

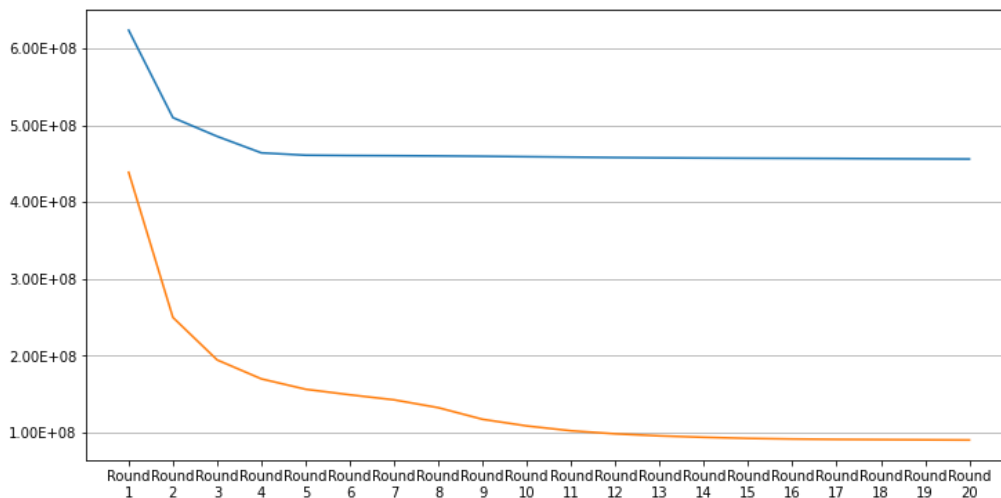
2021 Introduction to Massive Data Analysis

HW3 - K-means

107062117 李采蓉

(a.) Euclidean distance

1. A plot of cost 、iteration for c1 and c2



2. Percentage improvement values and explanation.

For c1: **26.885383** %

For c2: **79.437750** %

在 improvement 上 c2 有明顯較好的表現，因為 c1 的 centroid 是隨機選取，平均散佈在所有資料點中，因此一開始選到的 centroid 有機會不能代表所有分群或與實際分群狀況差異較大，導致分類的結果較差。c2 則在一開始就選距離遠的點當 centroid，因此更有可能代表實際分佈，有較好的分群結果。

	c1	c2
Round 1	6.24E+08	4.39E+08
Round 2	5.10E+08	2.50E+08
Round 3	4.85E+08	1.94E+08
Round 4	4.64E+08	1.70E+08
Round 5	4.61E+08	1.56E+08
Round 6	4.61E+08	1.49E+08
Round 7	4.60E+08	1.43E+08
Round 8	4.60E+08	1.32E+08
Round 9	4.60E+08	1.17E+08
Round 10	4.59E+08	1.09E+08
Round 11	4.58E+08	1.02E+08
Round 12	4.58E+08	9.83E+07
Round 13	4.58E+08	9.56E+07
Round 14	4.57E+08	9.38E+07
Round 15	4.57E+08	9.24E+07
Round 16	4.57E+08	9.15E+07
Round 17	4.57E+08	9.10E+07
Round 18	4.56E+08	9.08E+07
Round 19	4.56E+08	9.05E+07
Round 20	4.56E+08	9.02E+07

3. The Distances for all pairs of centroids.

(1.) The **Manhattan Distances** for all pairs of centroids with **c1**

[illegible]

(2.) The **Euclidean Distances** for all pairs of centroids with **c1**

[illegible]

(3.) The **Manhattan Distances** for all pairs of centroids with **c2**

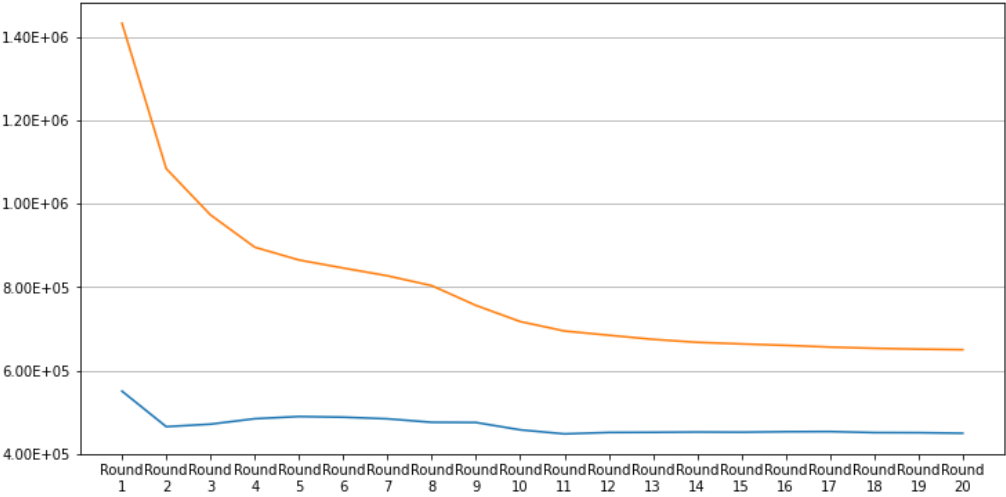
[illegible]

(4.) The **Euclidean Distances** for all pairs of centroids with **c2**

[illegible]

(b.) Manhattan distance

1. A plot of cost 、 iteration for c1 and c2



2. Percentage improvement values and explanation.

For c1: 18.378954 %

For c2: 54.685694 %

在 improvement 上 c2 有明顯較好的表現，與先前所述相同，c2 選取距離遠的點當做 centroid，比 c1 的隨機選取更貼近實際可能的分群狀況。

	c1	c2
Round 1	5.50E+05	1.43E+06
Round 2	4.65E+05	1.08E+06
Round 3	4.71E+05	9.73E+05
Round 4	4.84E+05	8.96E+05
Round 5	4.89E+05	8.65E+05
Round 6	4.88E+05	8.46E+05
Round 7	4.84E+05	8.27E+05
Round 8	4.75E+05	8.04E+05
Round 9	4.75E+05	7.56E+05
Round 10	4.57E+05	7.17E+05
Round 11	4.47E+05	6.95E+05
Round 12	4.51E+05	6.84E+05
Round 13	4.51E+05	6.75E+05
Round 14	4.52E+05	6.67E+05
Round 15	4.52E+05	6.64E+05
Round 16	4.53E+05	6.60E+05
Round 17	4.53E+05	6.56E+05
Round 18	4.51E+05	6.53E+05
Round 19	4.50E+05	6.51E+05
Round 20	4.49E+05	6.50E+05

3. The Distances for all pairs of centroids.

(1.) The **Manhattan Distances** for all pairs of centroids with **c1**

[illegible]

(2.) The **Euclidean Distances** for all pairs of centroids with **c1**

[illegible]

(3.) The **Manhattan Distances** for all pairs of centroids with **c2**

[illegible]

(4.) The **Euclidean Distances** for all pairs of centroids with **c2**

[illegible]