ISYE 6740 – Midterm 1

Nicholas Chao – [nchao3@gatech.edu](mailto:nchao3@gatech.edu)

1. **K-Means**

Euclidean Distance between two vectors, and :

Manhattan Distance between two vectors, and :

Geometric Mean of set of vectors:

Initial Centroid A:

Initial Centroid B:

**Using Euclidean Distance:**

1. Datapoint 1 -> Cluster B

Datapoint 2 -> Cluster A

Datapoint 3 -> Cluster B

Datapoint 4 -> Cluster B

Datapoint 5 -> Cluster A

Datapoint 1:

Distance from A:

Distance from B:

Datapoint 2:

Distance from A:

Distance from B:

Datapoint 3:

Distance from A:

Distance from B:

Datapoint 4:

Distance from A:

Distance from B:

Datapoint 5:

Distance from A:

Distance from B:

1. Cluster A’s centroid becomes (-1.5, -0.5)

Cluster B’s centroid becomes (1.667, 0.667)

Cluster A:

Geometric Mean:

Cluster B:

Geometric Mean:

1. The algorithm **will not** terminate after one step. Another iteration will result in the follow clusters:

Datapoint 1 -> Cluster B

Datapoint 2 -> Cluster A

Datapoint 3 -> Cluster B

Datapoint 4 -> Cluster A

Datapoint 5 -> Cluster A

Datapoint 1:

Distance from A: 4.30

Distance from B:

Datapoint 2:

Distance from A:

Distance from B:

Datapoint 3:

Distance from A:

Distance from B:

Datapoint 4:

Distance from A:

Distance from B:

Datapoint 5:

Distance from A:

Distance from B:

**Using Manhattan Distance:**

1. Datapoint 1 -> Cluster B

Datapoint 2 -> Cluster B

Datapoint 3 -> Cluster B

Datapoint 4 -> Cluster A

Datapoint 5 -> Cluster A

Datapoint 1:

Distance from A:

Distance from B:

Datapoint 2: (

Distance from A:

Distance from B:

Datapoint 3: (

Distance from A:

Distance from B:

Datapoint 4: (0

Distance from A:

Distance from B:

Datapoint 5: (-2

Distance from A:

Distance from B:

1. Cluster A’s centroid becomes (-1, -1.5)

Cluster B’s centroid becomes (1.33, 1.33)

Cluster A:

Geometric Mean:

Cluster B:

Geometric Mean:

1. The algorithm **will not** terminate after one step. Another iteration will result in the follow clusters:

Datapoint 1 -> Cluster B

Datapoint 2 -> Cluster A

Datapoint 3 -> Cluster B

Datapoint 4 -> Cluster A

Datapoint 5 -> Cluster A

Datapoint 1:

Distance from A: 6.5

Distance from B:

Datapoint 2:

Distance from A:

Distance from B:

Datapoint 3:

Distance from A:

Distance from B:

Datapoint 4:

Distance from A:

Distance from B:

Datapoint 5:

Distance from A:

Distance from B: