Data Mining HW2

**Clustering: K-means**

(a)

1.

Euclidean-random:

2.

Euclidean-random/Euclidean-farthest first:

3.

Euclidean-random: 196,114,105,85,63,37

Euclidean-farthest first: 196,186,118,41,34,25

(b)

1.

Manhattan-random:

2.

Manhattan-random/Manhattan-farthest first:

3.

Manhattan-random:195,108,106,91,52,48

Manhattan-farthest first:197,186,117,42,33,25,

(c)

Farthest first initial收斂比random initial快,可能原因為,random 選取的centroid,其分布可能與最終分群結果差很大,但通常使用Farthest first 的方式選取centroid ,其結果較接近最終分群結果,故收斂較快。

(d)

(e)

Yes,使用大的cluster數量可以將含有各種不同的特徵的點細分到不同的群,使得每群中的點同質性更高,不同群間的點異質性更高,故分群結果更好。

(f)使用weka