Group K GHC Task

PAM/k-Medoids

(to be presented on Nov. 7, 2023)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **A** | **B** | **C** | **D** | **E** | **F** | **G** |
| **A** | 0 | 1 | 4 | 5 | 6 | 5 | 8 |
| **B** |   | 0 | 2 | 3 | 3 | 5 | 9 |
| **C** |   |   | 0 | 1 | 8 | 9 | 7 |
| **D** |   |   |   | 0 | 7 | 4 | 9 |
| **E** |   |   |   |   | 0 | 3 | 6 |
| **F** |   |   |   |   |   | 0 | 8 |
| **G** |   |   |   |   |   |   | 0 |

Assume you apply PAM/K-Medoids to a dataset O containing 7 objects for k=4. The distance matrix of the objects in O is given above. Moreover, we assume that the initial set of representatives is {A, B, C, D}. What is the SSE of the corresponding clustering for this representative set? Next, run PAM for 1 iteration indicating all computations PAM performs. What new

representative set will be obtained? Will PAM terminate after the first iteration is completed.

Group L GHC Tasks

(to be presented on Nov. 9, 2023)

a) In which steps does the APRIORI algorithm take advantage of the APRIORI property?

b) Assume the APRIORI algorithm identified the following 9 4-item sets that satisfy a user given support threshold: **abcd, acde, acdf, acdg, adfg, bcde, bcdf, bcdg, cdef, cdeg;** what initial candidate 5-itemsets are created by the APRIORI algorithm; which of those survive subset pruning?

c) How does APRIORI algorithm create rules after frequent item sets have been computed.

d) Assume an association rule *if smoke then cancer* has a confidence of 86% and a high lift of 2.1. What does this tell you about the relationship of smoking and cancer?