Northeastern University, Data Mining Techniques - CS6220
Fall 2017

---

# Solutions to Homework 2, Part 2

---

Nakul Camasamudram

October 24, 2017

# 1. K-Means

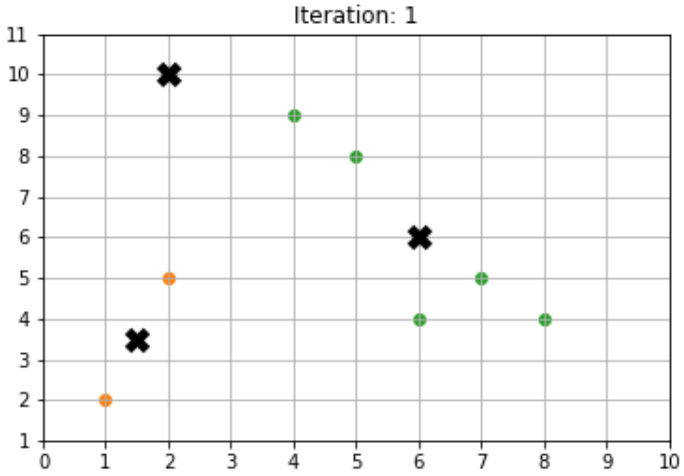
**Solution:**

- **Iteration 1:**
  The initial centroids are $C_1 = (2, 10)$ $C_2 = (1, 2)$ $C_3 = (5, 8)$. Let $\mathrm{D}(C_i)$ represent the **euclidean** distance between the respective point and the $i^{th}$ centroid.

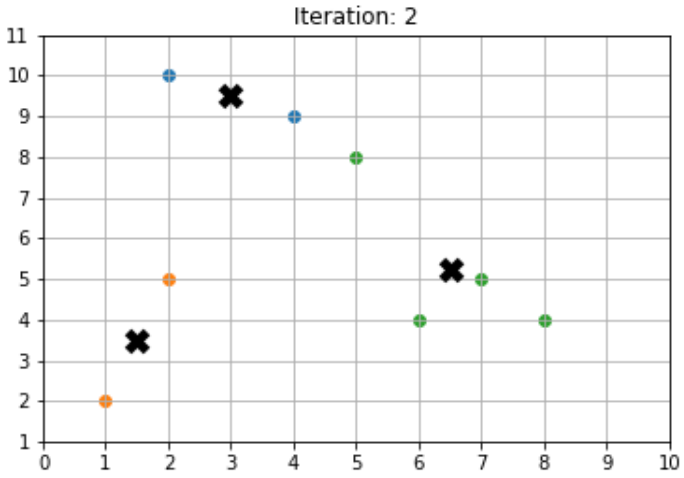
  The $E$-step:

| **Data Point** | **D($C_1$)** | **D($C_2$)** | **D($C_3$)** | **Optimal centroid** |
|---|---|---|---|---|
| (4,9) | 2.23606797749979 | 7.615773105863909 | **1.4142135623730951** | $C_3$ |
| (2,10) | **0.0** | 8.06225774829855 | 3.605551275463989 | $C_1$ |
| (1,2) | 8.06225774829855 | **0.0** | 7.211102550927978 | $C_2$ |
| (2,5) | 5.0 | **3.1622776601683795** | 4.242640687119285 | $C_2$ |
| (6,4) | 7.211102550927978 | 5.385164807134504 | **4.123105625617661** | $C_3$ |
| (8,4) | 8.48528137423857 | 7.280109889280518 | **5.0** | $C_3$ |
| (7, 5) | 7.0710678118654755 | 6.708203932499369 | **3.605551275463989** | $C_3$ |
| (5, 8) | 3.605551275463989 | 7.211102550927978 | **0.0** | $C_3$ |

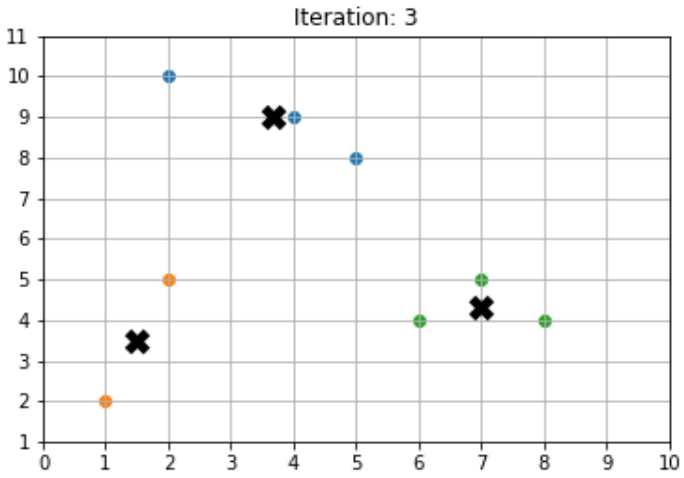
  The $M$-step:
  $C_1$ = mean[(2, 10)] = (2.0, 10.0)
  $C_2$ = mean[(1, 2), (2, 5)] = (1.5, 3.5)
  $C_3$ = mean[(4, 9), (6, 4), (8, 4), (7, 5), (5, 8)] = (6.0, 6.0)



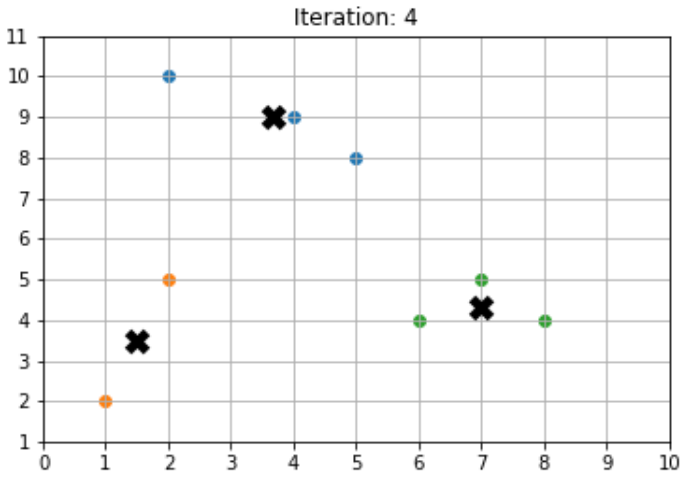
- **Iteration 2:**




- **Iteration 3:**




- **Iteration 4:**


Iteration: 4
11
10
9
8
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8
9
10

## 2. Agglomerative Hierarchical

**Solution:**

## 3. DBSCAN

**Solution:**