**The Hong Kong Polytechnic University**

**Department of Computing**

**COMP5511 – Introduction to Artificial Intelligence**

**Semester 1, 2017-18**

Assignment 2

Due Date: November 15, 2018

1. (50 *marks*) The ABC Telecom has collected more demographic data from its customers. A sample of 15 such data records were selected and are shown below. Each record is characterized by six attributes: (i) Age; (ii) Sex, (iii) Monthly Income; (iv) Marital status; (v) Service Plan (in mins.) and (vi) Extra Usage (in mins.).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **(0)** | **(i)** | **(ii)** | **(iii)** | **(iv)** | **(v)** | **(vi)** |
| **Ref** | **Age** | **Sex** | **Monthly Income** | **Marital**  **Status** | **Service**  **Plan** | **Extra**  **Usage** |
| 1 | 54 c | FEMALE | 3000 1 | YES | 100 A | 0 Ⅰ |
| 2 | 59 c | FEMALE | 4000 1 | NO | 600 B | 54 Ⅰ |
| 3 | 38 b | MALE | 7800 1 | NO | 200 A | 31 Ⅰ |
| 4 | 18 a | FEMALE | 8500 1 | NO | 600 B | 311 Ⅲ |
| 5 | 27 a | MALE | 14000 1 | YES | 100 A | 211 Ⅲ |
| 6 | 29 a | FEMALE | 31000 3 | YES | 1600 C | 25 Ⅰ |
| 7 | 17 a | MALE | 7500 1 | NO | 600 B | 254 Ⅲ |
| 8 | 22 a | FEMALE | 7900 1 | NO | 200 A | 31 Ⅰ |
| 9 | 34 b | MALE | 24700 2 | NO | 100 A | 7 Ⅰ |
| 10 | 46 c | FEMALE | 31110 3 | YES | 600 B | 0 Ⅰ |
| 11 | 39 b | FEMALE | 21000 2 | YES | 800 B | 64 Ⅰ |
| 12 | 35 b | FEMALE | 30000 3 | NO | 1600 C | 0 Ⅰ |
| 13 | 39 b | MALE | 40500 3 | YES | 1600 C | 50 Ⅰ |
| 14 | 18 a | MALE | 7800 1 | NO | 1000 C | 290 Ⅲ |
| 15 | 22 a | MALE | 18000 2 | YES | 400 A | 303 Ⅲ |

* 1. Find a clustering arrangement of records using the *k*-mode algorithm by setting k=3 and using the records 1, 8 and 15 as the initial cluster centers. For those numeric columns, divide the range into 3 intervals (use Equal-width and Equal-depth, i.e. you should have two sets of results). You must show the steps of the first iteration.
  2. Perform clustering by using the *k*-means algorithm but this time you can use Python to help find the results.
  3. Compare all the results you obtained with the different algorithms. Which one would you consider as the best? Why?

1. (50 *marks*) You are given the following graph representing ten individuals that are related to each other in some ways in a social network.
   1. Use Breadth First Search or Depth First Search to find all the shortest paths between the ten different vertices in the graph. Show your work as samples and the results.
   2. Compute the Edge-Betweenness values of the edges and put them in a table in the same way shown in the class notes. Show some samples of your work and the results.
   3. Use the Girvan-Newman algorithm to discover in the graph two communities. Show your work and the results.

\*\*\*\*\* END \*\*\*\***\***