

DA311 Machine Learning Lab

Assignment 8

Date: October 3rd, 2023

Please download the dataset from the following link:

https://github.com/tsharma12/IITG-DA311-Machine-Learning-Lab/tree/main/Week8_Dataset

Part A: Clustering Analysis

Given the two-dimensional data in the dataset file named `DBSCAN_data_with_outliers.csv`:

1. Implement the DBSCAN clustering algorithm on the data. Initially you can consider `eps = 0.1` and `min_points = 4`.
2. Further, vary the `eps` to 0.08, 0.12, 0.15 and 0.16 see how it affects the DBSCAN clustering.
3. Implement the KMeans clustering algorithm (using the Scikit-Learn library) on the same data. Compare the results with those obtained from DBSCAN.

Part B: Decision Tree Analysis

Given the tabular data in the dataset file named `DecisionTree_Fruits_data.csv`:

1. Calculate the information gain for all the attributes: [“Color”, “Taste”, “Size”, “Season”, “Texture”] for the first split in the decision tree. Which attribute provides the highest information gain?
2. Based on the attribute selected in the first step, further split the data and determine the potential leaf nodes. Discuss the criteria for your splits and the process you followed to arrive at the leaf nodes.