# 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)
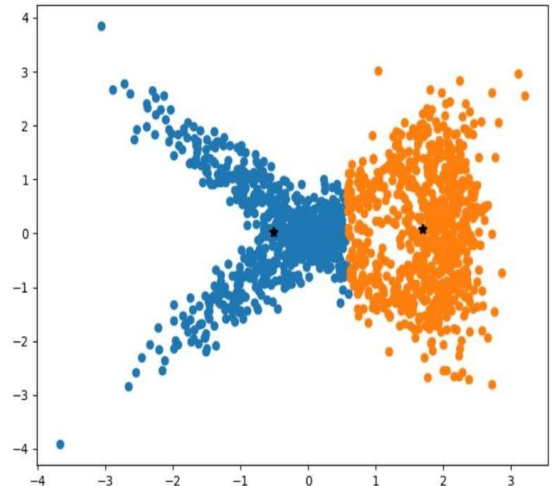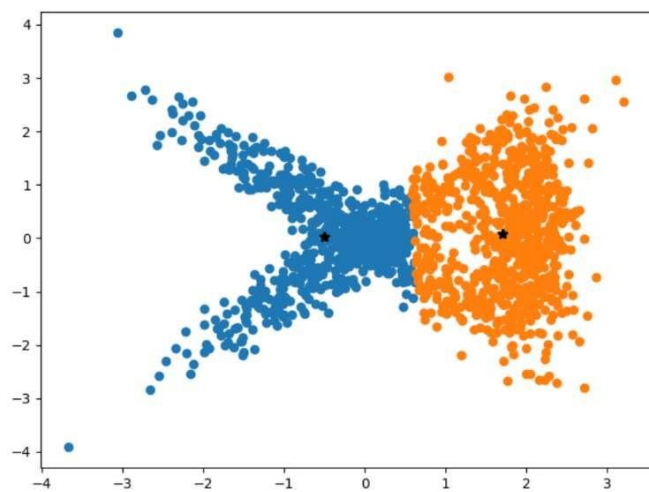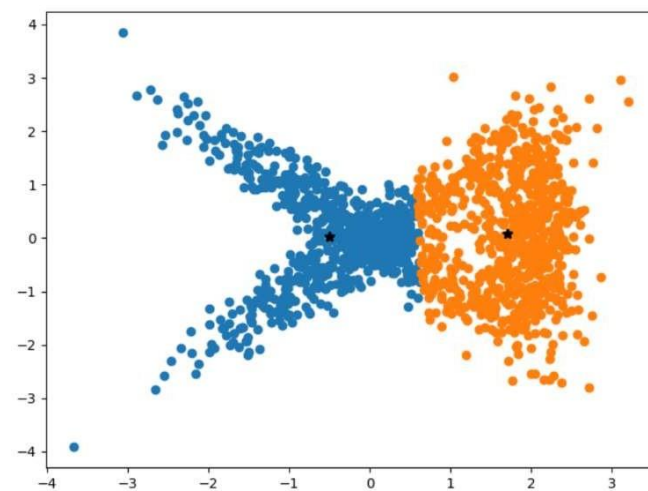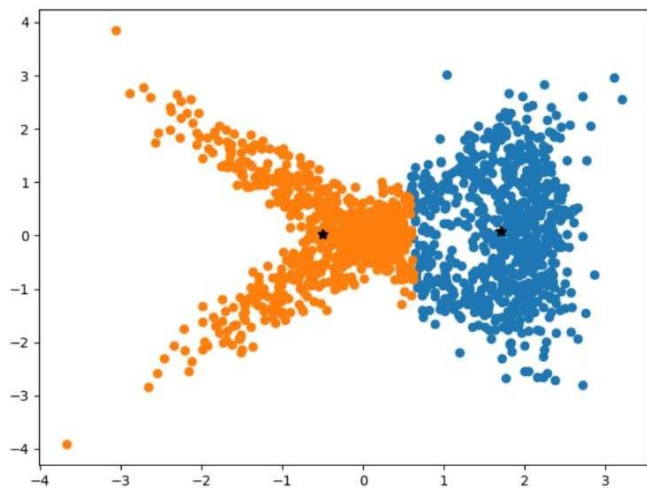
**Observations:**

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

    **r=1**                                                       **r=2**
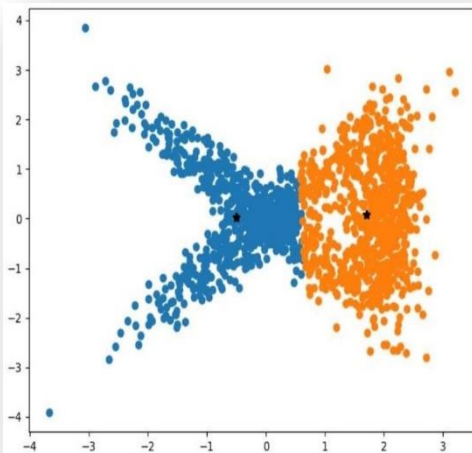
**r = 3**   **r = 4**

**r = 5**   **r = 6**
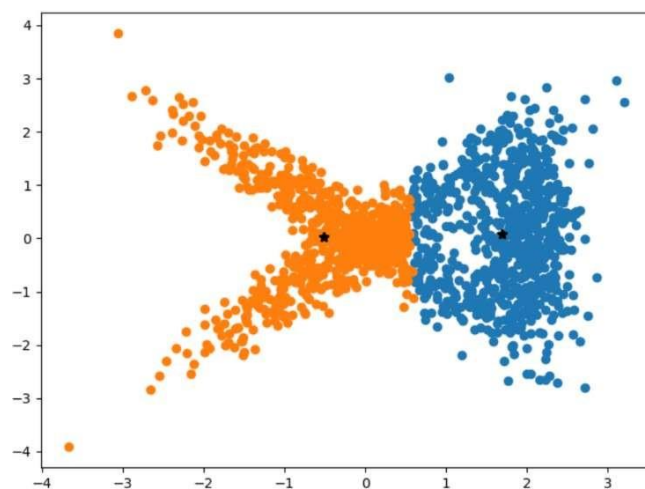
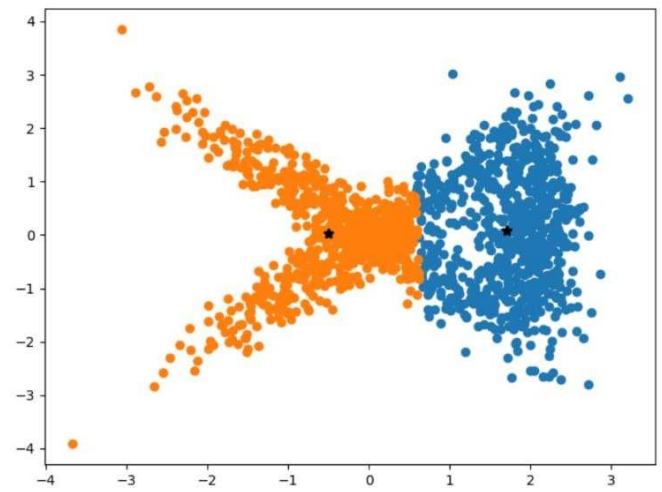**r = 7**   **r = 8**
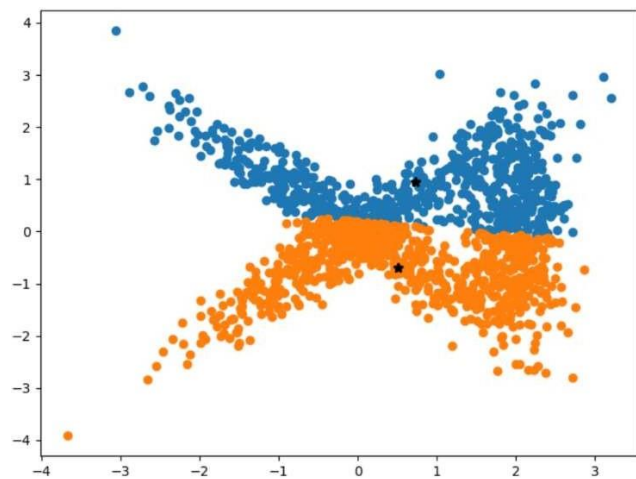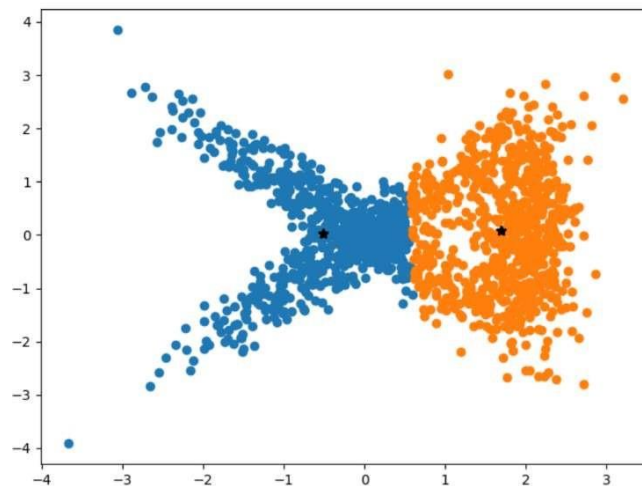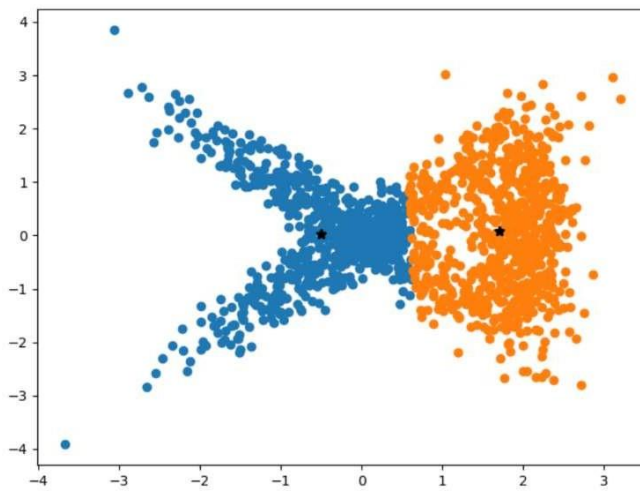
**r = 9**                                          **r = 10**
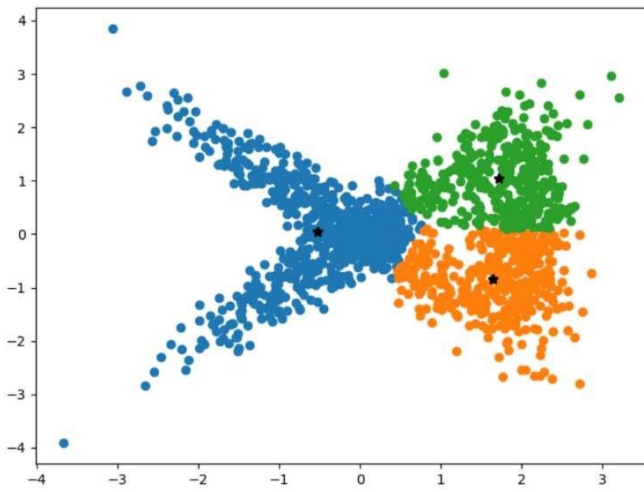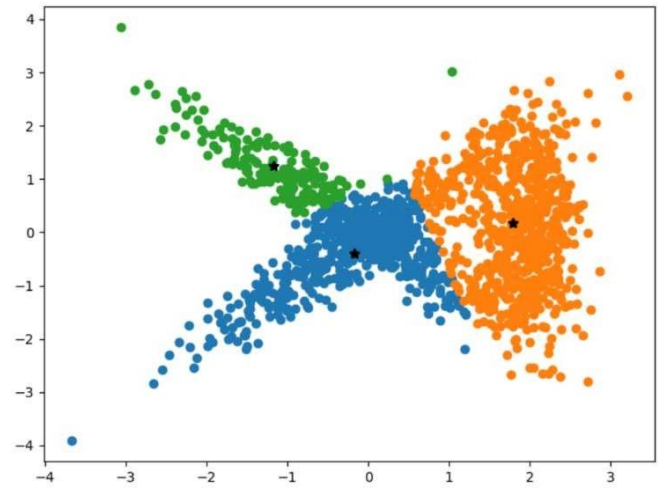


Errors for K=2

errors =  [2168.2788438938355, 2171.001969681453, 2171.001969681462, 2171.001969681453, 2171.001969681453, 2168.278843893843, 2171.001969681462, 2171.001969681453, 2171.001969681453, 2171.001969681453, 2171.001969681453]
min error when r = 1
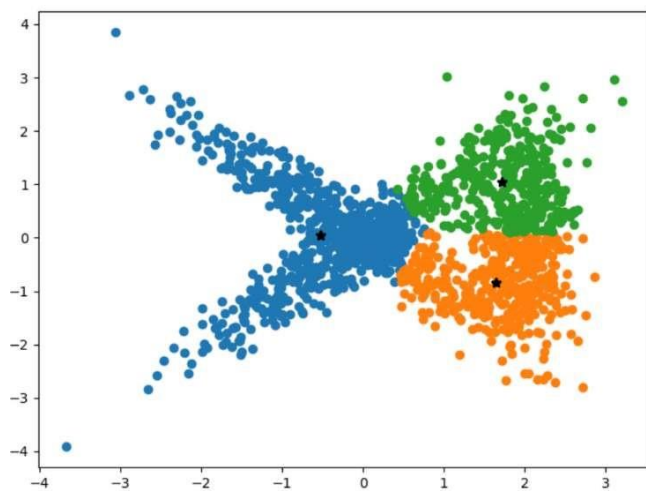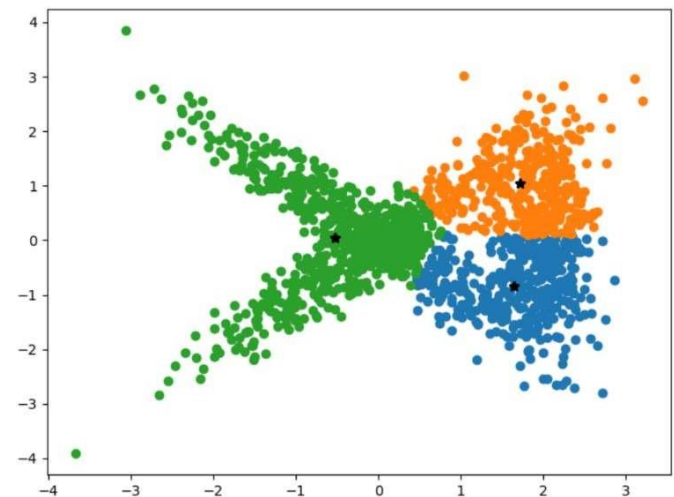error =  2168.2788438938355
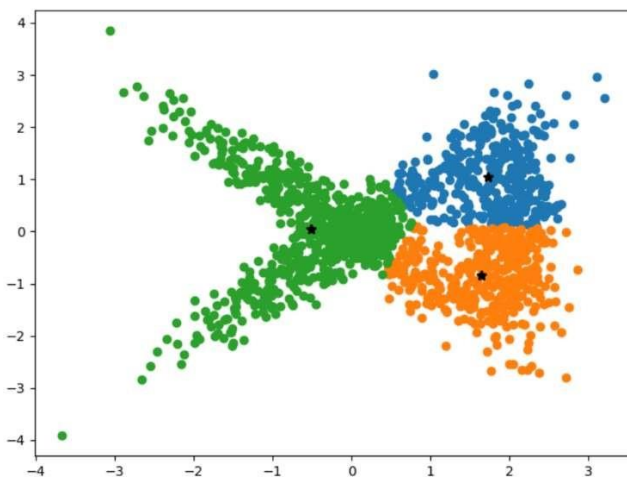
## 2. Number of Clusters K = 3

### r = 1



### r = 2



### r = 3



### r = 4



### r = 5



### r = 6

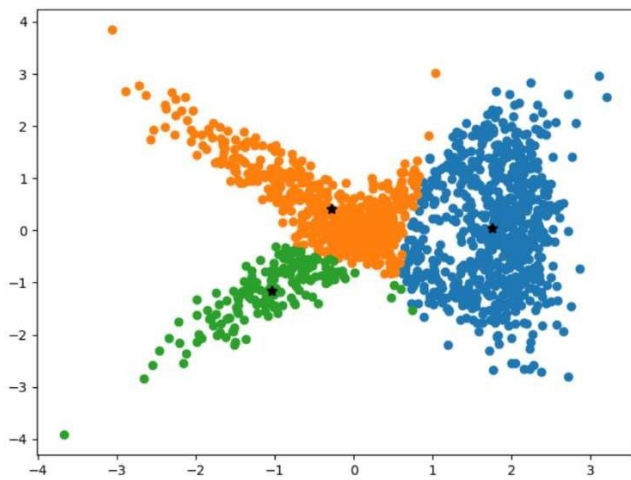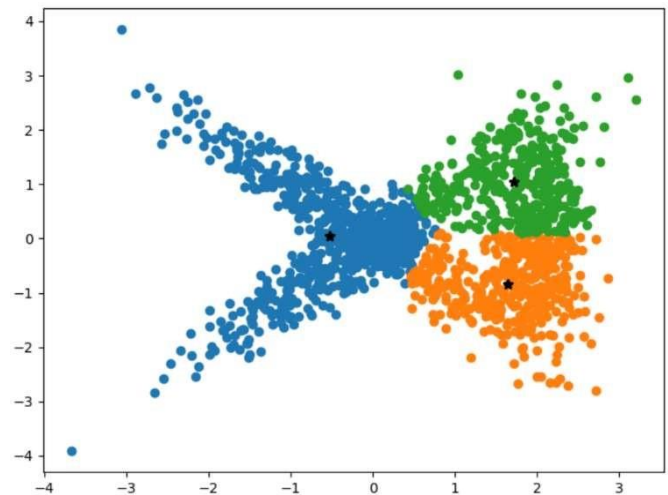## r = 7



## r = 8
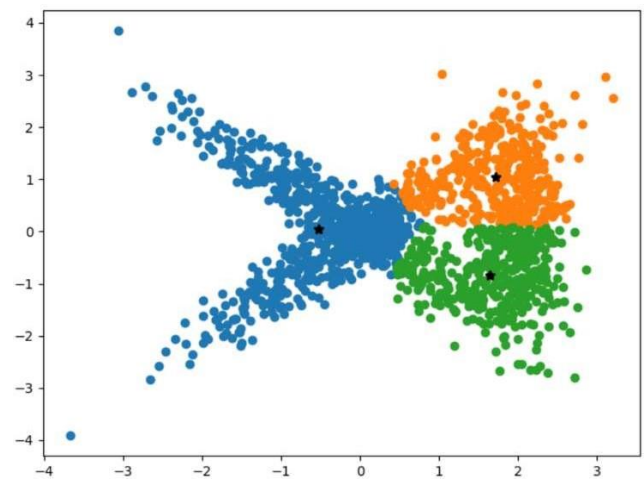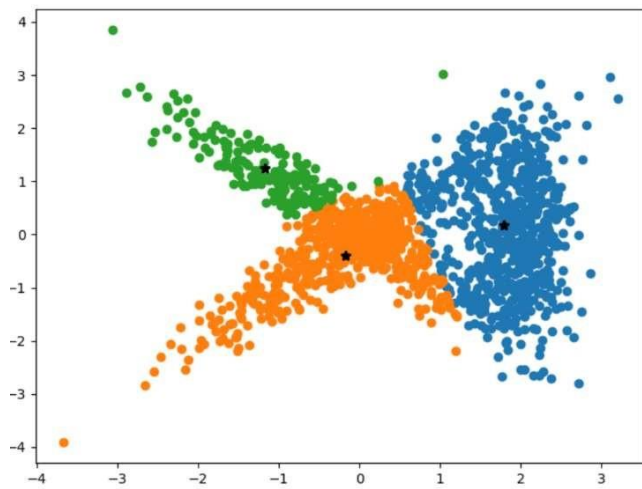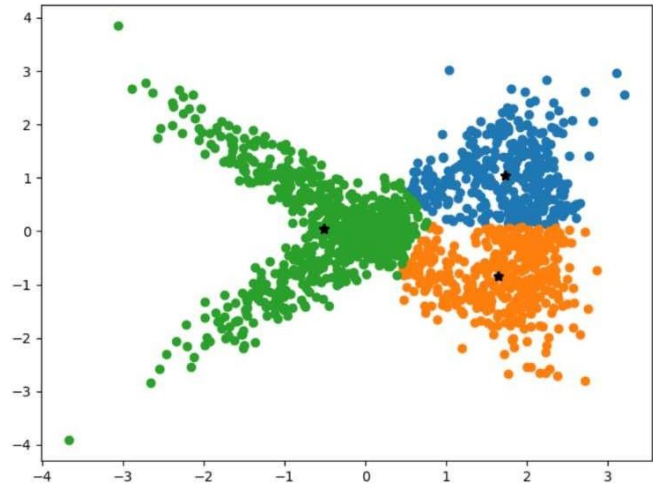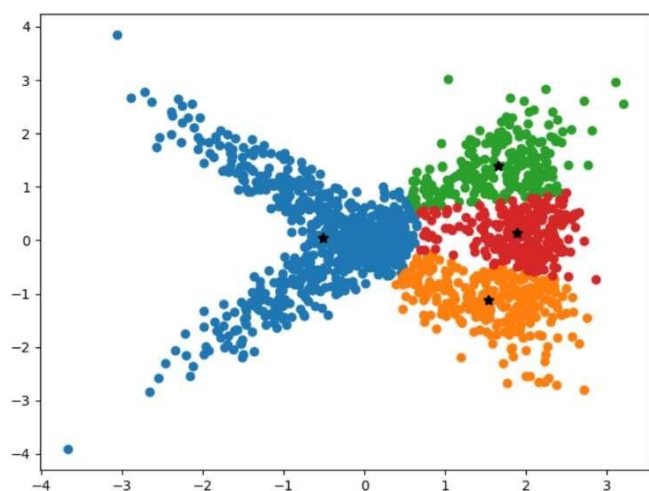


## r = 9



## r = 10



**Errors for K=3**

```
errors = [1912.9394977223042, 1479.5970650234513, 1478.4278229881193, 1912.939497722299, 1603.7708155559499, 1915.8717614424
424, 1478.4278229881165, 1478.4278229881143, 1479.5970650234558, 1912.939497722298]
min error when r = 8
error = 1478.4278229881143
```
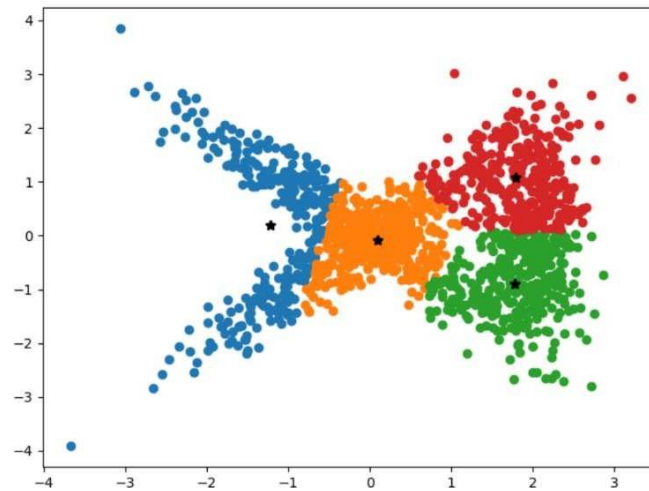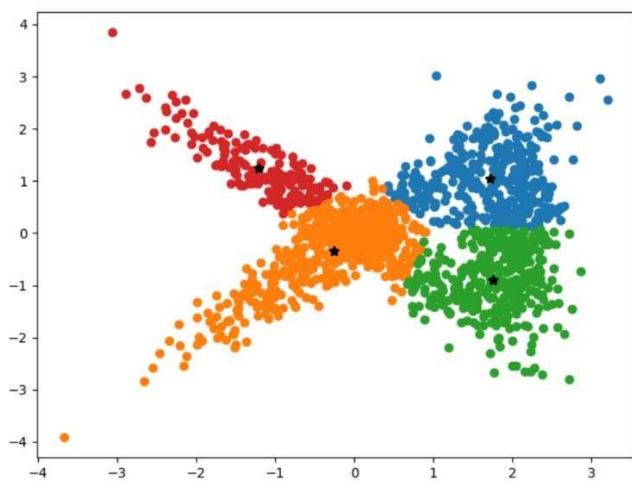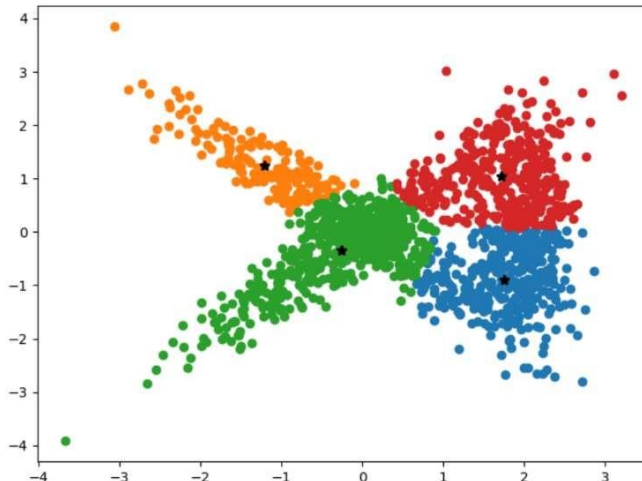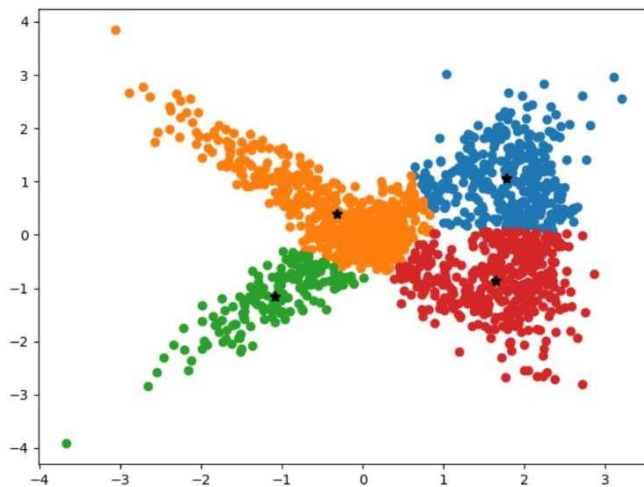
### 3) Number of clusters K = 4

**r = 1**



**r = 2**



**r = 3**



**r = 4**

**r = 5**

**r = 6**

**r = 7**

**r = 8**
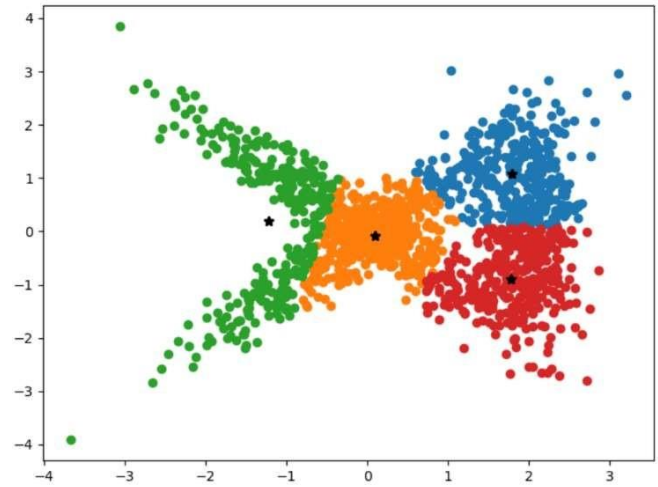
**r = 9**

**r = 10**

## Errors for K=4

```
errors = [1324.9651675045973, 1324.9651675045986, 1422.177645088136, 1420.470317927065, 1422.177645088134, 1293.510146347860
9, 1295.423959311399, 1325.0257551100406, 1422.177645088134, 1303.828957173732]
min error when r = 6
error = 1293.5101463478609
```

**4) Number of Clusters k = 5**

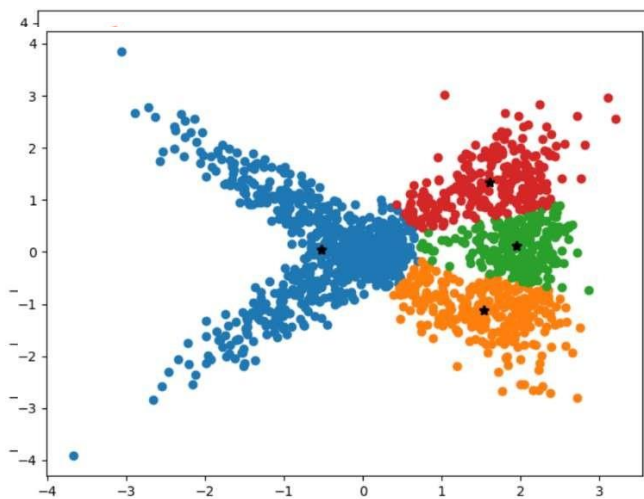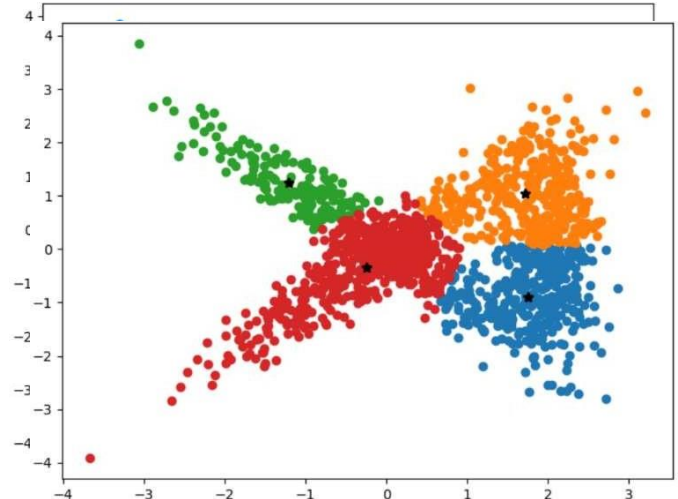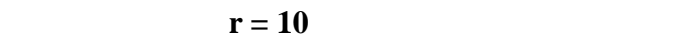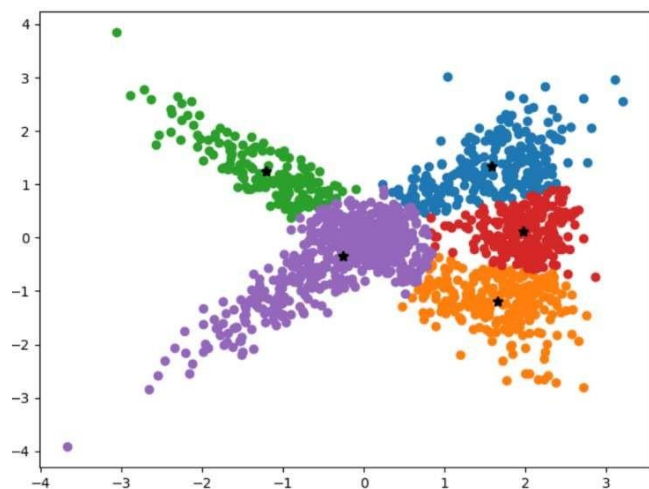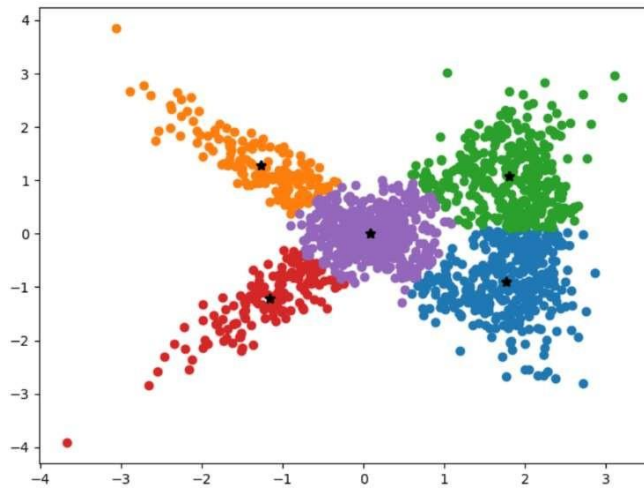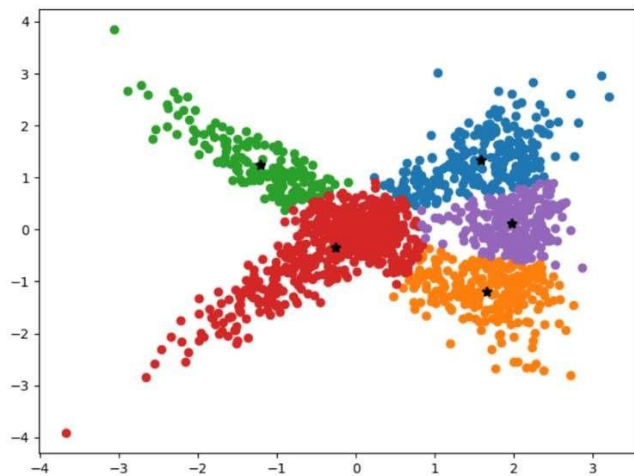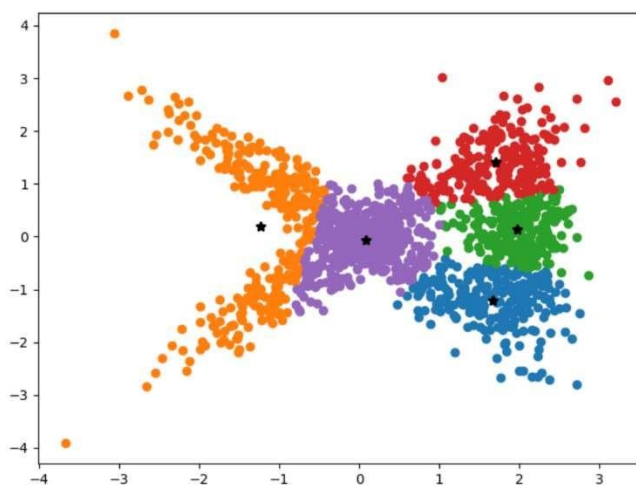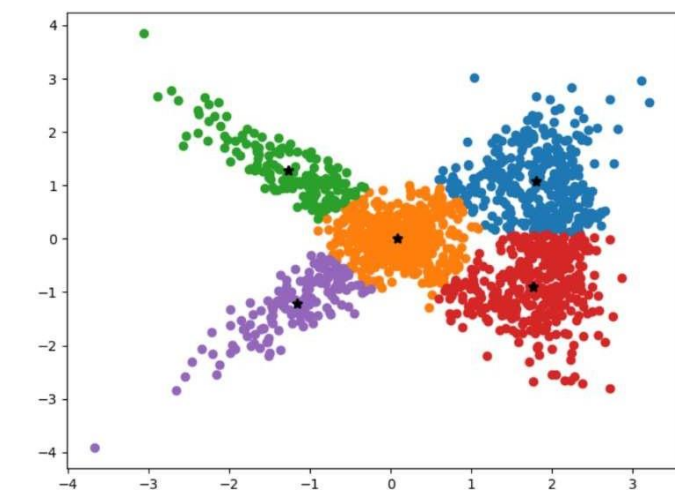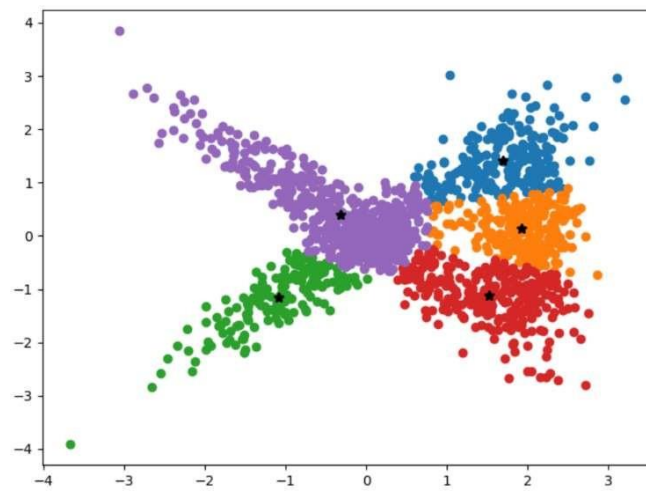r = 1

r = 2

r = 3

r = 4

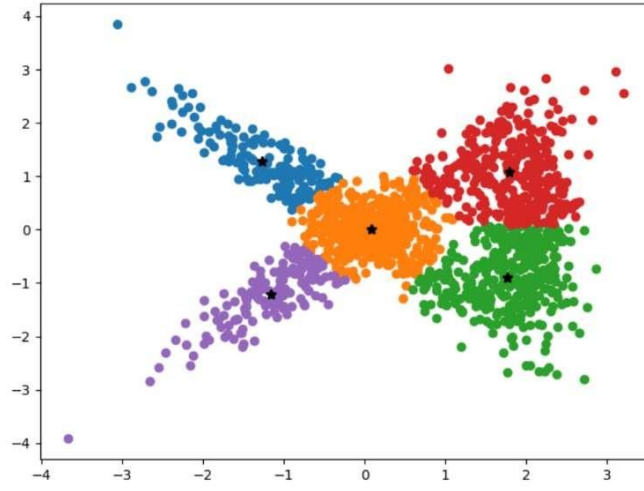r = 5      r = 6
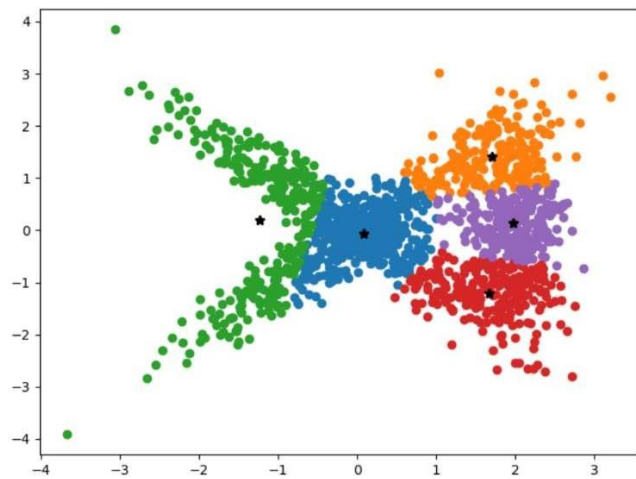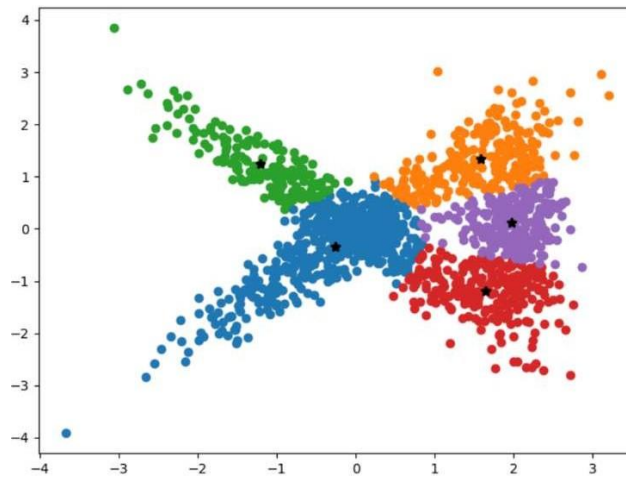
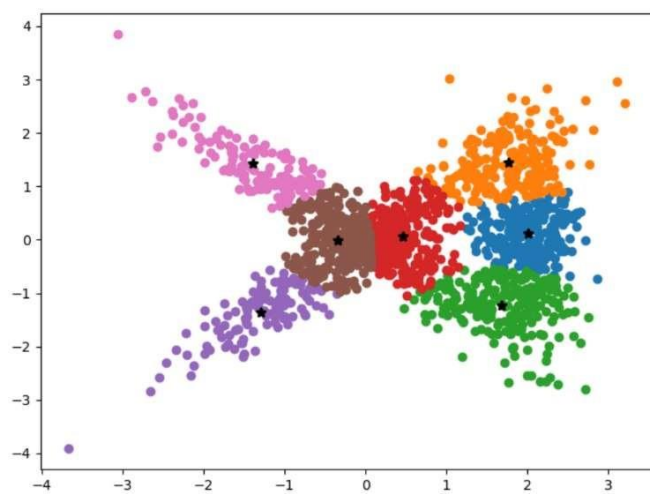r = 7      r = 8
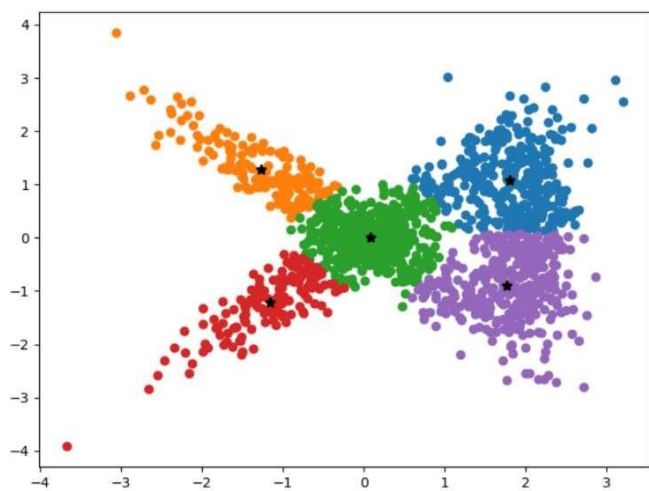
r = 9      r = 10

## Errors for K=5

```
errors =  [638.2989423487169, 1147.533249115753, 780.235642801088, 780.2356428010883, 1147.533249115754, 638.298942348717, 77
9.9152903382285, 1147.5332491157521, 1288.4819711122952, 1147.5332491157512]
min error when r = 1
error =  638.2989423487169
```
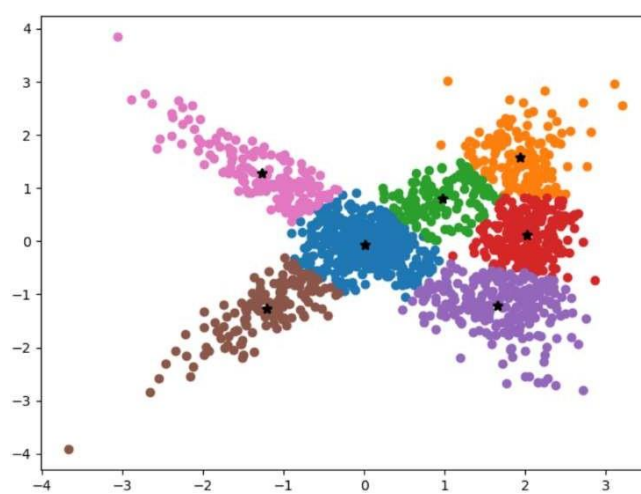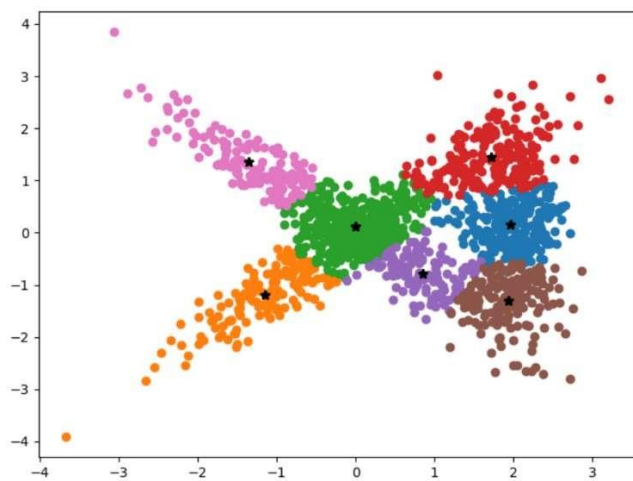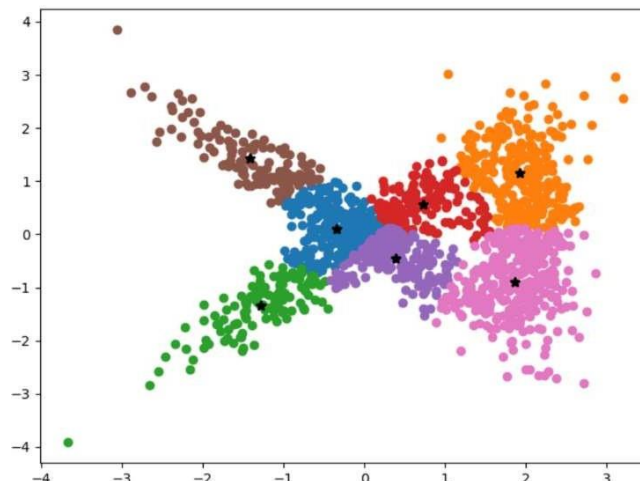
**5) Number of clusters k = 7**



r = 1

r = 2

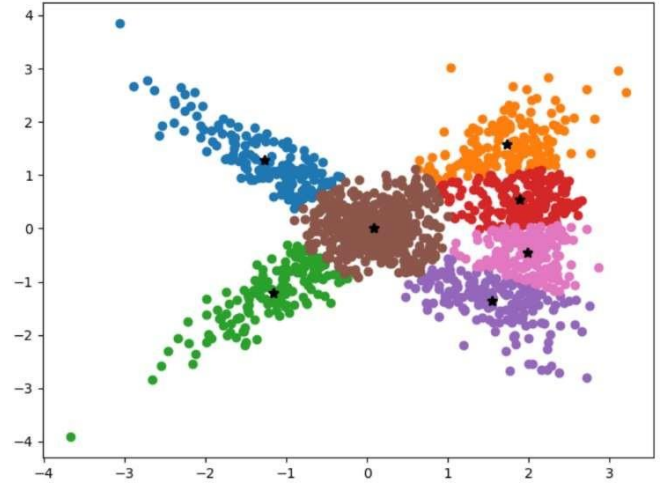r = 3

r = 4

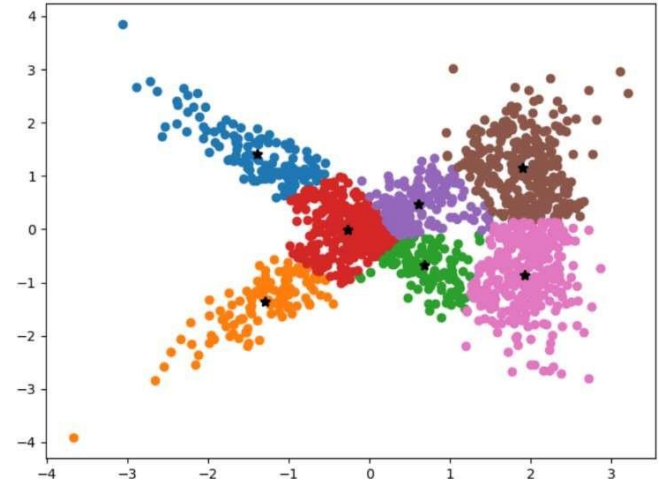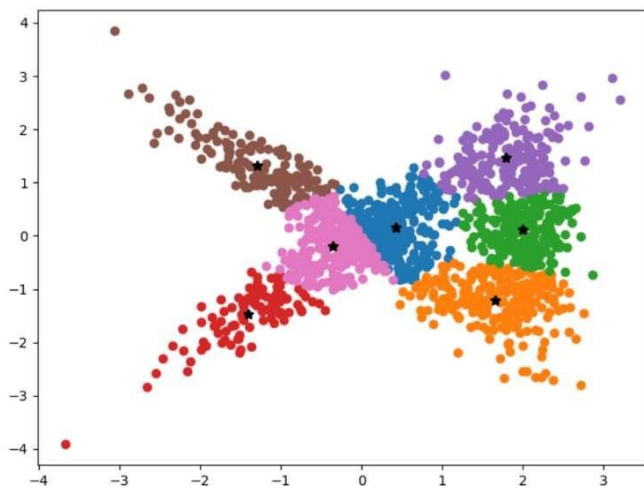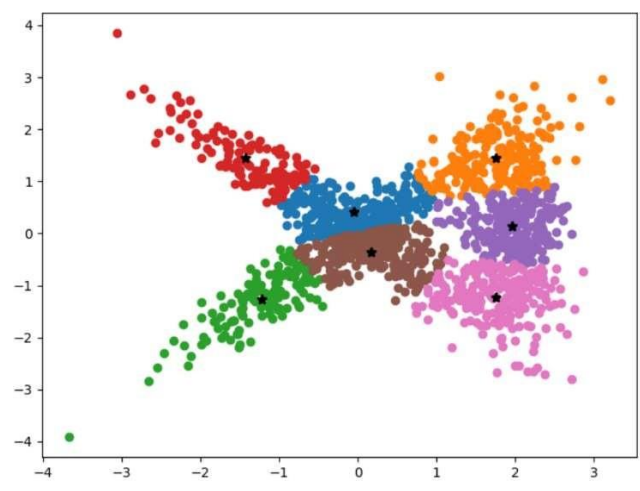**r = 5**   **r = 6**
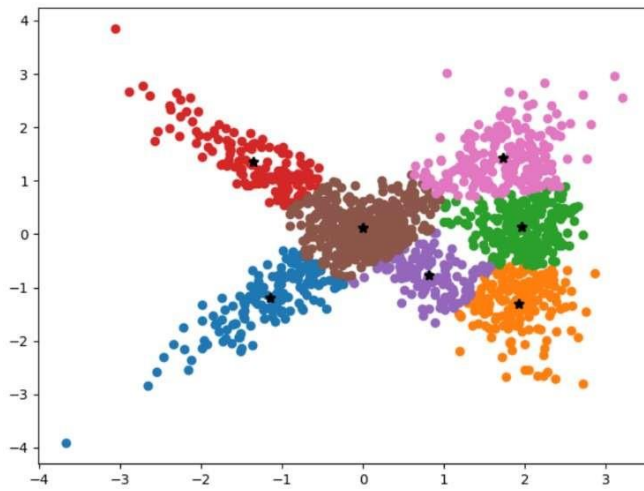
**r = 7**   **r = 8**
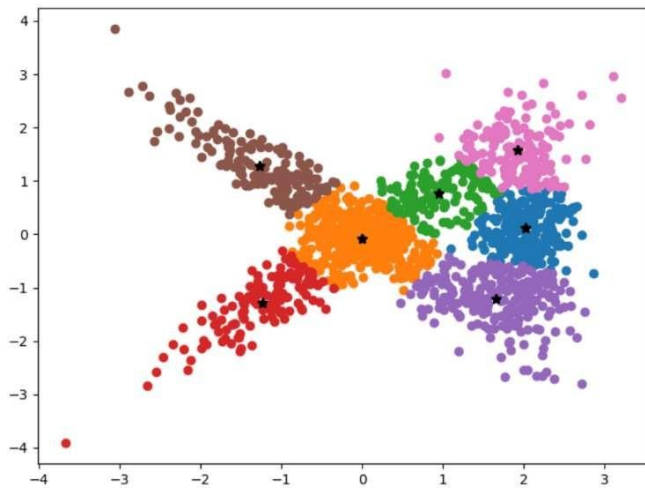
**r = 9**   **r = 10**

# Errors for K=7

```
errors =  [501.7170080682313, 613.35147061806, 566.8367585057923, 501.2878001639502, 617.7675609600572, 617.751406587475, 50
0.66599491088806, 610.4280482387062, 617.4867824312614, 617.751406587476]
min error when r = 7
error =  500.66599491088806
```

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 |