Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam – 603 110
(An Autonomous Institution, Affiliated to Anna University, Chennai)

UCS2612 Machine Learning Laboratory

Academic Year: 2023-2024 Even **Batch: 2021-2025**
Faculty In-charges: Y.V. Lokeswari & Nilu R Salim **VI Semester A & B**

A. No. : 6 . **K-Means Clustering Algorithm**

Download the Human Activity Recognition Using Smartphones dataset from the link given below:

https://www.kaggle.com/datasets/uciml/human-activity-recognition-with-smartphones?resource=download

Human Activity Recognition database built from the recordings of 30 subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors.

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 was selected for generating the training data and 30% 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 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.

Attribute information

For each record in the dataset the following is provided:

- Triaxial acceleration from the accelerometer (total acceleration) and the estimated body acceleration.
- Triaxial Angular velocity from the gyroscope.
- A 561-feature vector with time and frequency domain variables.
- Its activity label.
- An identifier of the subject who carried out the experiment.

Develop a python program to cluster the human activity using K-means clustering algorithm. Visualize the features from the dataset and interpret the results obtained by the model using Matplotlib library. **[CO1, K3]**

Use the following steps to do implementation:

1. Loading the dataset.
2. Pre-Processing the data (Handling missing values, Encoding, Normalization, Standardization).

3. Exploratory Data Analysis.
4. Feature Engineering techniques.
5. Split the data into training, testing and validation sets.
6. Cluster the data using Euclidean distance metric.
7. Measure the performance of the model.
8. Represent the clustered data using graphs.

Upload the code in GitHub and include the GitHub main branch link in the assignment PDF.

Hints to do the assignment:

Do the following:

1. Load the dataset.
2. Pre-Processing the data (Handling missing values, Encoding, Normalization, and Standardization).
3. Exploratory Data Analysis
4. Feature Engineering techniques.
   Refer to
   https://machinelearningmastery.com/feature-selection-machine-learning-python/
   https://www.analyticsvidhya.com/blog/2020/10/feature-selection-techniques-in-machine-learning/
   https://www.datacamp.com/tutorial/feature-selection-python

5. Apply K-Means Clustering algorithm on the input dataset and group them into clusters.
   https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html
   https://www.datacamp.com/tutorial/k-means-clustering-python

6. Implement K-Means Clustering algorithm with user-defined functions and Euclidean distance metric.
   https://medium.com/machine-learning-algorithms-from-scratch/k-means-clustering-from-scratch-in-python-1675d38eee42

7. Upload python project in GitHub and explore all git commands. Git Commands Tutorial : https://git-scm.com/docs/gittutorial

   Upload IPython to GitHub
   https://reproducible-science-curriculum.github.io/sharing-RR-Jupyter/01-sharing-github/

   Additional Reference:
   https://www.youtube.com/watch?v=LlrKTV4-ftI

Upload the code in GitHub and include the GitHub main branch link in the assignment PDF.

Upload python project in GitHub and explore all git commands.

Git Commands Tutorial : https://git-scm.com/docs/gittutorial