University of Dhaka

Department of Computer Science and Engineering

CSE-4255:Introduction to Data Mining and Warehousing Lab

Lab Assignment-3

Sihan Tawsik
Roll: 5
Session: 2016-17

Submitted to-

Dr. Chowdhury Farhan Ahmed

Professor
Department of Computer Science and Engineering,
University of Dhaka

This report contains the comparison result of two clustering methods: a) K-means and b) K-medoids.

- Language: python
- Libraries: classification report from sklearn, python-bcubed

Iris Dataset

- **Length:** 149
- **k:** 3

Target Values	Means	Medoid
Silhouette	0.507518648535588	0.510644686899777
BCubed-precision	0.5908573540280857	0.5747668845315904
BCubed-recall	0.6698666666666666	0.7589333333333333

lenses Dataset

- **Length:** 24
- **k:** 2

Target Values	Means	Medoid
Silhouette	0.5208840920245881	0.520884092024588

wine-quality-red Dataset

- **Length:** 1599
- **k: 2**

Target Values	Means	Medoid
Silhouette	0.5711898076042893	0.6075281800654608
BCubed-precision	0.37904922968215166	0.35808004599388493
BCubed-recall	0.7757738282616825	0.939595292902519

Wine quality white Dataset

- **Length:** 4898
- **k:** 2

Target Values	Means	Medoid
Silhouette	0.4298107076081132	0.4964294437426804
BCubed-precision	0.3350885059255205	0.3303216278358505
BCubed-recall	0.5328316247634576	0.8499072810852805

amphibians data set

- **Length:** 189
- **k:** 4

Target Values	Means	Medoid
Silhouette	0.41145744483931546	0.27999336075929

glass data set

- **Length:** 214
- **k:** 2

Target Values	Means	Medoid
Silhouette	0.6212797444547816	0.5211358075701247

mushroom data set

- **Length:** 8124
- **k:** 2

Target Values	Means	Medoid
Silhouette	0.3885068098745934	0.29050699879967073

Conclusion: From this, we can see that the K-medoids method performs better for outliers than the k-means. But the k-medoids method depends heavily on the first random initialization.