

## Exercise 1

(a)

The center of the first cluster after one iteration is [5.171428571428572, 3.1714285714285713].

(b)

The center of the second cluster after two iterations is [5.3, 4.0].

(c)

The center of the third cluster when the clustering converges is [6.2, 3.025]

(d)

2 iterations are required for the clusters to converge.

注：代码见文件T1.ipynb

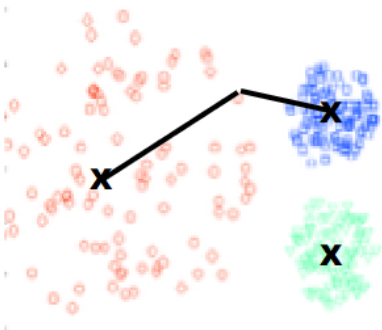
## Exercise 2

(a) A2

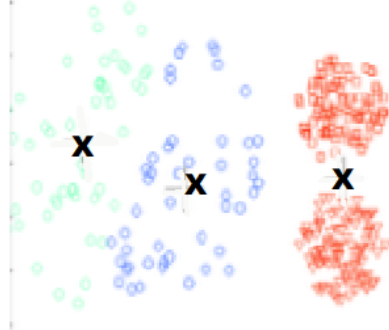
(b) B2

可以看到 B1 中那个红色点明显离蓝色簇的质心距离更近，所以 B2 更像是K-Means聚类结果。

B 1

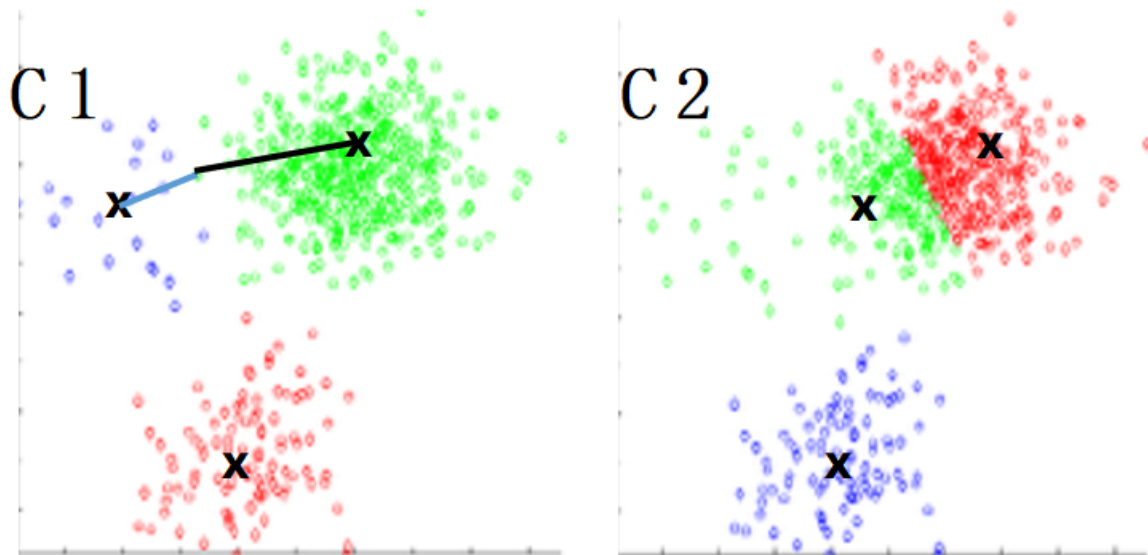


B 2



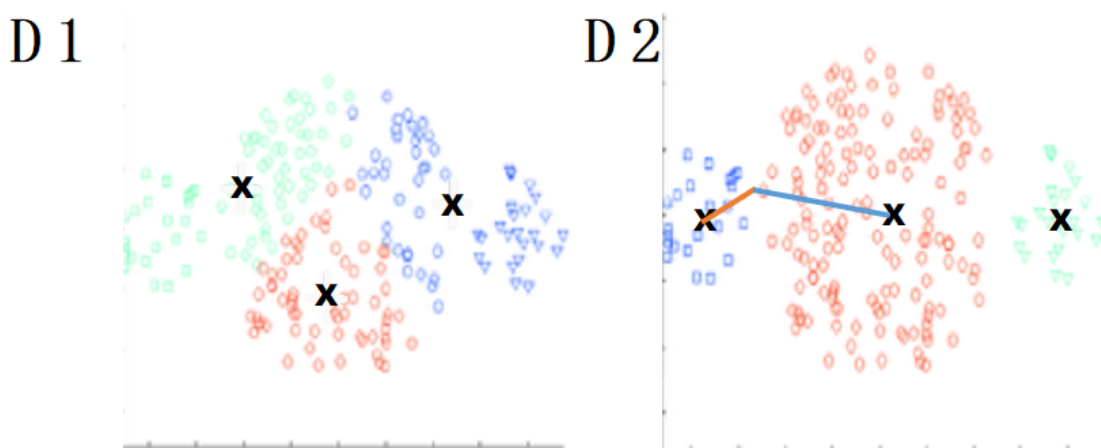
(c) C2

可以看到 C1 中那个绿色点明显离蓝色簇的质心距离更近，所以 C2 更像是K-Means聚类结果。



(d) D1

可以看到 D2 中那个红色点明显离蓝色簇的质心距离更近，所以 D1 更像是K-Means聚类结果。



(e) E2

(f) F2

(g)

- 根据 K-Means 的迭代过程，收敛后，每个点都应该离它的质心最近，而不是离其他簇的质心近。
- K-Means 的聚类结果中，两个簇通常会有明显的中垂线。
- K-Means 只能捕捉数据点之间的距离，不能捕捉数据的结构，因此 K-Means 无法识别具有复杂结构的数据。只能根据点的距离判断聚类结果是否是 K-Means 算法的结果。

(h)

K-Means 在 F 数据中表现不佳。K-Means 只能捕捉数据点之间的距离，不能捕捉数据的结构，因此 K-Means 无法识别具有复杂结构的数据。

DBSCAN 算法是个处理 F 图数据更好的算法。它考虑了数据点的邻居结点信息，这样就能捕捉到数据的结构信息。

## Exercise 3

- 信息检索的应用程序例子有搜索结果聚类。其中搜索结果是指为响应查询而返回的文档。信息检索中搜索结果的默认表示是一个简单的列表。用户从上到下扫描列表，直到找到他们要查找的信息。相反，搜索结果聚类会对搜索结果进行聚类，以便类似文档一起显示。扫描几个连贯的组通常比许多单个文档更容易。如果搜索词具有不同的词义，则此功能特别有用。
- 数据挖掘中的应用例子有实现精准销售。通过分析不同消费者在消费同一类商品或服务时的不同特点，可以指定更好的营销组合，从而获取最大的消费者剩余。