Cluster Analysis II

Learning Outcomes

Upon successful completion of this workshop, you will have demonstrated the abilities to:

- Applying *DBSCAN* clustering different benchmark data sets
- Compare and analyze the results

Instructions:

1. Read the **tutorial** **(http://www.cse.msu.edu/~ptan/dmbook/tutorials/tutorial8/tutorial8.html)**
2. Download the following data sets (the first two columns are feature and the 3rd column is class label):
   - **compound (https://learn.ontariotechu.ca/courses/19275/files/2375096?wrap=1)** (https://learn.ontariotechu.ca/courses/19275/files/2375096/download?download_frd=1)
   - **flame (https://learn.ontariotechu.ca/courses/19275/files/2375094?wrap=1)** (https://learn.ontariotechu.ca/courses/19275/files/2375094/download?download_frd=1)
   - **pathbased (https://learn.ontariotechu.ca/courses/19275/files/2375097?wrap=1)** (https://learn.ontariotechu.ca/courses/19275/files/2375097/download?download_frd=1)
   - **spiral (https://learn.ontariotechu.ca/courses/19275/files/2375289?wrap=1)** (https://learn.ontariotechu.ca/courses/19275/files/2375289/download?download_frd=1)
3. Remove the 3rd column from the data sets
   - Note: In real-world examples, we usually do not have access to actual labels, so we cannot calculate the *Accuracy*

Part I:

1. Perform k-means clustering on different data sets using different **k** (e.g., 2-6)
2. Visualize the clustering results using scatter plots for different **k**
3. Calculate the SSE and plot sum-of-squared errors (SSE) versus ***k***
4. Discuss the results of part (2) and (3)

Part II:

1. Perform DBSCAN clustering on different data sets using different Eps and MinPt
2. Draw the distance to k nearest point versus k (for different k)
   - Try to find the estimation of Eps and MinPt (see DBSCAN: Determining EPS and MinPts slide)

Part III:

1. Compare the results of part I and II and discuss the advantages and limitations

Report:

1. Your report should have a cover letter including the group member names
2. Organize all your *diagrams* and *interpretations* in your lab report *(PDF format)*
3. Include your code and report in a folder (you can zip the folder) and submit it