

Data Mining

Clustering III - Cluster Evaluation (Part B)

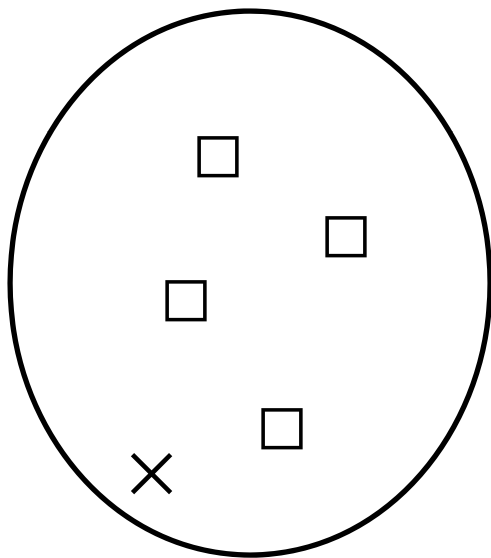
**Dr. Jason T.L. Wang, Professor
Department of Computer Science
New Jersey Institute of Technology**

Where am I?

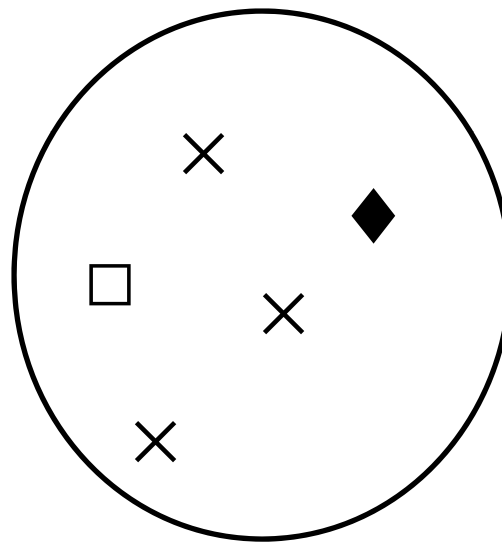
- Part A explains Silhouette coefficient, which is a good evaluation measure for clustering points in Euclidean space.
- Part B explains Purity, which is a simple evaluation measure.

Purity

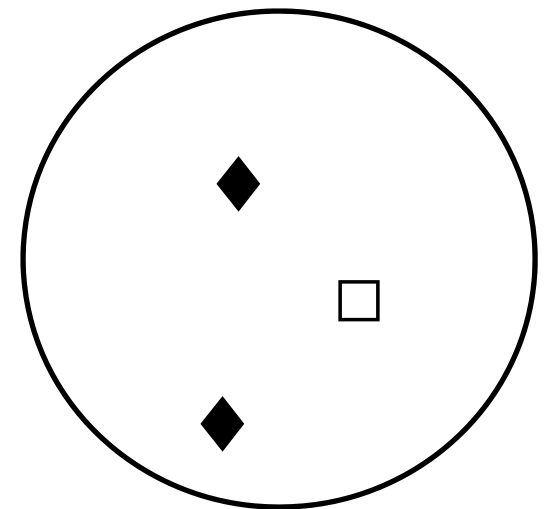
Each cluster is assigned to the class which is most frequent in the cluster: cluster 1 (\square), cluster 2 (\times), cluster 3 (\blacklozenge). Purity is the number of correctly assigned objects divided by the total number of objects, so purity is $(4+3+2)/13 \approx 0.69$.



Cluster 1



Cluster 2



Cluster 3

A bad clustering has a purity of 0.

A perfect clustering has a purity of 1.

End of Cluster Evaluation Module (Part B)