CPS 844 Lab 9

Part 1: Density-Based Clustering methods (50 points)

Using the data from the file ‘chameleon.data’, we identify high-density clusters separated by regions of low-density. In the popular DBScan method, data points are classified into 3 types:

- Core points
- Border points
- Noise points

Classification is applied as a function of two parameters:

- The radius of the neighborhood size (eps)
- The minimum number of points in the neighborhood (minpts).

Part 2: Anomaly detection (50 points)

Anomaly detection is the task of identifying instances whose characteristics differ significantly from the rest of the data. Although this lab can be completed in only a few lines of code, what’s making it challenging is that you are given only high-level instructions to complete it. Rely on the knowledge you accumulated over the semester, search online, and re-use the code presented during class.

Follow the instructions described in lab9.py. Submit the .py file on D2L. Please note that if you submit your file in some other format besides .py or (.txt should you meet an issue), then your mark will at most be 60%.