Dear Researcher,

Thank you for using this code and datasets. I explain how GEPFCM code related to my paper "Generalized entropy based possibilistic fuzzy C-Means for clustering noisy data and its convergence proof" published in Neurocomputing, works. The main datasets mentioned in the paper together with GEPFCM code are included.

If there is any question, feel free to contact me at:

bas\_salaraskari@yahoo.com

s\_askari@aut.ac.ir

Regards,

S. Askari

Guidelines for GEPFCM algorithm:

1. Open the file GEPFCM Code using MATLAB. This is relaxed form of the algorithm to handle noisy data.
2. Enter or paste name of the dataset you wish to cluster in line 15 after "load". It loads the dataset in the workplace.
3. For details of the parameters cFCM, cPCM, c1E, c2E, eta, and m, please read the paper.
4. Lines 17 and 18: "N" is number of data vectors and "D" is number of independent variables.
5. Line 26: "C" is number of clusters. To input your own desired value for number of clusters, "uncomment" this line and then enter the value. Since the datasets provided here, include "C", this line is "comment".
6. Line 28: "ruopt" is optimal value of ρ discussed in equation 13 of the paper. To enter your own value of ρ, "uncomment" this line. Since the datasets provided here, include "ruopt ", this line is "comment".
7. If line 50 is "comment", covariance norm (Mahalanobis distance) is use and if it is "uncomment", identity norm (Euclidean distance) is used.
8. When you run the algorithm, first FCM is applied to the data. Cluster centers calculated by FCM initialize PFCM. Then PFCM is applied to the data and cluster centers computed by PFCM initialize GEPFCM. Finally, GEPFCM is applied to the data.
9. For two-dimensional plot, "uncomment" lines 419-421 and "comment" lines 423-425. For three-dimensional plot, "comment" lines 419-421 and "uncomment" lines 423-425.
10. To run the algorithm, press Ctrl+Enter on your keyboard.
11. For your own dataset, please arrange the data as the datasets given here.

It is explained by an example. Consider the following dataset with 3 variables and 10 data vectors. Your data should be inputted as a 3×10 matrix named "yd" as follows:

