Pattern Recognition / NCTU Spring 2019 Program Assignment #3 (due: 6/11/2019)

The goal of this program assignment is to let you experiment with clustering algorithms. You can use Matlab, python, or C++. You are not allowed to use existing libraries/modules/toolboxes. You can continue to use the datasets mentioned in the previous assignments. (The instructor will continue to upload mainly 2-D simulated datasets for you to play with, so that you can visualize your clustering results.)

For the clustering algorithms to implement, choose one of the two tracks below. Things that you can do are listed under each track. Most of the experiments are optional; you don't need to do them all, but you should do some comprehensively. You can also think of experiments that are not listed here.

Track 1: Hierarchical Agglomeration (single-link and complete-link) and DBSCAN.

- You need to convert the datasets to relational data using some similarity or dissimilarity measures.
- Experiment with different options of computing similarity or dissimilarity measures.
- Experiment with different parameter settings, such as radius and density threshold in DBSCAN.
- Try to see if you can determine stable clusterings from the dendograms of hierarchical agglomeration.
- Add different levels of random noise to the data and see how the clustering results are affected.

Track 2: The c-means family: HCM, FCM, and PCM

- Experiment with different parameter settings, such as fuzzification factors in FCM and HCM.
- These all give results that are dependent on initialization. Try to devise measures for the variability of the clustering results (example: pairwise ARI).
- Add different levels of random noise to the data and see how the clustering results are affected.
- Try to see if you can determine the number of clusters using some validity measure.
- Add different levels of random noise to the data and see how the clustering results are affected.

For either track, try to evaluate your results by comparing them to the actual class labels. Use at least one objective evaluation method (Adjusted Rand Index, Normalized Mutual Information, or Hungarian Algorithm); for this part you are allowed to use existing source codes or libraries that you can find.

Write a report (limited to 10 pages, single-spaced) describing

- Methods you have implemented.
- Experiments you have done, and the results.
- Analysis - Are the results what you expect? Why?

The report (Word or PDF formats) is to be submitted electronically through e3. Also include the program listing as an appendix to your report (in the same file, not counted toward the page limit).

Late submission policy: 10% credit deduction for each day late; up to 7 days late accepted.