Lab Exercises 18

Date: 07 Nov 2025

Implementation of DBSCAN Clustering Algorithm

from sklearn.cluster import DBSCAN

18. Implement DBSCAN clustering on the **Air Quality Dataset** to group environmental readings with similar pollution levels. Identify core points, border points, and noise points. Visualize clusters and interpret the air quality behavior across clusters.

DBSCAN Evaluation and Analysis:

1. **Cluster Summary**: After running DBSCAN, how many clusters were formed? How many data points were marked as noise?
2. **Silhouette Score:** Compute the **Silhouette Score** for the clustering result. What does the score indicate about cluster quality?
3. **Parameter Sensitivity**: Repeat DBSCAN with different eps (e.g., 0.5, 1.0, 1.5). How does the number of clusters and noise points change? Which parameter combination yields the best cluster separation?
4. **Outlier Interpretation**: Examine the noise points identified by DBSCAN. What do these represent — extreme pollution readings, sensor errors, or natural outliers? Give a short justification.