PHASE 2

Designed by: Srikar Kodavati

Instructor: Dr. Wei Ding

The Weightage for term project is as follows:
phase 1 would be 20%
Phase 2 would be 40%
Phase 3 would be 40%, it will include the in class presentation.

Data description:

Naval Air Training and Operating Procedures Standardization (NATOPS) Dataset.

- 6 of 24 body-hand gestures used when handling aircraft on the deck of an aircraft carrier
- Six classes of actions:

I have command; All clear; Not clear; Spread wings; Fold wings; and Lock wings.

The data generated in phase 1 has table (time steps * features) with multiple rows corresponding to the same sample id (sid). Use the given file NATOPS_sid_TRAIN.csv provided as the phase 1 output data. We find the atomic units from the data, which are some basic units shown in the data repeated, and we are trying to find them by applying an unsupervised clustering algorithm. It could be k-Mean or any other clustering method.

NOTE: use .ipynb notebook to write your code and submission.

Tasks:

1. Find clusters:

Find the clusters for each row of the data using unsupervised learning methods.

2. Generate atomic units data:

Then, for each sample(sid), find the ratio of each cluster.

For example, we have a sample A, it has 10 time steps, after applying a clustering method, each step will be assigned to a cluster (assume we are trying to find 3 clusters here), then calculate the ratio of each cluster among these 10 time steps. These ratios will be the new ratio features for the next phase (classification).

The final data considers each cluster as features. If data is grouped into 10 clusters the number of features are 10 with the ratio of the corresponding cluster as the value. The columns in the final data are test/train, sid, class attribute and the features. For each timestep in phase one data there is a row which only has above given columns.

3. The final step

   The data generated should be displayed as two outputs:

- Save the table as a csv file.
- Print the first 5 rows with column names in the program.

  Grading Criteria:

Full grade(40%): The final data is generated as mentioned above.

Half grade(20%): create a unsupervised learning algorithm for the above task.

Submission:

Submit the notebook as .ipynb file.Upload this file on blackboard in the Phase2 section.