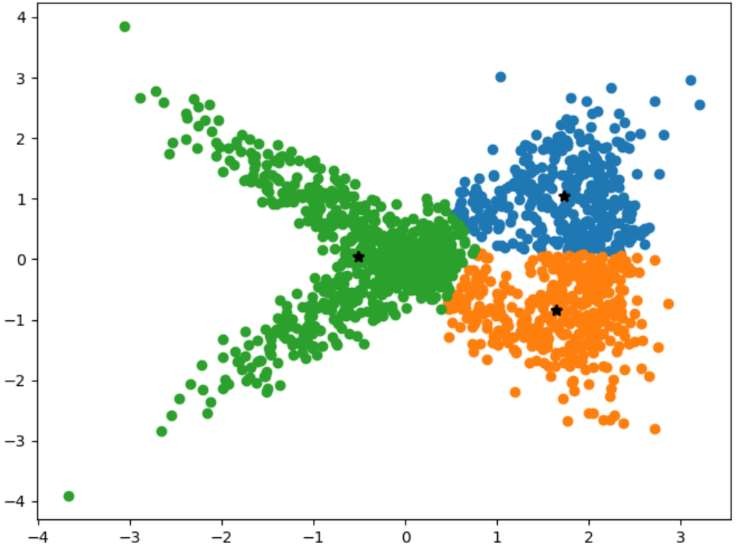
Errors for K=2



1. **Number of Clusters K = 3**

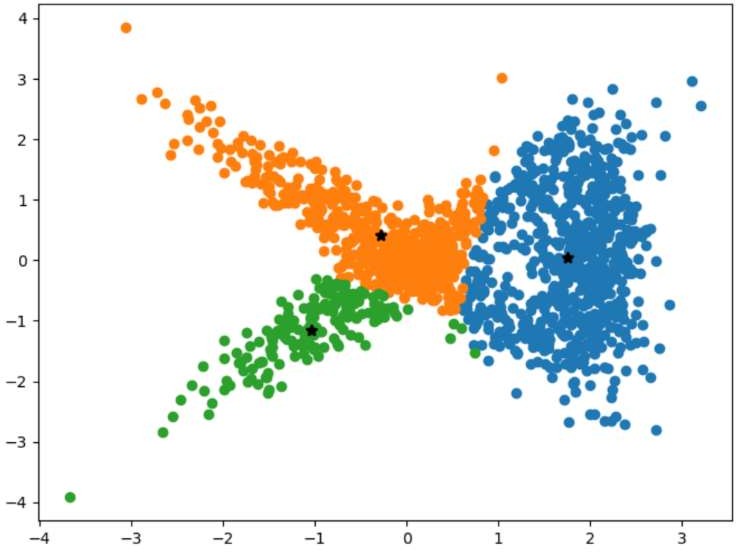
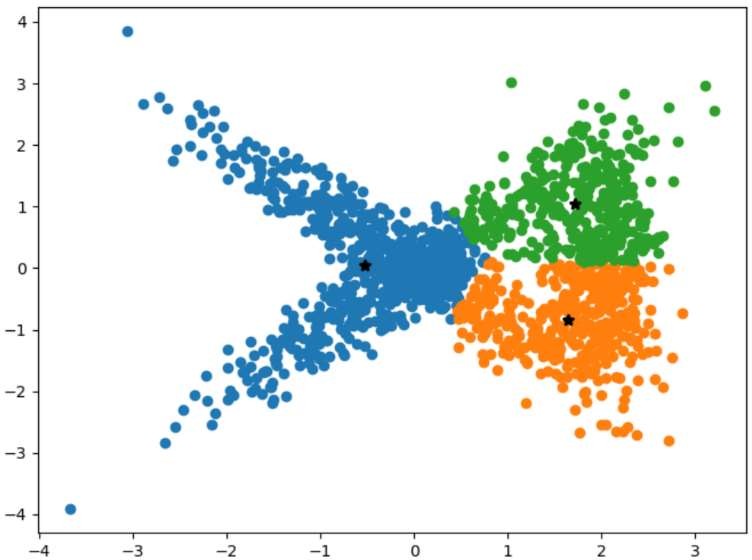
**r = 1 r = 2**

# r = 3 r = 4

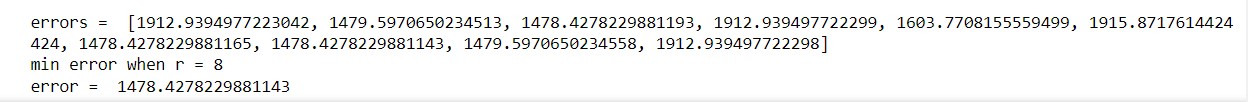
**r = 5 r = 6**

# r = 7 r =8

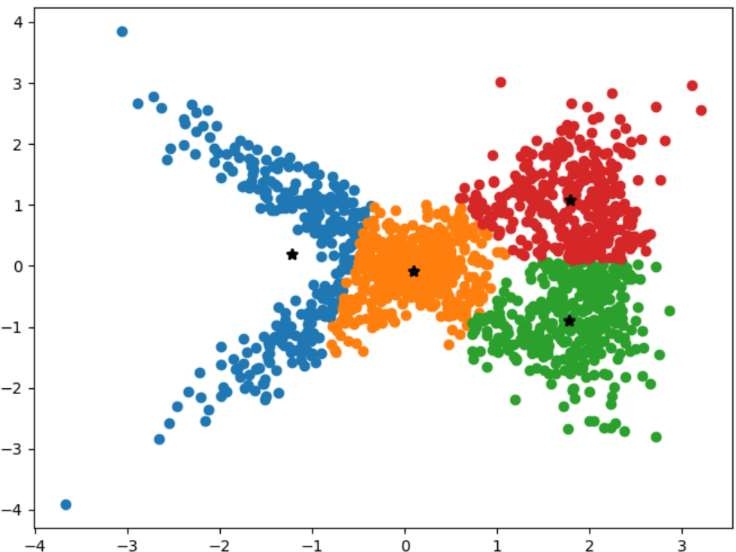
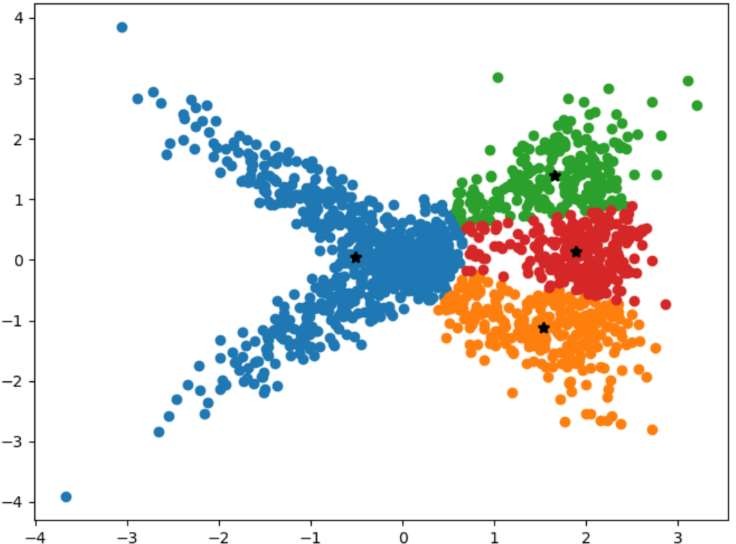
**r = 9 r = 10**

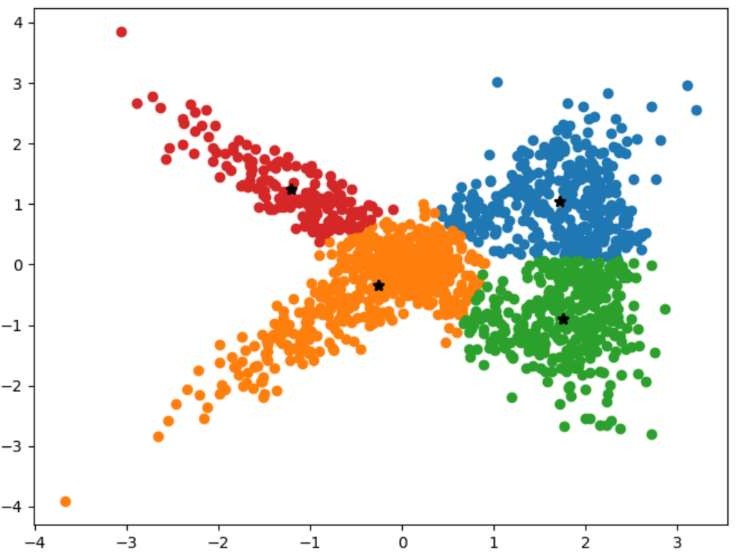
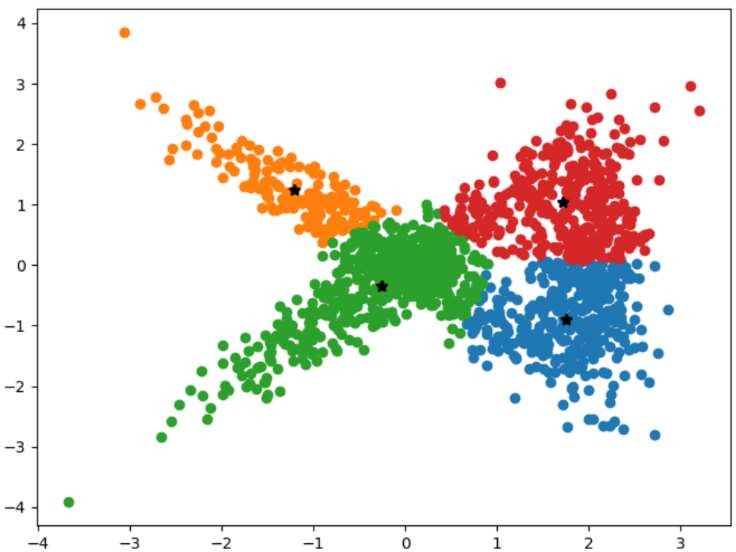
**Errors for K=3**



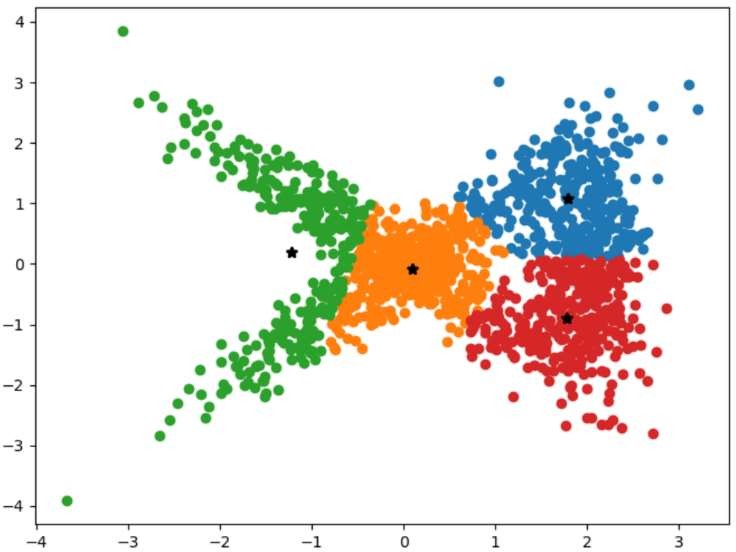
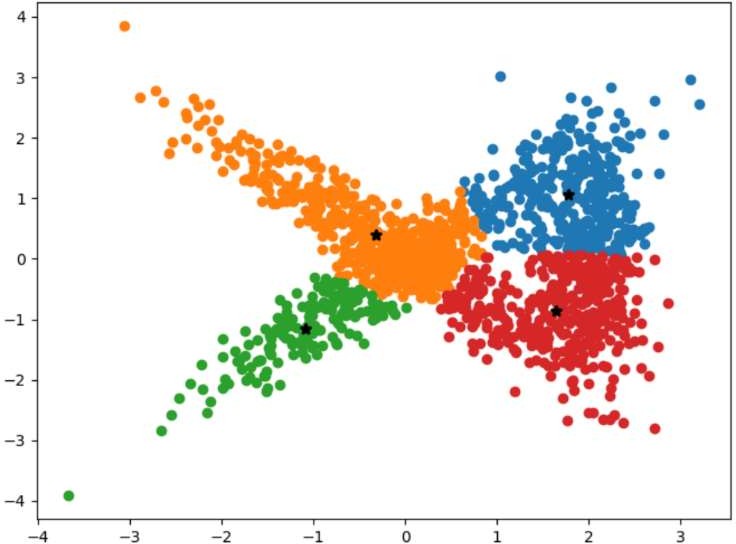
# Number of clusters K = 4

**r = 1 r = 2**

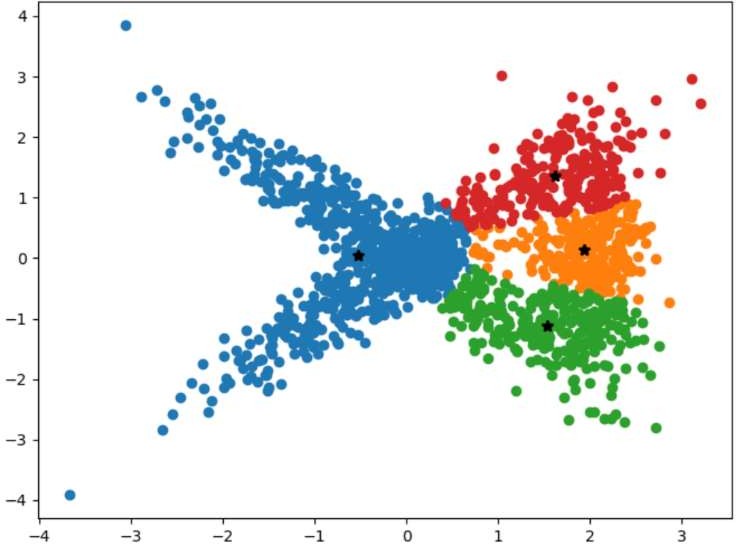
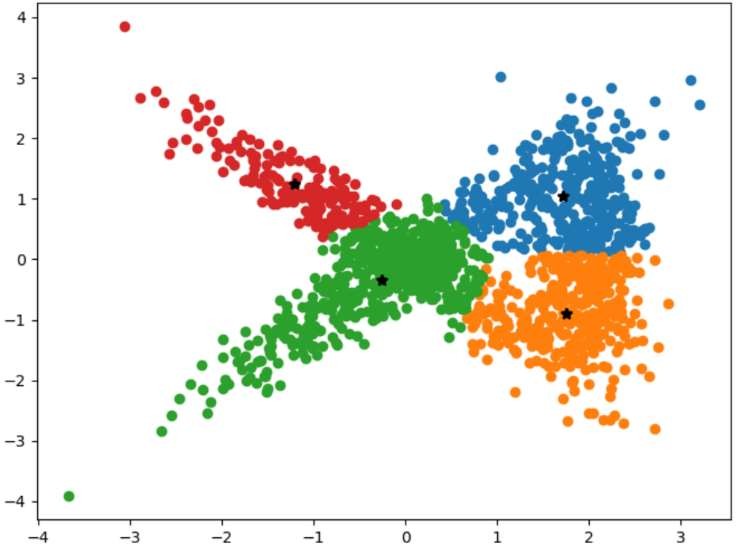
# r = 3 r =4

# r = 5 r = 6

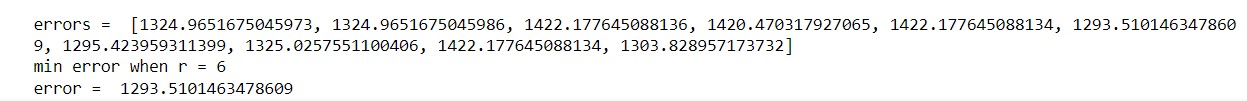


**r = 7 r = 8**

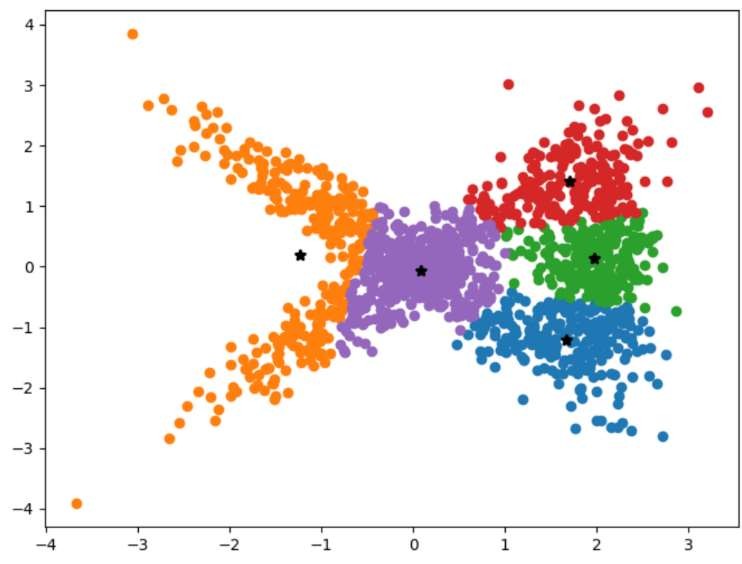
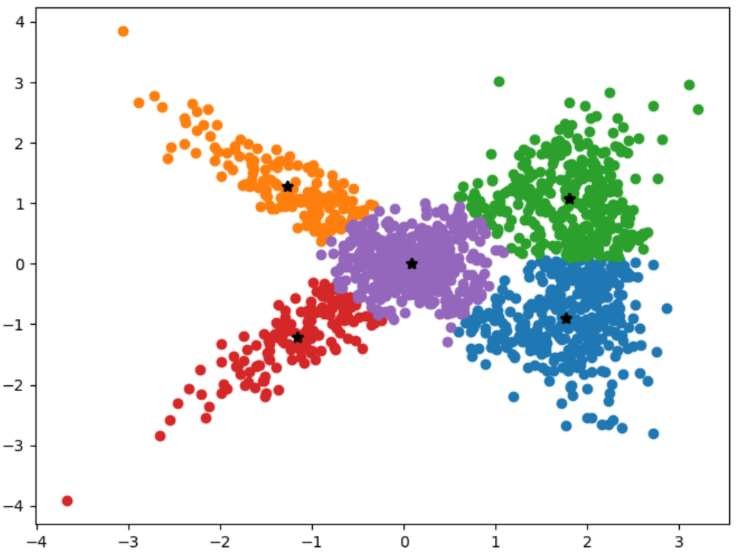
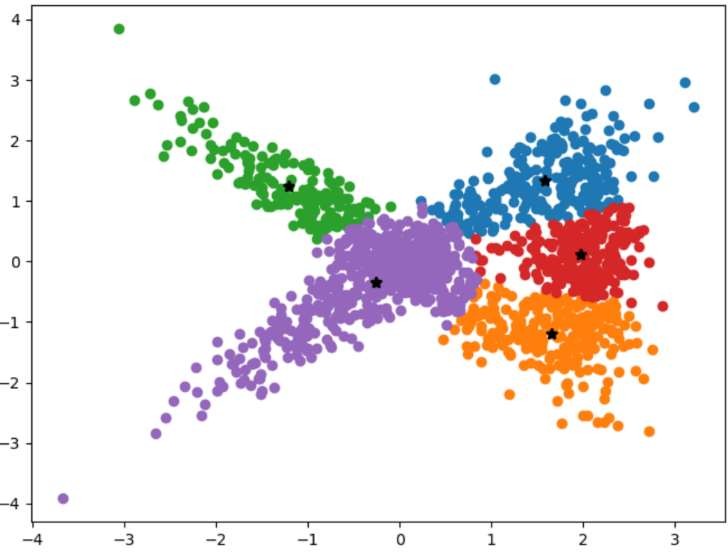


# r = 9 r = 10

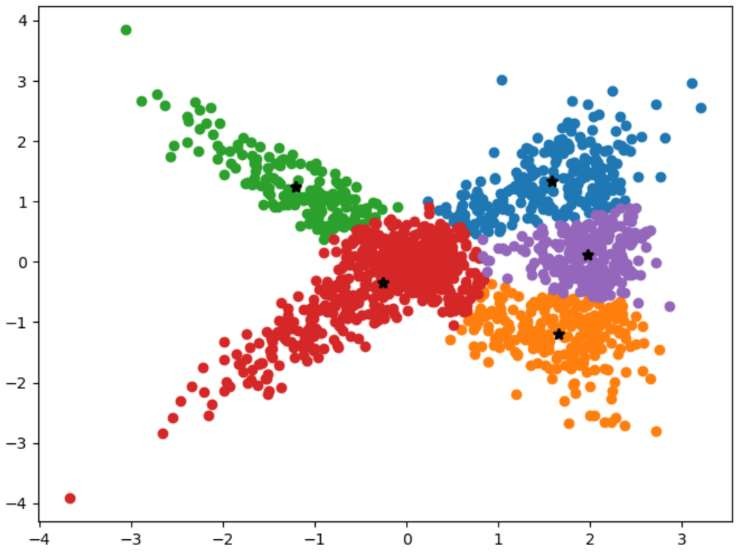
**Errors for K=4**



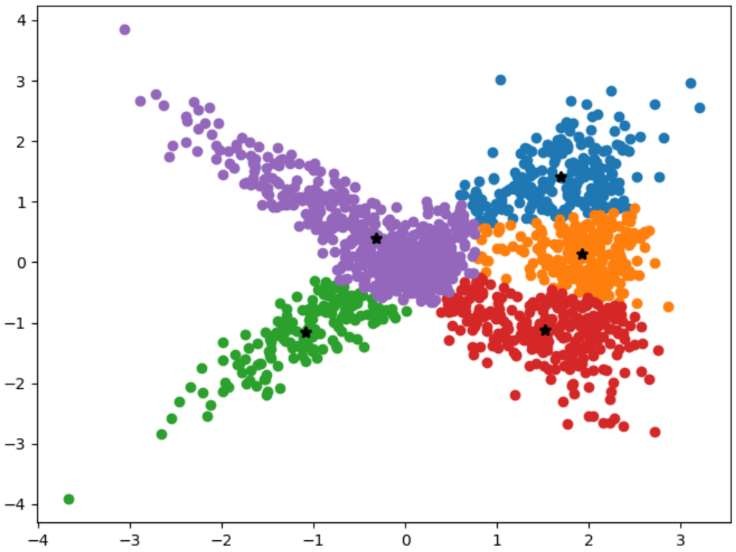
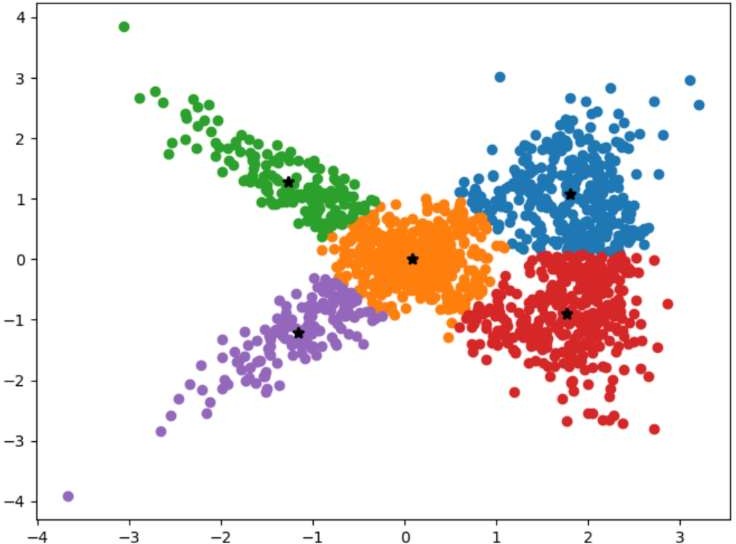
1. **Number of Clusters k = 5**

**r = 1 r = 2**

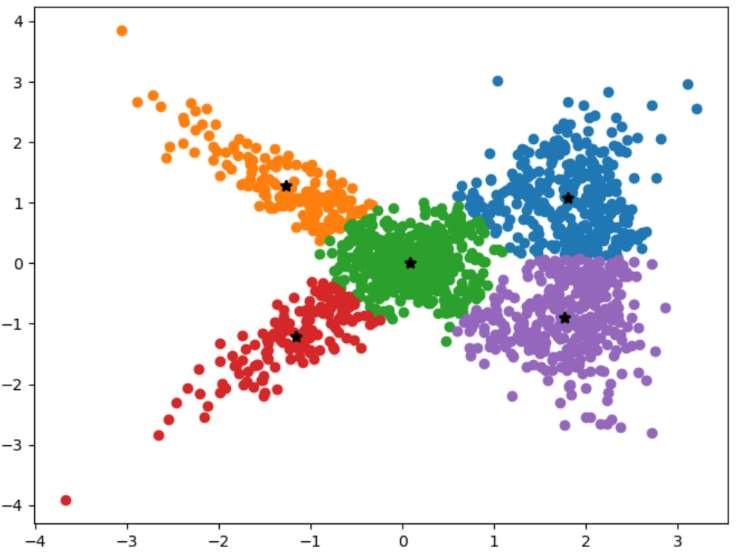
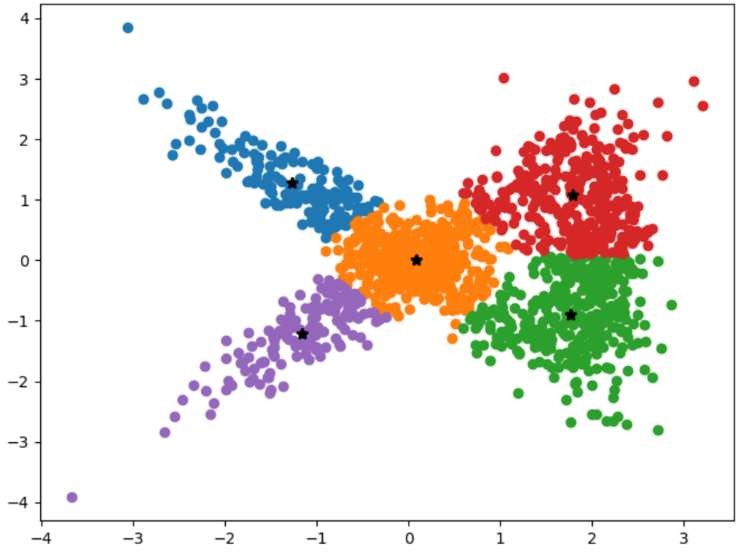
# r = 3 r = 4



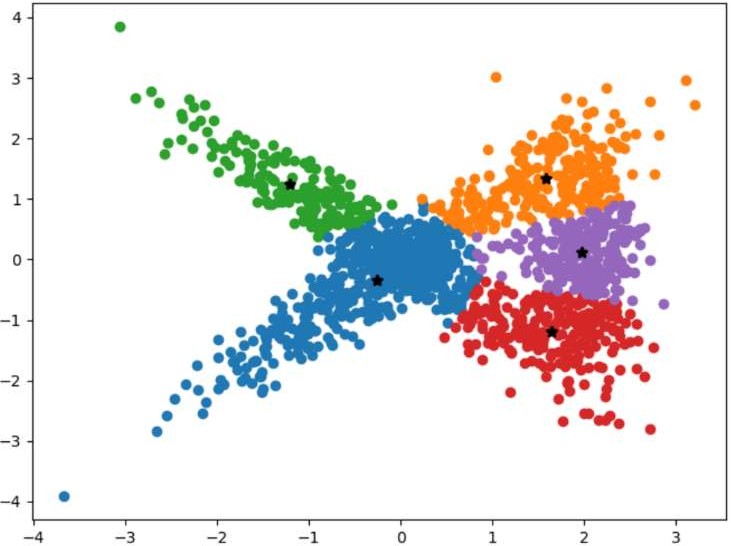
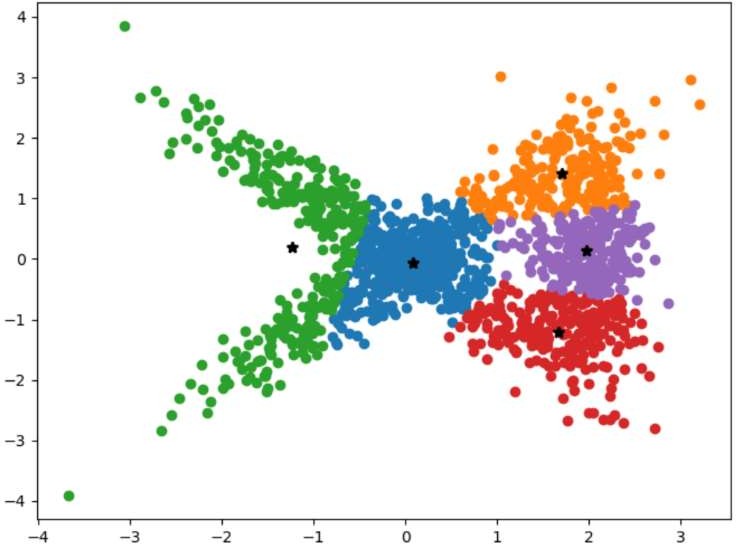
**r = 5 r = 6**



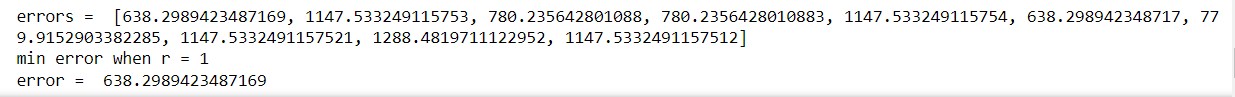
# r = 7 r = 8

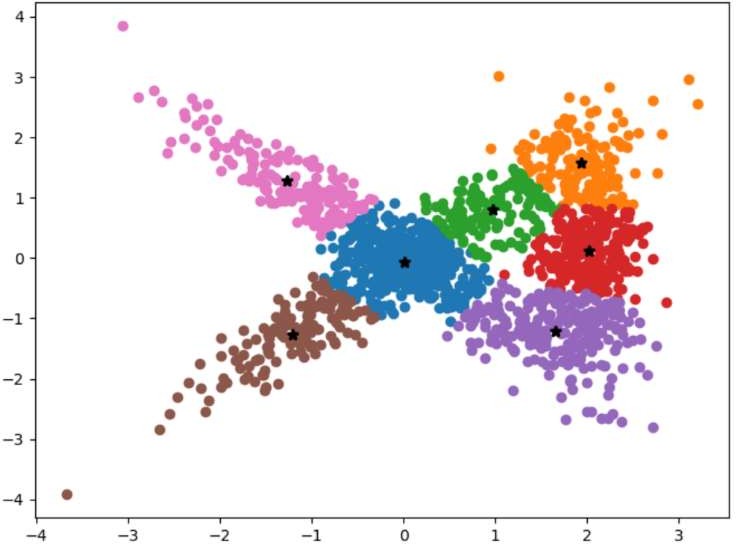
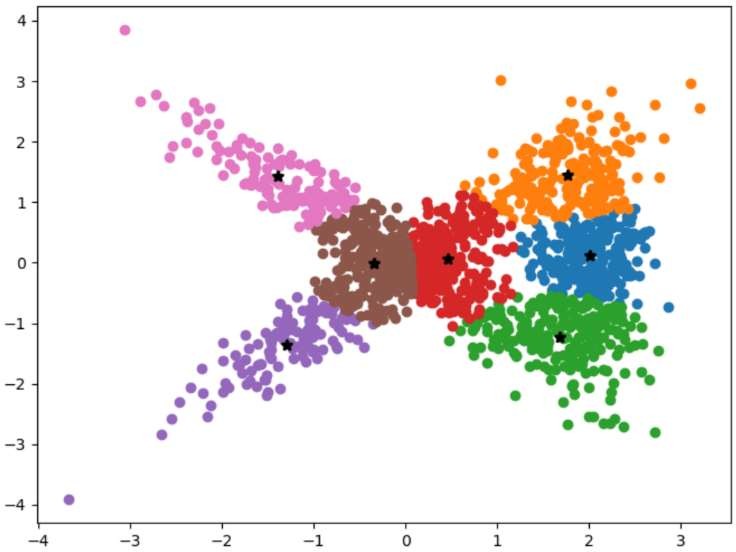
# r = 9 r = 10



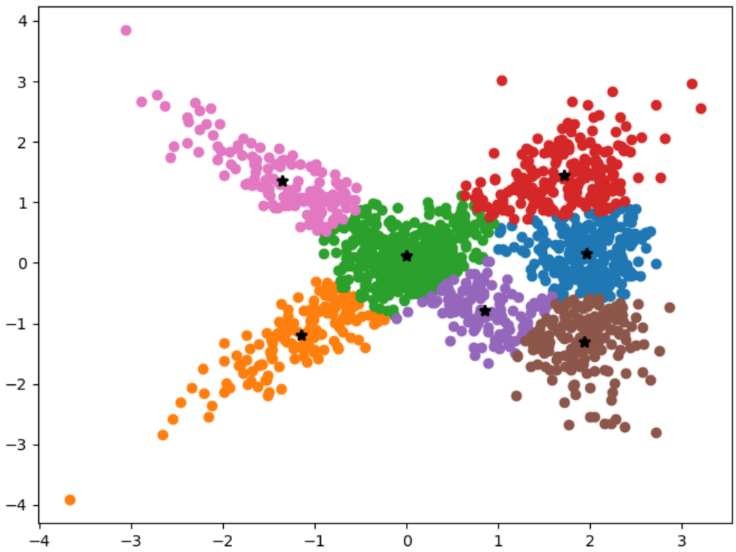
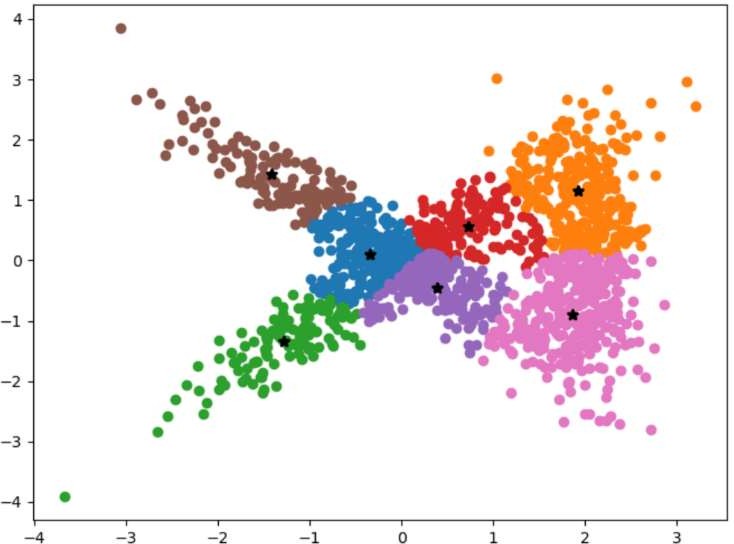
**Errors for K=5**



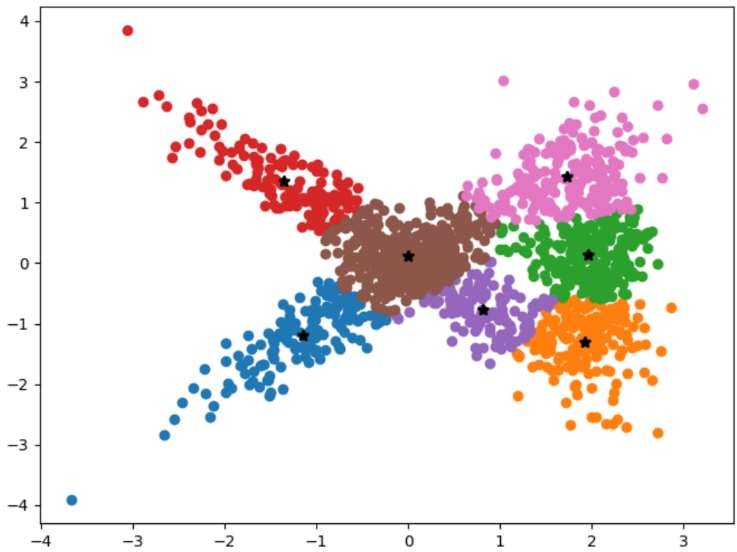
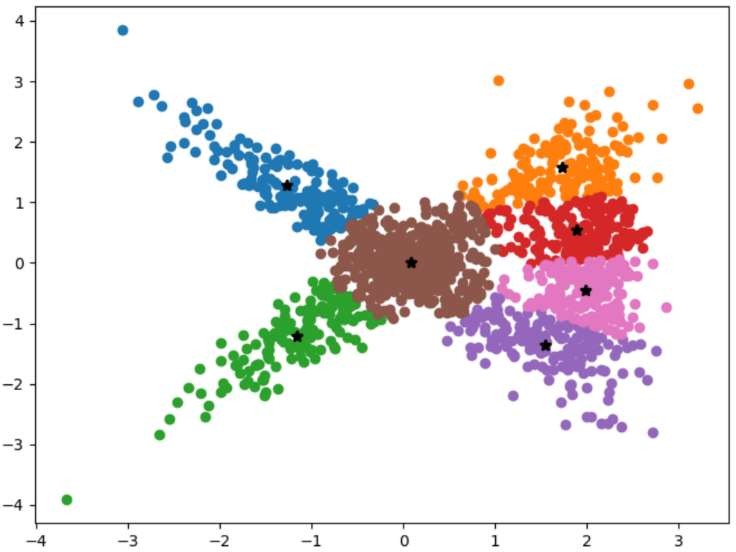
1. **Number of clusters k = 7**

**r = 1 r = 2**

# r = 3 r = 4

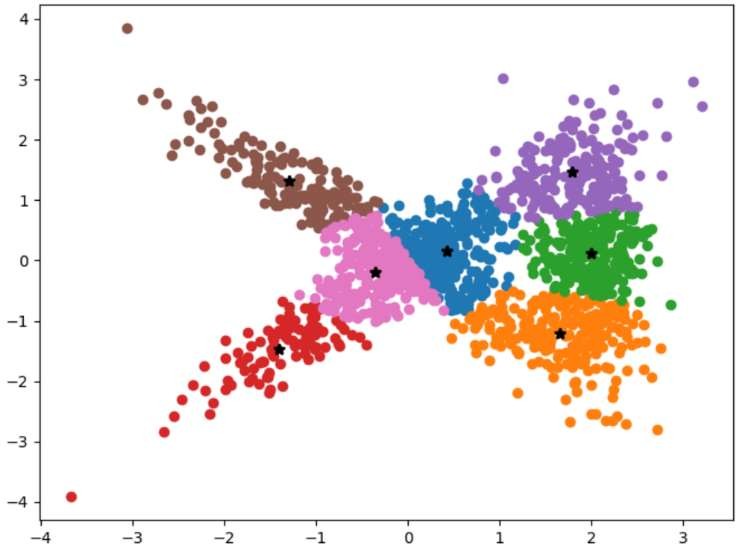
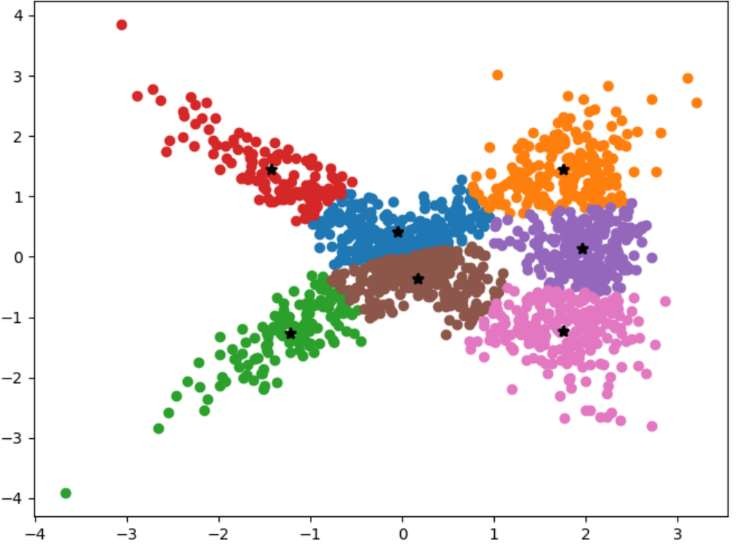
 

# r = 5 r = 6

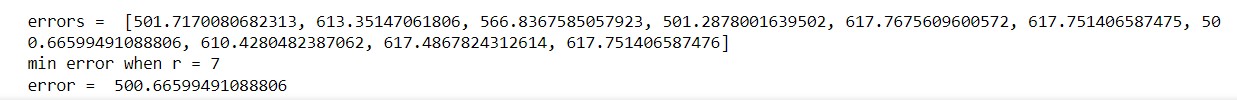
 

# r = 7 r = 8

**r = 9 r = 10**

**Errors for K=7**



For different ‘K’ values (k = 2 to 7). And for each ‘k’ value it is run for ‘r’ (10) times. Below table shows selected models sum-square-error for each ‘k’ value and ‘r’ value at which that model occurred.

|  |  |  |
| --- | --- | --- |
| **K Value** | **Minimum Sum-Square Error** | **Model occurred at ‘r’** |
| K = 2 | 2168.278 | 5 |
| K = 3 | 1478.427 | 1 |
| K = 4 | 810.717 | 5 |
| K = 5 | 638.298 | 6 |
| K = 6 | 627.896 | 4 |
| K = 7 | 500.78 | 7 |