Assignment 3

DATA 622 / MDCH 615
Winter 2025

In this assignment, you will generate a simulated data set and perform clustering on it.

Your Python script must meet the following specifications:

- Use ***numpy.random.multivariate_normal*** to generate a random data set consisting of samples from the following 3 bivariate Gaussian distributions:
    - Cluster 1: 1000 samples from ***N***([1, 2], [[0.1, 0.05], [0.05, 0.2]])
    - Cluster 2: 500 samples from ***N***([0, 0], [[0.3, -0.1], [-0.1, 0.2]])
    - Cluster 3: 1500 samples from ***N***([-2, 3], [[1.5, 0], [0, 1.5]])
- Apply DBSCAN to the data set via ***sklearn.cluster.DBSCAN***. Tune *eps* and *min_samples* by trying the following parameter values (see below for performance evaluation):
    - *eps*: from 0.1 to 1, inclusive, with a step of 0.1
    - *min_samples*: from 1 to 10, inclusive, with a step of 1
- Since we know the ground truth, use ***sklearn.metrics.adjusted_rand_score*** to compute the adjusted Rand index to evaluate clustering performance.
- You do NOT need to partition the data; use the entire data set for hyperparameter tuning. Print to the screen the best *eps* and *min_samples* values, as well as the corresponding adjusted Rand index value.
- Make a scatter plot that shows the clustering results from the best *eps* and *min_samples*. Make sure each cluster and outliers are shown in different colours. Show a figure legend.
- Your code should be in a file named ***assignment3_studentid.py***
    - Replace “studentid” in the file name with your student ID
    - Your code should **not** be in a Jupyter notebook

Grading Scheme

- This assignment is worth **5% of your total grade** in the course.
- Submitted Python code must run **error-free**. If your code results in errors and does not execute to completion, a grade of zero will be given. Please test your code before submission.
- If your code executes without error, your code, clustering results, and outputs will be inspected, tested and evaluated. Partial marks will be given if your submission fails to meet some of the specifications.

Deliverable and Deadline

- Submit your ***assignment3_studentid.py*** file to the **Assignment 3 Dropbox on D2L**
- Due at **5:00pm on Thursday, Feb. 13**.