**Course: Data Mining**

**Student ID: 20C14001**

**Student Name: Le Duong Tuan Anh**

**Final Exam**

**Initial clustroids (C):**

**C1 (Jim): (0, 0, 0, 0)**

**C2 (Mary): (1, 0, 1, 0)**

Iteration 1:

Clustroids:

C1: (0, 0, 0, 0)

C2: (1, 0, 1, 0)

Jaccard distance (JD) between every patient to clustroids:

|  |  |  |
| --- | --- | --- |
| **Name** | **JD to Clustroid 1** | **JD to Clustroid 2** |
| Jack |  |  |
| Mary |  |  |
| Jim |  |  |
| Rose |  |  |
| Tom |  |  |

Depend on JD, we have two clusters:

Cluster 1: (Jim, Tom)

Cluster 2: (Jack, Mary, Rose)

Note: Tom can be belonged to both.

**New clustroids :**

**Cluster 1:**

This cluster has 2 samples so any is also valid if decided as clustroid. We take Jim as clustroid

* **Clustroid: Jim (total dist: 1)**

**Cluster 2:**

JDist(Jack, Mary): 0.5

JDist(Jack, Rose): 1

JDist(Mary , Rose): 0.75

* **Clustroid: Marry (total dist: 1.25, other total dist are 1.5 (Jack) and 1.75 (Rose)).**

**Clustroids are not updated 🡪 Terminate clustering process.**