|  |  |  |
| --- | --- | --- |
| CIS 635 | Data Mining | name\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

Homework 8

1. normalized values

|  |  |  |
| --- | --- | --- |
|  | x1 | x2 |
| 101 |  |  |
| 102 |  |  |
| 103 |  |  |
| 104 |  |  |
| 105 |  |  |
| 106 |  |  |

1. manual assignments: fill in the tables for the calculation of centroids and assignment of clusters.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| means | iteration 1 | | iteration 2 | |
| cluster | x1 | x2 | x1 | x2 |
| 1 |  |  |  |  |
| 2 |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | first assignment | | | second assignment | | |
| id | dist. to  cluster 1 | dist. to  cluster 2 | cluster  assign. | dist. to  cluster 1 | dist to  cluster 2 | cluster  assign. |
| 101 |  |  |  |  |  |  |
| 102 |  |  |  |  |  |  |
| 103 |  |  |  |  |  |  |
| 104 |  |  |  |  |  |  |
| 105 |  |  |  |  |  |  |
| 106 |  |  |  |  |  |  |

1. Agglomerative clustering
   1. distance matrix

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | c1 | c2 | c3 | c4 | c5 | c6 |
| c1 |  |  |  |  |  |  |
| c2 |  |  |  |  |  |  |
| c3 |  |  |  |  |  |  |
| c4 |  |  |  |  |  |  |
| c5 |  |  |  |  |  |  |
| c6 |  |  |  |  |  |  |

* 1. answers:
  + what two clusters would be joined next: \_\_\_\_\_ \_\_\_\_\_\_
  + what would be the new distance: \_\_\_\_\_\_\_\_\_\_\_\_

1. R analysis
   1. sizes \_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_
   2. descriptions  
      * cluster 1 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
      * cluster 2 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
      * cluster 3 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
      * cluster 4 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
      * cluster 5 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_