Nick Palacio

Raj Dasgupta

October 25, 2018

Introduction to AI

Homework 3

Mathematics Part

**Question 1.**

*For the medicine data set discussed in class, use K-means with* ***the Manhattan distance metric*** *for clustering analysis by setting K=2 and initializing seeds as C1 = A and C2 = C. You should show the steps for the calculations made by the K-means algorithm to get full points. Then, answer the three questions below. Submit your answers as an electronically written file in Word or pdf format, via Canvas.*

1. *How many steps were required for convergence?*
2. *What are the memberships of two clusters after convergence?*
3. *What are the centroids (coordinates) of two clusters after convergence?*

This is the plot of the medicine data set. The stars represent our centers **C1** and **C2**. The first step is to assign all of our data points to the closest cluster center. In order to do this we need to calculate the distance of every point to all centers and then we can assign each point to the closest center. Here is a table of the Manhattan distances (=**sqrt((x1-x2)2 + (y1-y2)2)**) between each point and each center (the closer center has its distance in bold):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **A** | **B** | **C** | **D** |
| **C1** | **0** | **1** | 3.61 | 5 |
| **C2** | 3.61 | 2.83 | **0** | **1.42** |

With these values our clusters would look like:

C1 = [A, B]

C2 = [C, D]

Next we recalculate our centers by using the average values for every dimension from the points in each cluster. So for **C1** our new center will be the average of **A** and **B** and the same for **C2** except with **C** and **D**.

This makes our new plot look like this, again with stars as centers:

Now we reassign every point to its closest center which means we need to recalculate the distance of every point to its center which can be found here:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **A** | **B** | **C** | **D** |
| **C1** | **.5** | **.5** | 3.2 | 4.61 |
| **C2** | 4.3 | 3.54 | **.7** | **.71** |

Now we can reassign each point to its closest cluster:

C1 = [A, B]

C2 = [C, D]

Since the memberships of every cluster remained the same we are actually done.

1. *How many steps were required for convergence?* **2 steps**
2. *What are the memberships of two clusters after convergence?* **C1 = [A,B]; C2 = [C,D]**
3. *What are the centroids (coordinates) of two clusters after convergence?*