TÉCNICO LISBOA

Intelligent Systems

Master Degree in Mechanical Engineering

Assignment 1 [EN]

Authors:

Henrique Branquinho (ISTID 087203) henrique.branquinho@tecnico.ulisboa.pt
Diogo Dias (ISTID 096705) diogohbdias@tecnico.ulisboa.pt

Group 7

2024/2025 – 1st Semester, P1

Contents

1 Exercise 1

This exercise consisted in modelling a single input single output (SISO) fuzzy rule-based system applied to the Hair Dryer dataset. This set can be visualized in the image bellow (figure 1). Prior to the implementation of the model, we preformed an analysis to make sure there were no missing values and we were ready to proceed to the data processing.

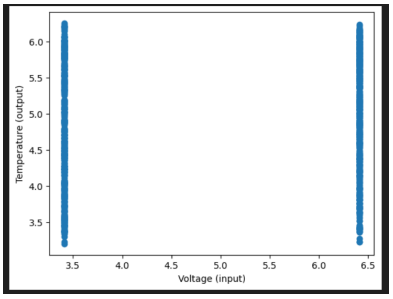

Figure 1: Hair Dryer Dataset

1.1 Creating the model

To create our model we experimented to different methods. Firstly it was implemented was FCM (Fuzzy C-means clustering method), in which data points can belong to various clusters, with different degrees of membership. Then we implemented the GK (Gustafson-Kessel clustering method), where we have clusters of varying shape.

After choosing the previous best method, we also trained models using local and global fit, and measured their performances.

As an evaluation metric, we used the Mean Squared Error (MSE).

1.1.1 Results

To study the models performances we plotted the following graphs:

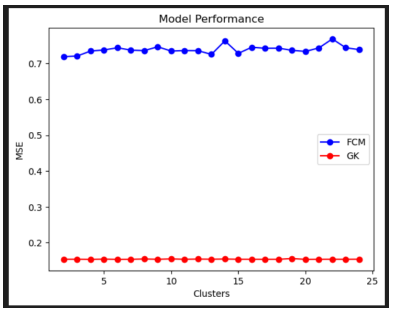

Figure 2: GK vs FCM

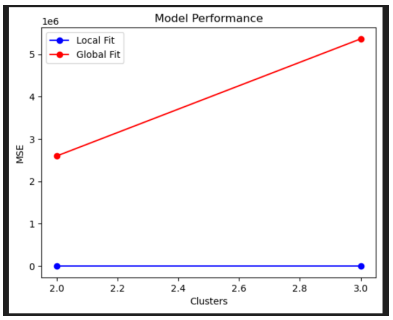

Figure 3: Local vs Global Fit

As it can be observed from the figure 2 the GK method had a better performance, no matter the number of clusters. Using this same method the local optimization seemed to produce much better results than the global fit (figure 2). This was the expected behavior since the data is separated into two very defined regions with strong local tendencies.

In the end, considering the performance and simplicity of the model, we chose the model with the Gustafson-Kessel clustering method, using 2 Clusters, and local fit.

The results obtained were the following:

Metric	Best Result	Cluster Count
MSE	0.15	2

2 Exercise 2

2.1 Creating Model

In this part of the assignment, we are dealing with a MISO (multiple input, single output) dataset. Similarly, we start by analysing if our dataset is complete (no missing values in the cells), then removed all the rows contain non-numeric values.

Once the preprocessing of the data was done, we studied different fuzzy models, exactly as we did for the previous dataset.

For this models we used the F1-score as a metric, since it works nicely for unbalanced datasets.

2.1.1 Results

After testing all the models created we obtained these plots:

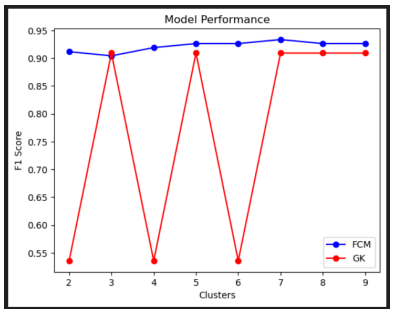

Figure 4: GK vs FCM

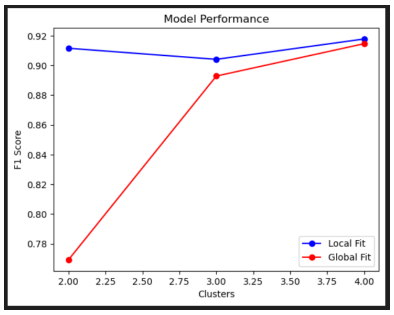

Figure 5: Local vs Global Fit

Since breast cancer is a very complex that take many factors into consideration we cannot define rigid groups to subdivide our data. Therefore, as observed in figure 4, the FCM method worked much better, despite the numbers of clusters used.

For the Fuzzy C-means clustering method we discovered that, given enough clusters, both the local and global fit achieved incredible results.

The model chosen as final ended up being the one that used 4 FCM clusters, and trained with the local fit (figure 5).

From that model we obtained the following results:

Metric	Best Result	Cluster Count
F1 Score	0.91	4

3 GitHub Link

You can find the code on GitHub at:
https://github.com/diogohbd/SInt/tree/master/Assignments/Assignment%201