Intelligent Systems Assignment 1

Group nr: 2	Student 1: Álvaro Lopes	IST nr: 96148
	Student 2: André Lopes	IST nr: 96351

Exercises

1. What is the difference between a characteristic function and a membership function?

A characteristic function is associated with classical set theory, which determines in a binary way wether or not an element belongs to a set. The membership function is a concept in fuzzy logic which dictates the degree of membership of said element in a fuzzy set. While the characteristic function has values of 0 or 1, the membership function can take any value between 0 and 1. The higher the value, the higher is the degree of membership to the fuzzy set. This allows fuzzy logic to model uncertainty, unlike classical set theory.

2. Consider the two fuzzy sets in the Universe of Discourse X ={8, 6, 4, 2, 0, 2, 4, 6, 8}:
$\mu_A(x) = \frac{1}{1+|x|}$ **and** $\mu_B(x) = 1 - \frac{|x|}{20}$

a) Are the membership functions valid in the given Universe?

Yes, because the membership functions result in values between 0 and 1 regardless of the element considered from the universe of discourse X.

b) Compute the α-cuts of A and B for α = 0.3

The results from computing the alpha-cuts on each of the membership functions are below and are illustrated in figures 1 and 17.

$$\mu_{A|\alpha=0.3} = \{-2, 0, 2\} \quad (1)$$

$$\mu_{B|\alpha=0.3} = \{-8, -6, -4, -2, 0, 2, 4, 6, 8\} \quad (2)$$

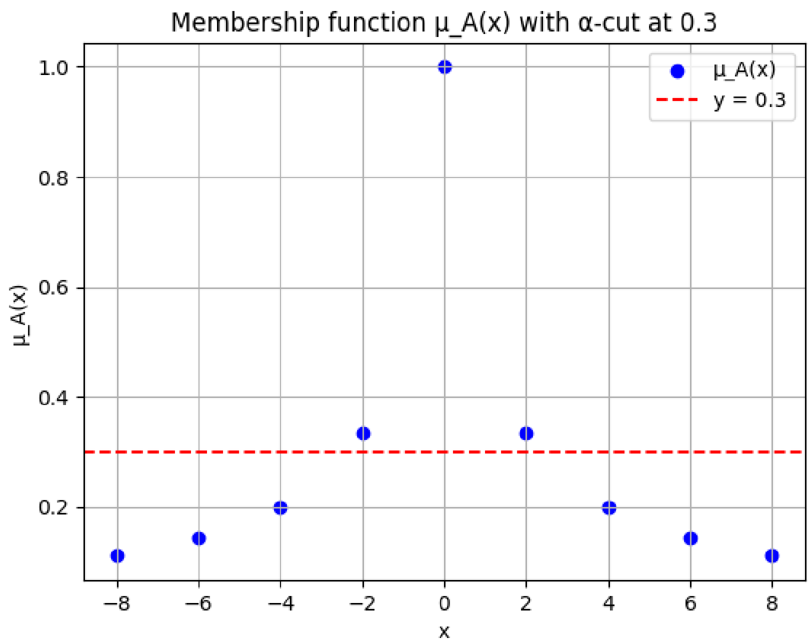

Figure 1: Membership function A - α-cut at 0.3

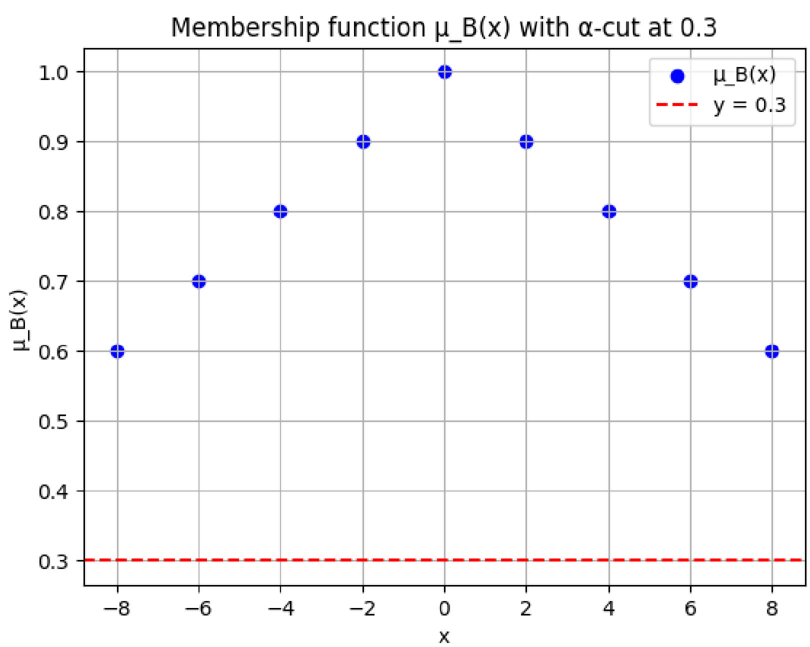

Figure 2: Membership function B - α-cut at 0.3

Wine Classification using Fuzzy Modeling

It was proposed to develop a Takagi-Sugeno fuzzy model to classify the provided dataset, which consists of results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. Each sample has measured 13 associated features: 1) Alcohol; 2) Malic acid; 3) Ash; 4) Alcalinity of ash; 5) Magnesium; 6) Total phenols; 7) Flavanoids; 8) Nonflavanoid phenols; 9) Proanthocyanins; 10) Color intensity; 11) Hue; 12) OD280/OD315 of diluted wines; 13) Proline.

If it all goes well, by building a model and training it with a portion of the dataset, it will be possible to predict what type of wine a given sample is, by feeding the associated attributes of the sample.

Pre-processing

The data was imported to a python script and the first thing to do was to check if the data had missing values, duplicate rows and duplicate columns. This is relatively common when dealing with a very large dataset. Ours only had 178 entries though, and this problem was not present.

Two plots were made in order for us to have a better understanding of the data we were dealing with. One of them is a bar plot, to see how many samples of each type the data has. The other one is a box plot, which represents in a graphical way the distribution of the dataset's features. Both of them can be seen bellow.

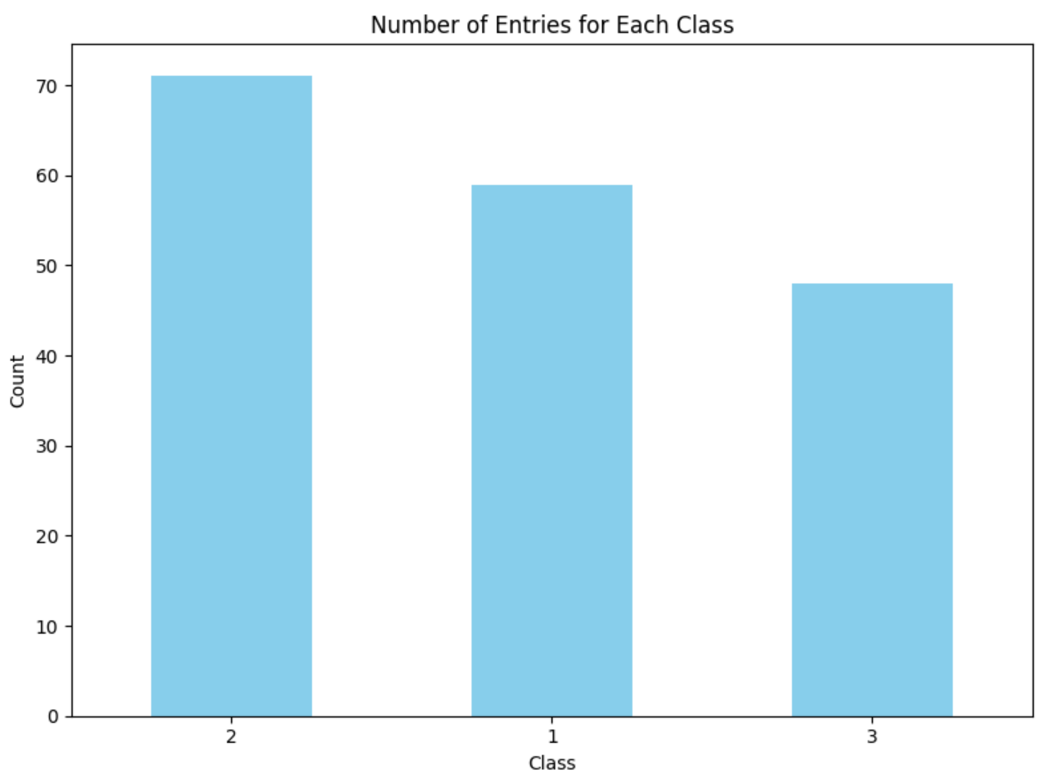

Figure 3: Class occurrence bar plot

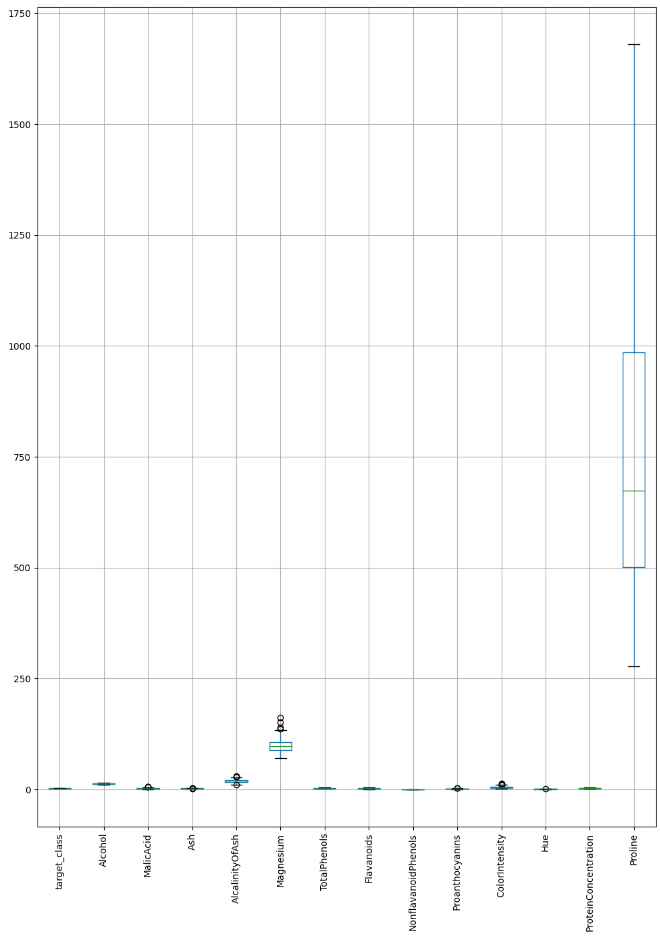

Figure 4: Features data distribution

The dataset was partitioned into training and test sets to conduct an empirical study. Various partitioning ratios were explored during the experimentation phase. However, it was observed that due to the relatively small size of the dataset and its inherent simplicity, a conventional training size of 80% yielded perfect results, demonstrating excellent model performance.

To investigate the impact of the number of clusters on model accuracy, a deliberate decision was made to employ a reduced training size of 20%. Subsequently, an 80% portion of the dataset was designated for testing purposes. This partitioning strategy was chosen to introduce a controlled level of complexity and variability into the study, allowing for a comprehensive examination of how varying cluster numbers influence model accuracy.

Results

A type 1 Takagi-Sugeno fuzzy model was used in order to classify the data. As for fuzzy clustering, both Fuzzy c-means and Gustafson-Kessel FCM were applied, but FCM gave better results, so we decided to analyze those instead.

As already mentioned, by using a 80% training set size, the obtained results were perfect. In the next page we can see that the number of clusters in this case is completely irrelevant to the results.

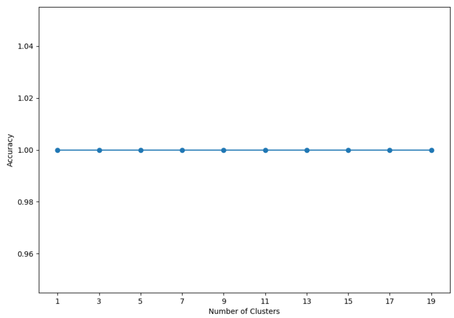

Figure 5: Accuracy score for different number of clusters

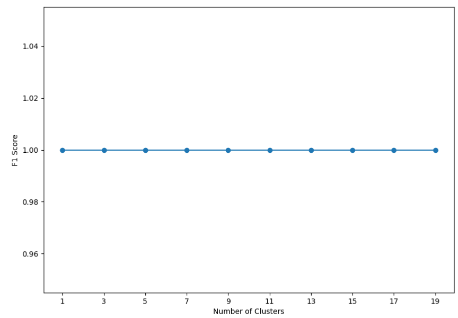

Figure 6: F1 score for different number of clusters

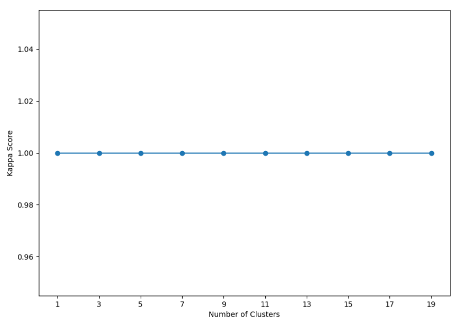

Figure 7: Kappa score for different number of clusters

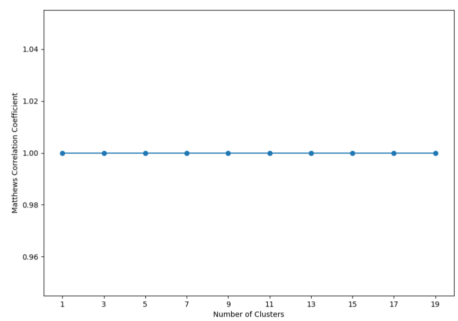

Figure 8: Matthews score for different number of clusters

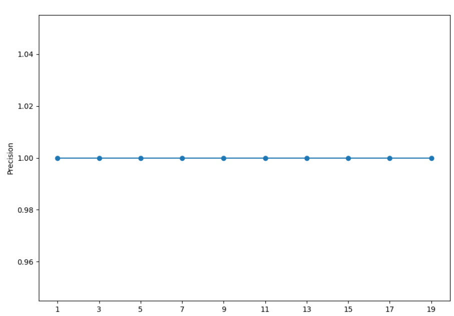

Figure 9: Precision score for different number of clusters

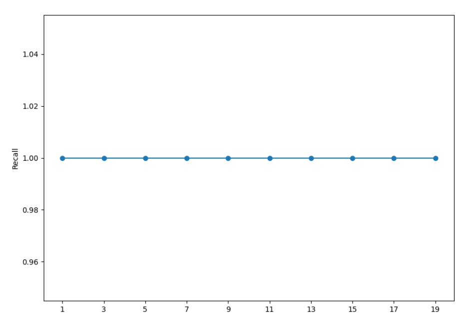

Figure 10: Recall score for different number of clusters

After these excellent results, a training set with just 20% of the data was used, as we had hope that we could see a convergence in these scores with the increase of the number of clusters. The results can be seen in the next page.

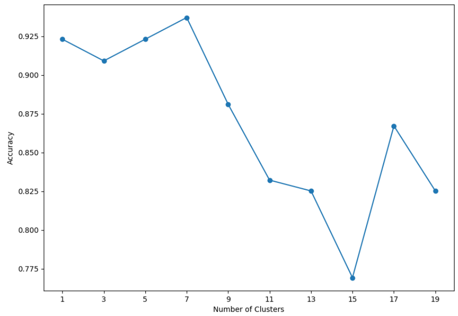

Figure 11: Accuracy score for different number of clusters

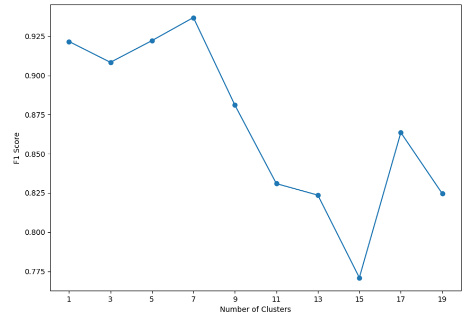

Figure 12: F1 score for different number of clusters

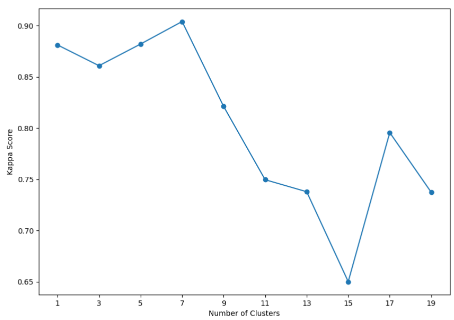

Figure 13: Kappa score for different number of clusters

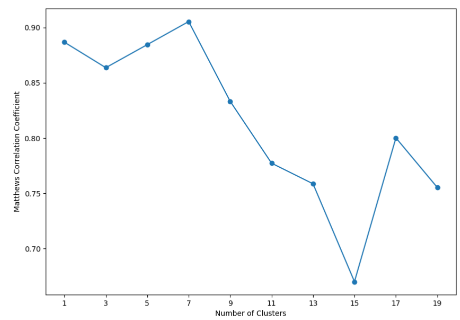

Figure 14: Matthews score for different number of clusters

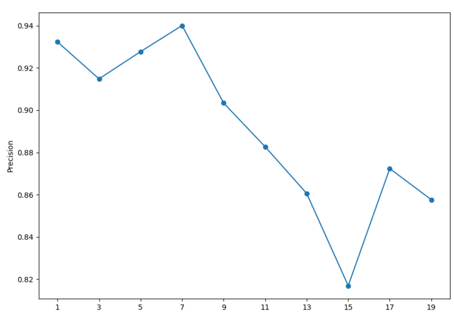

Figure 15: Precision score for different number of clusters

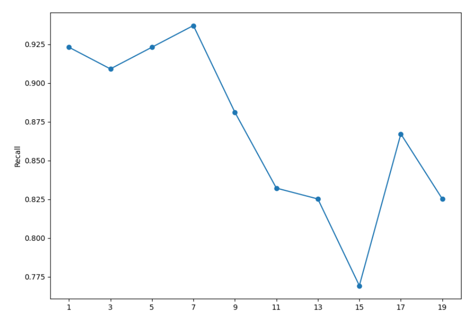

Figure 16: Recall score for different number of clusters

As it is possible to observe, until 7 clusters, there is a tendency to slightly improve the model, but from there, it generally gets worse (with some fluctuations). With the increase in the number of clusters, the model becomes more complex and can capture noise and random fluctuations instead of meaningful patterns. Therefore, there is a loss of generalization. Our dataset is quite simple, so it does not need a complex model for the predictions to perform well. We had 13 features for each sample, but for example, thinking in an example from real life, if we go to the supermarket and buy 3 different wines, they all can have different alcohol percentages, and it can be possible to distinguish them just by their alcohol content. If this feature is enough for distinguish them, all the others can be inducing error or noise.

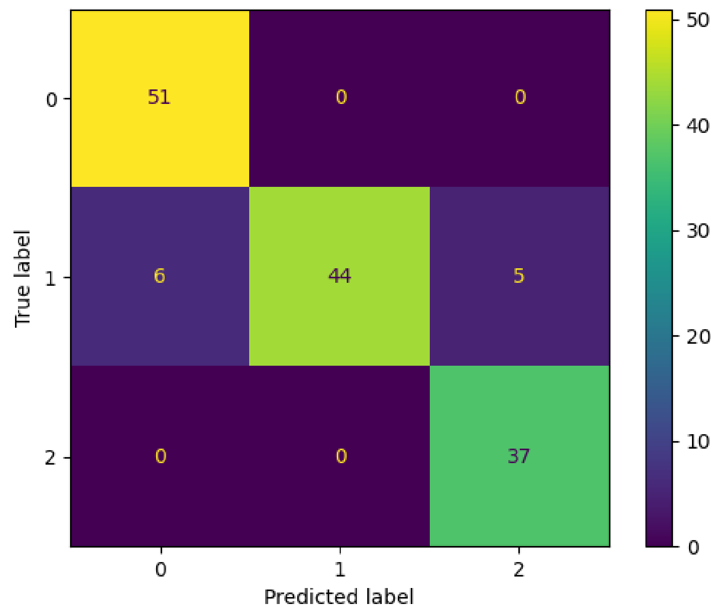

Figure 17: Confusion matrix for 7 clusters

Conclusion

With this assignment it was possible to apply the concepts learned in class to a practical situation. The dataset that we worked on was quite simple, so the model constructed with a typical train size of 80% gave perfect results. We made the model produce worse results on purpose, by giving it a train size of only 20%, to check if there was a convergence of accuracy by increasing the number of clusters. This did not happen, and that can be attributed to overfitting of the model.

Link to group GitHub: `https://github.com/andrelopes2001/Intelligent_Systems`