Mini Project 3

Dataset – The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (**WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING**) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers were selected for generating the training data and 30% for the test data.

The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low-frequency components, therefore a filter with a 0.3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain.

Link to the dataset:

https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

Tasks

Task 1 – Your main task is to use K-Means and DBSCAN to do clustering on the given dataset. Your code needs to consider the following aspects, and this also should be reflected in your final report.

- How do you choose the number of clusters in K-Means, is it the same number of clusters for DBSCAN?
- How do you find the optimal parameters' values?
- What data processing steps do you apply and why?

Task 2 – Use a dimensionality reduction technique before using K-Means and DBSCAN on the dataset.

- What is the dimensionality reduction technique that you choose, and why?

- Does it have any effect on your code efficiency, both in terms of computational efficiency and clustering output?
- How do you compare the outcome of this model with the model where the dimensionality reduction technique was not applied to the dataset?

Task 3 – Visualize your clustering.

- Have you applied any dimensionality reduction techniques? Why?

Task 4 – Write a scientific report which includes

- Introduction (what is the problem you are solving?)
- Data processing (what are the choices you made in data processing and how you performed it?)
- Modeling (make sure you have answered all the questions in Tasks 1-3)
- Conclusion (How do you interpret the identified clusters? What do they represent? What were the “scientific” bottlenecks? How did you overcome them?)

You need to hand in your Python code (preferably Jupyter Notebook or Google Colab notebook) alongside a written report.