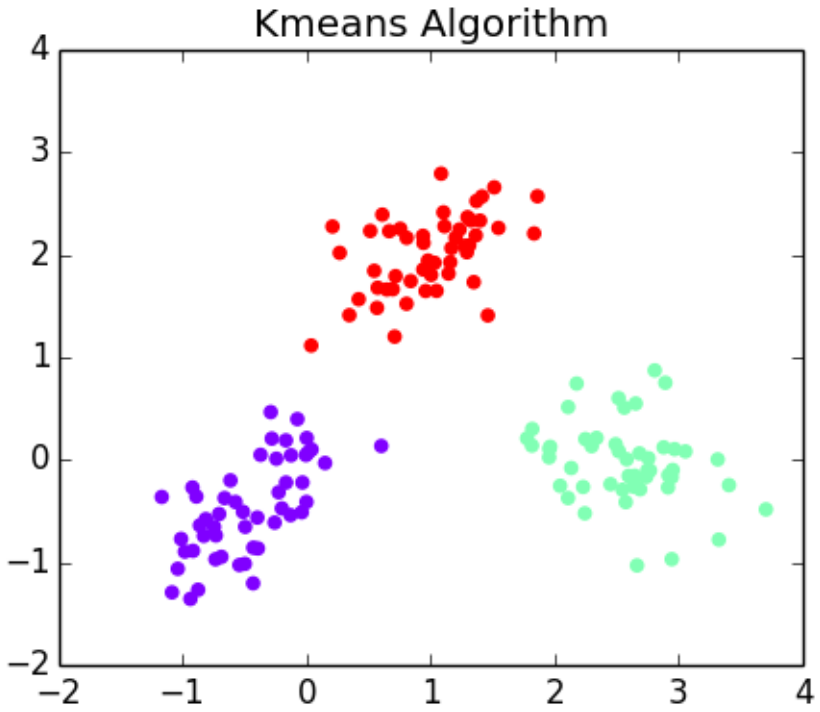
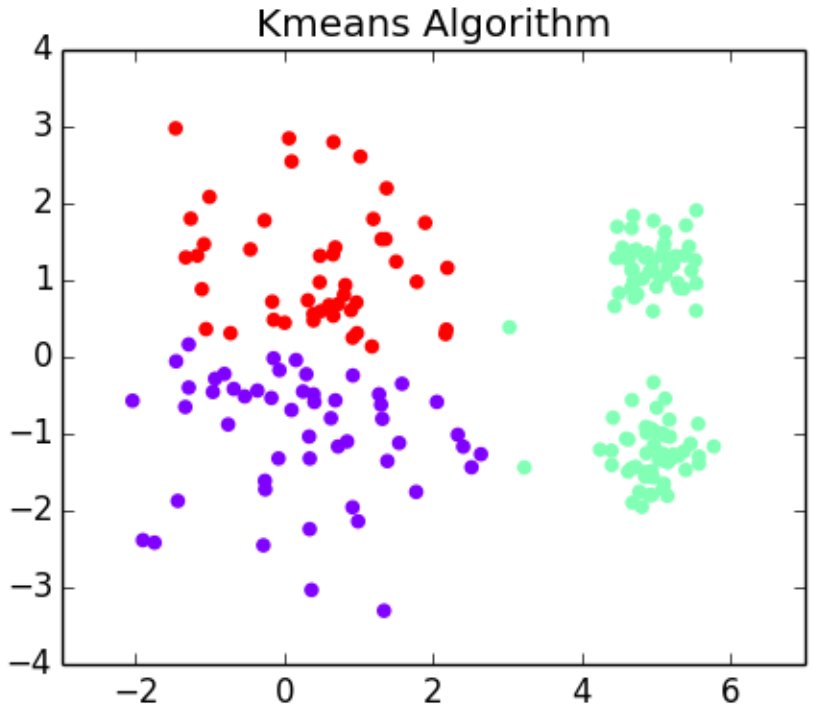
Data Mining HW-3

K-mean

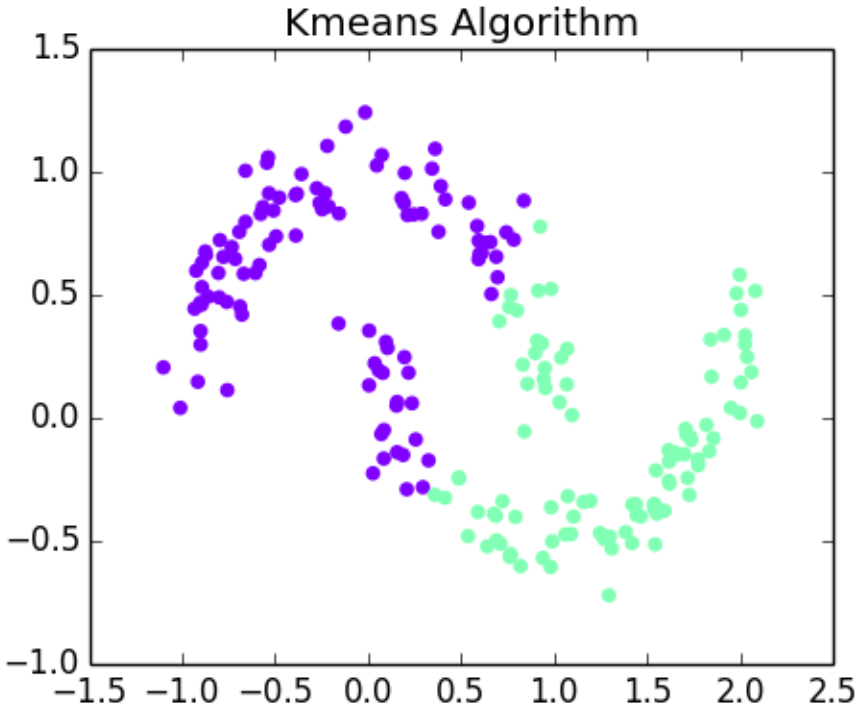
Dataset-1



Dataset-2


Kmeans Algorithm
4
3
2
1
0
−1
−2
−3
−4
−2
0
2
4
6

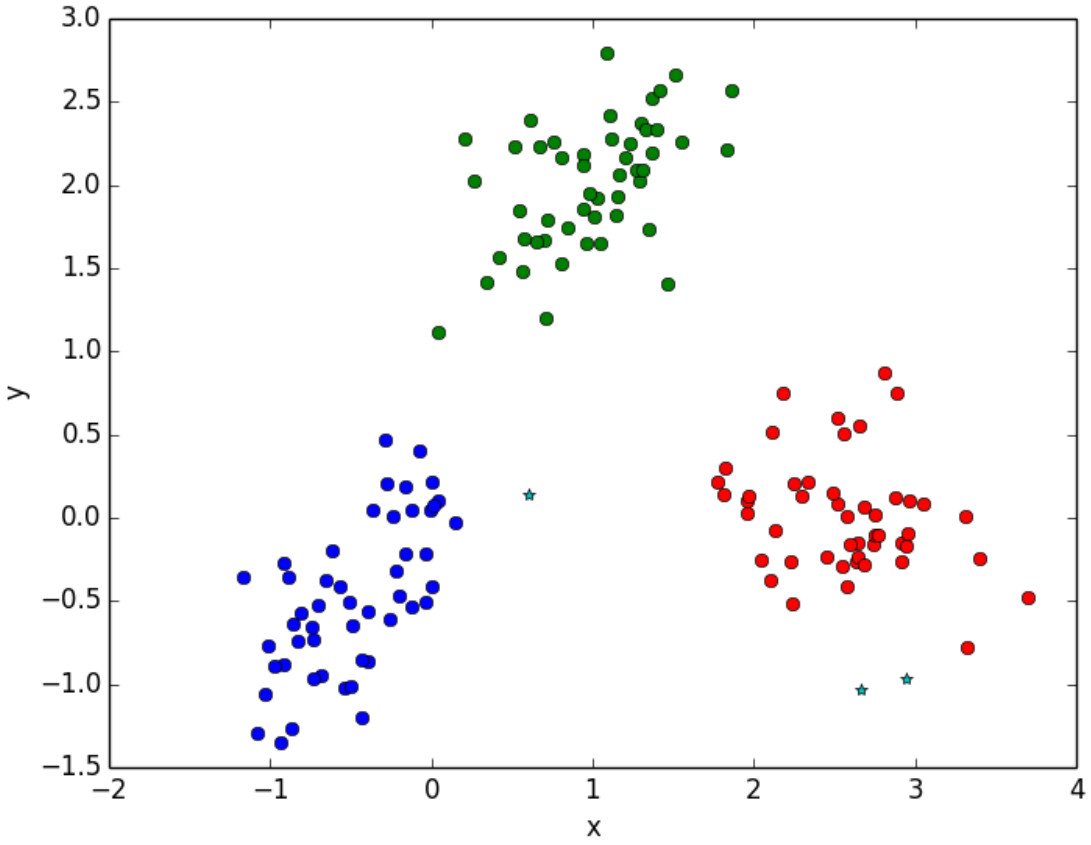
Dataset-3


Kmeans Algorithm
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5

Data	**Purity**	**Nmi**
Dataset-1	1	**1.5**
Dataset-2	**0.76**	**0.92**
Dataset-3	**0.78**	**0.23**

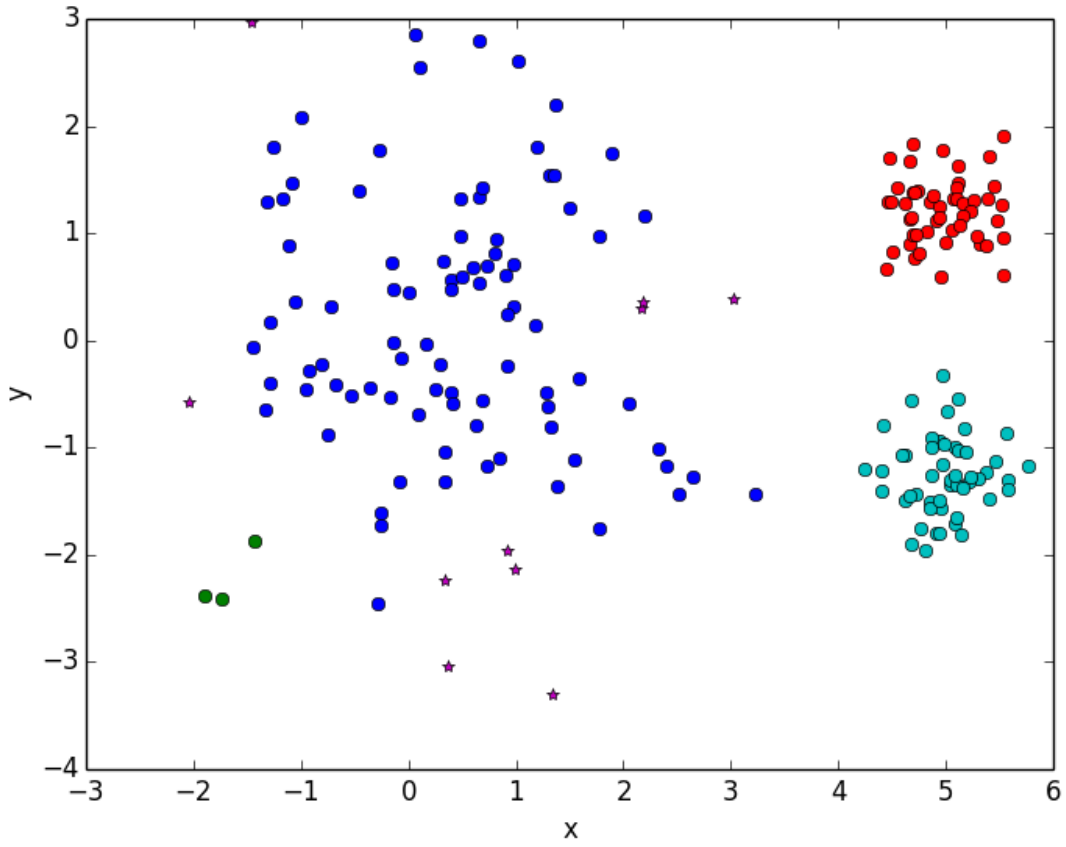
DBScan:

Dataset-1
Minpts = 3
Epsilon = 0.227



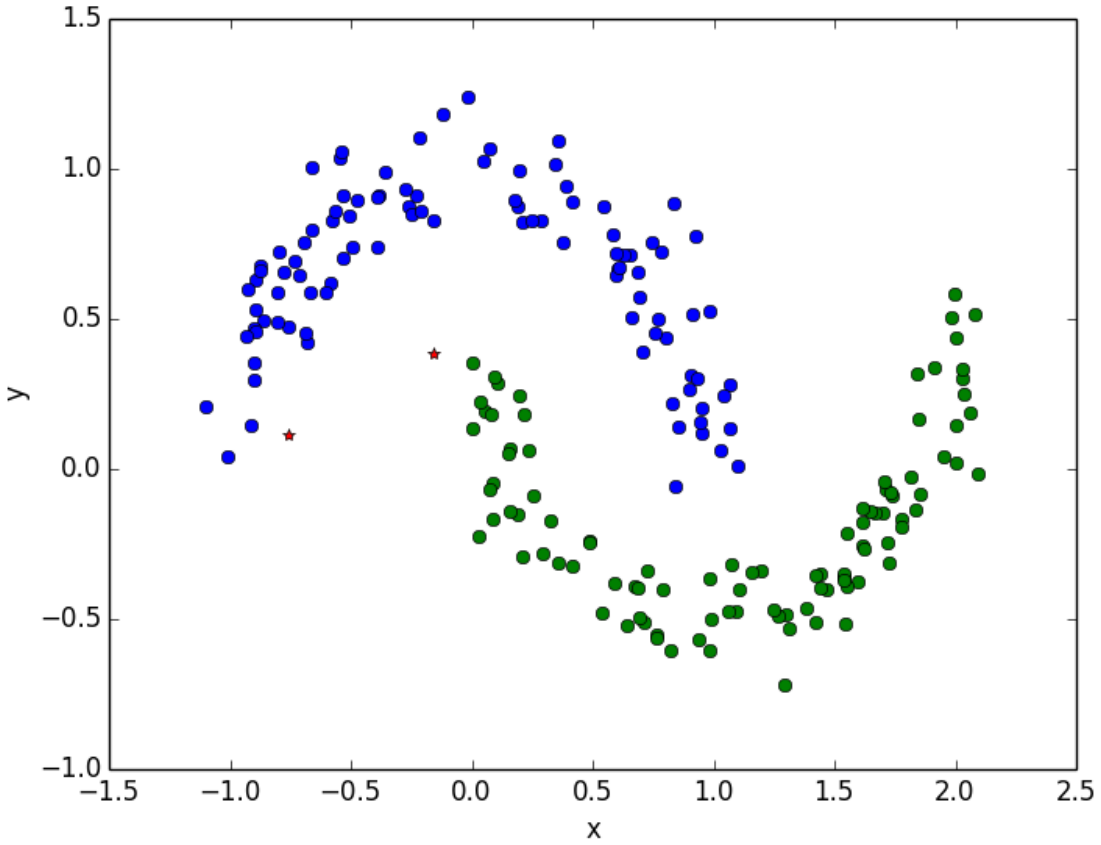
---

dataset-2
minpts = 3
epsilon = 0.64



---

dataset-3
minpts = 3
epsilon = 0.13



Data	**Purity**	**Nmi**
Dataset-1	1	**1.5**
Dataset-2	**1**	**1.5**
Dataset-3	**1**	**1**

Analysis:

By observing the data above we find that k-means is better for small size data and which is distributed evenly but dbscan is better for large size data and uneven distribution and whose data can be separated in non-linear fashion.