Yuzuncu Yil University
Journal of the Institute of Natural & Applied Sciences
https://dergipark.org.tr/en/pub/yyufbed

Research Article

A Novel Membership Function Definition for Fuzzy Classification

Nur UYLAŞ SATI

Muğla Sıtkı Koçman University, Bodrum Vocational High School, Department of Motor Vehicles and Transportation Technologies, 48100, Muğla, Türkiye
Nur UYLAŞ SATI, ORCID No: 0000-0003-1553-9466
Corresponding author e-mail: nursati@mu.edu.tr

Article Info

Received: 20.01.2023
Accepted: 16.03.2023
Online August 2023

DOI:10.53433/yyufbed.1239769

Keywords
Data mining,
Fuzzy classification,
Mathematical optimization,
Membership functions,
Polyhedral conic functions

Abstract: In this paper, a novel membership function is defined for fuzzy sets using a supervised learning approach. Firstly, the training dataset is separated using the previously defined polyhedral conic functions in a supervised learning approach. Then obtained polyhedral conic functions are used for defining a new membership function. After that, a new fuzzy classification algorithm is formed to classify fuzzy sets with a similar structure. The algorithm with all suggested methods is implemented on real-world datasets, and the performance values are compared with the state of art classification algorithms.

Bulanık Sınıflandırma için Yeni Bir Üyelik Fonksiyon Tanımlaması

Makale Bilgileri

Geliş: 20.01.2023
Kabul: 16.03.2023
Online Ağustos 2023

DOI:10.53433/yyufbed.1239769

Anahtar Kelimeler
Bulanık sınıflandırma,
Çok yüzlü konik fonksiyonlar,
Matematiksel optimizasyon,
Üyelik fonksiyonu,
Veri madenciliği

Öz: Bu çalışmada gözetimli öğrenme yaklaşımı kullanılarak bulanık kümeler için yeni bir üyelik fonksiyonu tanımlanmıştır. İlk olarak, gözetimli öğrenme yaklaşımında, eğitim veri kümesi, önceden tanımlanmış çokyüzlü konik fonksiyonlarla ayrılmış ve daha sonra elde edilen bu çokyüzlü konik fonksiyonlar yeni bir üyelik fonksiyonu tanımlamak için kullanılmıştır. Sonrasında ise bu fonksiyon kullanılarak benzer yapıda bulanık kümeleri sınıflandırmak için yeni bir bulanık sınıflandırma algoritması tanımlanmıştır. Önerilen tüm yöntemler bir algoritmada birleştirilerek, veri kümeleri üzerinde denenmiş ve performans değerleri, literatürde yer alan sınıflandırma algoritmalarıyla karşılaştırılmıştır.

1. Introduction

Machine learning is an important research area of artificial intelligence, which uses the information gained from data and algorithms to learn as humans learn. Various methods are used in machine learning, such as learning associations, classification, regression, unsupervised learning (clustering) and reinforcement learning. Fuzzy classification is one of these machine learning techniques based on fuzzy logic and aims to classify fuzzy sets by a defined machine-learning technique (Alpaydın, 2010).

Defining membership functions or fuzzy rules is the most critical work in constructing a fuzzy classification system. Fuzzy membership functions and fuzzy rules can be formulated based on expert knowledge or data-driven approaches. These approaches utilize a machine learning process based on training instances. The expert knowledge approach can be non-confidential since the experts can define so different rules or membership functions for the same case. From this point of view, the automatic construction of membership functions is preferable. Various methods based on machine learning and statistics are proposed in the literature to generate membership functions (Mendes et al., 2001; Sanz et al., 2010; Makrehchi & Kamel, 2011; Borkar & Rajeswari, 2013; Jamsandekar & Mudholkar, 2014; Bhattacharyya & Mukherjee, 2020; Rapheal & Bhattacharya, 2020; Satı, 2020; Xie et al., 2021; Azam et al., 2021).

Xie et al. (2021) proposed a novel polynomial membership function approach for polynomial fuzzy system stability analysis. Conservatism is reduced in this new approach, and comparisons present the effect of this reduction with other methods.

An approach for generating fuzzy Gaussian and triangular membership function using fuzzy c-means was considered by Azam et al. (2021). The problem related to sunspot prediction was considered and its accuracy was calculated. It is evident from the results that the proposed technique of generating membership functions using fuzzy c-means can be adopted for predicting sunspots.

In the paper of Bhattacharyya & Mukherjee (2020), 33 different membership function evaluation methodologies published between 1971 and 2016 were discussed. Three different case studies were done to check the applicability and tractability of the method.

Rapheal & Bhattacharya (2020) predicted stock market prices by defining a new membership function that uses the minor error of prediction in a fuzzified repeated incremental pruning to produce error reduction (RIPPER) hybrid model. They used rule based system (RBS) and RIPPER in a hybrid model. Gaussian, triangle and trapezoidal memberships of the fuzzy rule based system (FRBS) were regarded. The performance analysis considered mean absolute percentage error (MAPE) for all the membership functions.

In the paper of Satı (2020), by using the labeled data's means and medoids novel membership functions were defined for labeling unsupervised ones and also defined membership values were used in the classification phase in the linear programming problem. The suggested algorithm was tested on real-world datasets and compared with the state-of-the-art semi-supervised methods. The results indicated that the suggested algorithm is effective in classification and worth studying.

This paper presents a new method for constructing membership functions using polyhedral conic functions in optimization-based programming techniques. The main novelty of this work lies in defining a new membership function using polyhedral conic functions (PCFs) and mathematical techniques. In addition to making an essential contribution to the scientific literature, the presented method is also useful for many fields, such as control systems engineering, image processing, power engineering, industrial automation, robotics, education, consumer electronics, and optimization, where fuzzy logic is used (Singh et al., 2013)

In Section 2, the basic concepts of fuzzy sets and the techniques that will be used in fuzzy classification are explained in detail. The proposed method is presented by related instructions and pseudocode of the suggested algorithm in Section 3. In Section 4, the experimental results are presented and discussed. In Section 5, the paper is finalized and future works are suggested.

2. Material and Methods

2.1. Datasets

In this paper, two different typed datasets are used. Firstly in the supervised training process, a binary dataset is used. Binary datasets only have two (usable) values: 0 or 1. These labels can be used in health data for disease diagnosis, as 0 for healthy and 1 for unhealthy patients. The used binary dataset is given below with mathematical symbols.

$$A=\left\{(a^i,0)\middle|a^i \in R^n\right\}, i=1,....,m, B=\left\{(b^j,1)\middle|b^j \in R^n\right\}, j=1,..., p. \quad (1)$$

In the study's second phase, fuzzy sets will be used in implementations and experiments. A fuzzy set consists of objects with different grades of membership. Membership functions determine the grades (between 0 and 1). The theory of fuzzy sets is used in various research areas. For example, in health studies, it is sometimes hard to say whether a patient is strictly sick. It may be more appropriate to give the rate of getting the disease a certain percentage. The used fuzzy set is given below with mathematical symbols.

$$C=\left\{(c^l,\mu_c(c^l))\middle|c^l\in R^n\right\},\ 0\le\mu_c(c^l)\le 1,\ l=1,...,r \tag{2}$$

Here "*a*", "*b*" and "*c*" are *n- dimensional* vectors (point or data) in which every dimension represents an attribute of the data. "*m*", "*p*" and "*r*" are the number of the points in the datasets respectively in *A*, *B* and *C*.

2.2. Polyhedral conic functions (PCFs)

In this paper, PCFs approach is utilized in the training phase to find a function that separates two distinct datasets.

PCFs were defined in 2006 by Gasimov & Ozturk to distinguish two different labeled point sets (Gasimov & Ozturk, 2006).

Polyhedral functions are identified as a function in the paper of Gasimov & Ozturk (2006):

$$g_{(w,\xi,\gamma,a)}:IR^n\to IR=w'(x-a)+\xi\|x-a\|_1-\gamma \tag{3}$$

where x is an n-dimensional data, $x,w,a\in IR^n,\xi,\gamma\in IR, w'x=w_1x_1+...+w_nx_n, \|x\|_1=|x_1|+...+|x_n|$.

Definition 1 and Lemma 1 quoted below were given and proofed in the paper of Gasimov & Ozturk (2006).

Lemma 1: A graph of the function $g_{(w,\xi,\gamma,a)}$ defined in (1) is a polyhedral cone with a vertex at $(a,-\gamma)\in IR^n\times IR$. This cone is called a polyhedral conic set and "*a*" its center.

From Lemma 1, every polyhedral function given in (3) acts as a polyhedral conic function (PCF).

Definition 1: A function $g:IR^n\times IR$ is called polyhedral conic if its graph is a cone and all its level sets $S_\alpha=\left\{x\in IR^n:g(x)\le\alpha\right\},\alpha\in IR$ are polyhedrons.

2.3. MinMax normalization technique

In this paper, for membership function construction besides PCFs, the min-max normalization technique is used. In the Min-max algorithm, a linear transformation is done on the original data (*O*). xmin and xmax are defined as the minimum and the maximum of a variable in the samples. The Min-max algorithm uses the following formula to map a value, *y*, to a value, *y*′. (Cao et al., 2016):

$$y'=\frac{y-\min_o}{\max_o-\min_o}(new\_\max_o-new\_\min_o)+new\_\min_o \tag{4}$$

3. A Novel Algorithm for Fuzzy Classification

A novel algorithm is defined in this paper for fuzzy classification. There are two main procedures in the algorithm: training and finding the separative PCFs with the binary labeled data and constructing the membership function by the obtained PCFs for fuzzy classifying the fuzzy data. The first process is executed by a previously defined PCFs approach for separation. In the second process, a novel membership is defined using the min-max normalization technique given in Section 2.3.

Let A, B and C be given sets containing $m \in Z^+$, $p \in Z^+$, and $r \in Z^+$ n-dimensional points, respectively:
$A=\left\{a^i \in R^n, i \in I\right\}$, $B=\left\{b^j \in R^n, j \in J\right\}$ and $C=\left\{c^t \in R^n, t \in T\right\}$ where $I=\{1,...,m\}$, $J=\{1,...,p\}$ and $T=\{1,...,t\}$.

Algorithm 1: An Algorithm for Fuzzy Classification

Step 0. Let l=1, I_l=I, A_l=A and go to step 1.
Step 1. Let a^l be an arbitrary point of A. Solve subproblem (P_l) .

$$(P_l) \quad \min\left(\frac{y' e_m}{m}\right) , \tag{5}$$

$$w'(a^i - a^l) + \xi \left\| a^i - a^l \right\|_1 - \gamma + 1 \le y_i, \quad \forall i \in I_l, \tag{6}$$

$$-w'(b^j - a^l) - \xi \left\| b^j - a^l \right\|_1 + \gamma + 1 \le 0, \quad \forall j \in J, \tag{7}$$

$$y = (y_1, ..., y_m) \in R^m_+, w \in R^n, \xi \in R, \gamma \ge 1 \tag{8}$$

Let $w^l, \xi^l, \gamma^l, y^l$ be a solution of (P_l). Let

$$g_l(x) = g_{(w^l, \xi^l, \gamma^l, a^l)}(x) \tag{9}$$

Step 2. $I_{l+1} = \left\{i \in I_l : g_l(a^i) + 1 > 0\right\}, A_{l+1} = \left\{a^i \in A_l : i \in I_{l+1}\right\}, l = l+1$. If $A_l \neq \varnothing$ go to Step 1.
Step 3. Define the function $g(x)$ (dividing the sets A and B) as

$$g(x) = \min_l g_l(x) . \tag{10}$$

Step 4. Define bound as the threshold value:

$$\text{bound} = \max_j g(x_j), \; j = 1,.., p . \tag{11}$$

Step 5. Define the membership function for C set:

$$\mu_C(x) = \begin{cases} \dfrac{g(x)}{bound}, & \text{bound} > g(x) > 0, \\ 0, & g(x) \le 0, \\ 1, & g(x) \ge bound. \end{cases} \tag{12}$$

Step 6. Define the class of the point (x):

$$C(x) = \begin{cases} 1, & \mu_c(x) > 0.5 \\ 0, & \mu_c(x) \le 0.5. \end{cases} \tag{13}$$

Stop.

In the defined algorithm, the novelty takes part in Steps 4, 5, and 6. The before steps are previously defined PCFs classification algorithm. This algorithm was modified in the papers of Uylaş Satı (2015) and Ozturk & Ciftci (2015) to decrease the running time and avoid the over-fitting by allowing the miss-classifications. However, in this approach, the first defined algorithm in the paper of Gasimov & Ozturk (2006), which guarantees a separation with 100% accuracy, is used since the points with $g(x) \leq 0$ are labeled with a strict 1.

In step 4, bound is defined as a threshold of the 0 labeled points such that the points $g(x) \geq bound$ are labeled with a strict 0. (see Equation (11))

In step 5, the essential part of the fuzzy logic, the membership function, is defined using obtained PCFs and the bound value.

For the fuzzy points(x) which ensures "$\text{bound} > g(x) > 0$" the membership value is defined by using the min-max normalization technique. In Equation (4) mentioned in Section 2.3, minA is defined as 0 and maxA is defined as bound, new_minA is defined as 0 and new_maxA is defined as 1. After identifying these parameters, Equation (4) is defined as "$\frac{g(x)}{bound}$" in the membership function $\mu_C(x)$. If the membership is nearly 1, it can be said that the point is so near to being labeled 1. For the points(x) ensures "$g(x) \leq 0$" and "$g(x) \geq bound$" is labeled respectively with a strict 1 and 0 (see Equation (12)).

Finally, in step 6, a simple defuzzification is defined for getting crisp values (class labels) to use in performance analysis. In this defined algorithm, since binary classification has been experimented with and the used datasets are binary, just a line is used for separation (see Figure 2). If the membership value ($\mu_C(x)$) of a point is nearly 1, then it can be told that the point is so near to be labeled as 1, from this point of view, the equation defined in (9) is used for defuzzification. Here the 0.5 value of equation (13) can be defined due to the problem structure. For example, this value can be increased in health cases since mortal decisions are required. In other words, experts can define this parameter for different problems.

4. Numerical Experiments and Results

Numerical experiments are presented in this section to discuss the suggested algorithm's efficiency. Accuracy is used in the numerical experiments for the performance analysis.

Before implementing the algorithm on real-world datasets for comparison, a simple synthetic dataset with one attribute is experimented with, and the results are shown in the figures for better realization. The used datasets are given below:

Training dataset={(1,1),(2,0),(3,0),(8,0),(0,1),(4,1),(5,1),(7,1),(6,1),(10,1),(11,1)}
Test (Fuzzy) dataset={(1.5,0),(0.5,1),(3.5,0),(12,1),(4.5,0),(6.5,0)}

In the algorithm, the training dataset is defined as *A* and *B* sets, and the fuzzy set is defined as *C*. The steps 0, 1, 2, and 3 training datasets are used for getting separative PCFs. The obtained functions (g_1,g_2) and training dataset (blue circles defines "0" labeled data and red circles defines "1" labeled data) are shown in Figure 1. After finding the *g(x)* (separative function), in steps 4 and 5, the membership function is defined via obtained *g(x)* and determined bound value. The obtained membership values of the test (fuzzy) set are shown in Figure 2(a). In step 6, defuzzification for getting crisp values (i.e., labels), equation (13) isplied. For separation y=0.5 line is used (see Figure 2(a)). In Figure 2(b), the classes (labels) of the fuzzy set are shown.

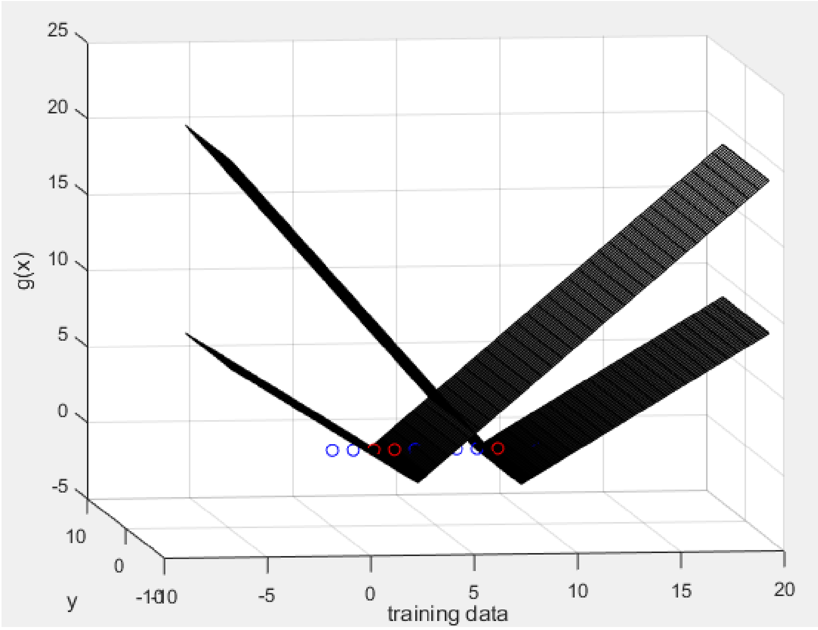

Figure 1. Obtained PCFs (g(x)) with training data.

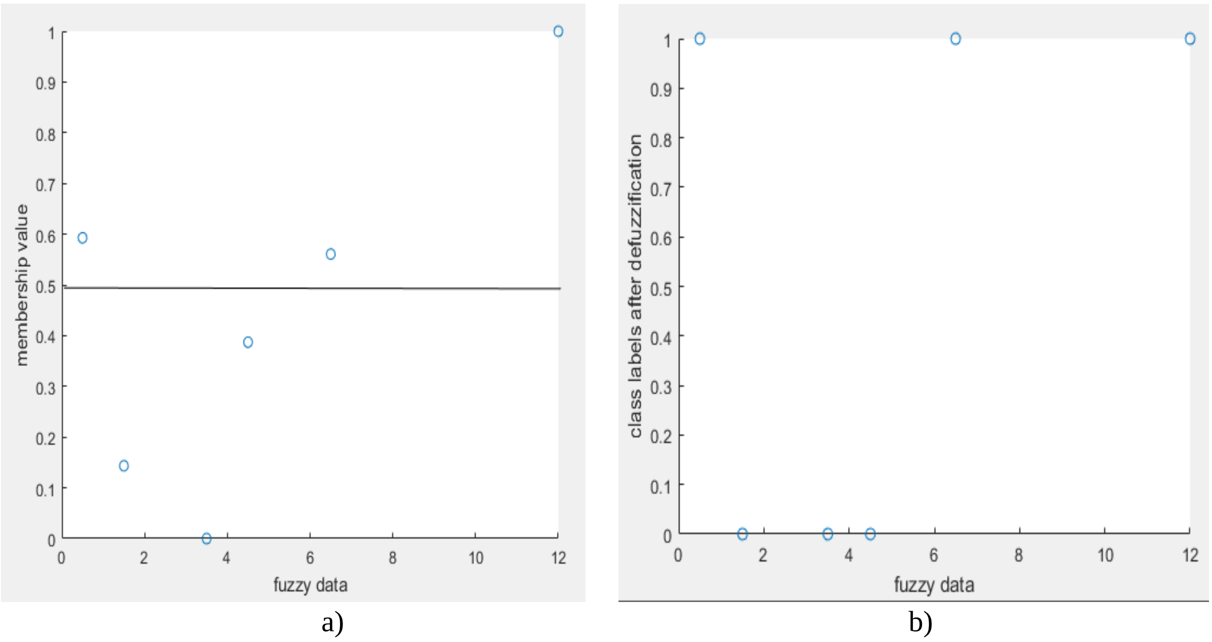

Figure 2. a) Membership values of fuzzy set and defuzzification using line y=0.5 b) fuzzy data and labels after defuzzification.

We make binary classification and use just one discriminant line in defuzzification to get an accurate result. According to the problem structure, the number of classes can be increased by increasing the number of discriminant lines at the defuzzification stage. For this example the obtained accuracy for the test set (fuzzy set) is 83.33% where 5 out of 6 elements are classified correctly. Furthermore, 63.64% for the training set where 7 out of 11 elements are classified correctly.

Besides the presented example of the synthetic dataset, real-world datasets obtained from UCI Machine Learning Repository are applied to the defined algorithm. For comparison, the state of art classification algorithms is applied to the same datasets. Used dataset details are presented in Table 1, and obtained results are presented in Table 2.

Table 1. Details of real-world dataset

Datasets	Number of Instances	Number of Attributes
Vehicle	946	18
Ecoli	336	8
Iris	150	4
Haberman	306	3
Breast Cancer	286	9

The suggested algorithm code is written and implemented on matrix laboratory (MATLAB) program, and the state of art classification algorithms are implemented on Waikato Environment for Knowledge Analysis (WEKA) open-source machine learning software. For each experiment respectively 20% and 80% of the data is used for training and testing.

Table 2. Accuracy results of algorithms

Datasets	Fuzzy Classification	Naive Bayes	Logistics	Classification with Regression
Vehicle	69.81	64.54	94.83	92.9
Ecoli	85.9	91.44	88.74	86.23
Iris	96.00	100	100	100
Haberman	66.08	74.28	74.28	74.28
Breast Cancer	83.21	93.18	89.89	91.86

The results show that the suggested fuzzy classification is not the most effective in the state of art classification algorithms when it acts as a standard classifier that assigns a single crisp label. However, its success is not deniable, and it should not be forgotten that the novelty of this research is defining a new membership function. With this defined membership function, it can be determined which class the fuzzy data will be closer to.

For example, considering the disease processes, by using the available patient data (0-patient, 1-not patient), the sieges (disease proximity processes) can be considered as a cluster, and the cluster (process) of the fuzzy data can be estimated. Here for defining the cluster in step 6, the used equation should be changed according to the defined number of clusters (number of sieges).

5. Conclusion

This article aims to use fuzzy logic in classification by using accessible labeled data. In the supervised learning process, the previously defined PCFs approach for separation is applied, and the obtained PCFs define a novel membership function. For fuzzy classification, defuzzification is determined with the membership values. In order to measure performance during the test stages, binary classification is studied with the obtained membership values. However, according to the problem structure, the number of classes can be increased by increasing the number of discriminant lines at the defuzzification stage, and also suggested membership function can be used in fuzzy logic research for future work.

Acknowledgments

This research has received no external funding.

References

Alpaydın, E. (2010). *Introduction to Machine Learning (2nd ed.).* Cambridge, MA, USA: The MIT Press.

Azam, M. H., Hasan, M. H., Kadir, S. J. A., & Hassan, S. (2021). Prediction of sunspots using fuzzy logic: A triangular membership function-based fuzzy C-means approach. *International*

Journal of Advanced Computer Science and Applications, 12(2), 357-362. doi:10.14569/IJACSA.2021.0120245

Bhattacharyya, R. & Mukherjee, S. (2020). Fuzzy membership function evaluation by non-linear regression: An algorithmic approach. *Fuzzy Information and Engineering*, 12(4), 412-434. doi:10.1080/16168658.2021.1911567

Borkar, S. & Rajeswari, K. (2013). Predicting students academic performance using education data mining. *IJCSMC International Journal of Computer Science and Mobile Computing*, 2(7), 273-279.

Cao, X. H., Stojkovic, I., & Obradovic, Z. (2016). A robust data scaling algorithm to improve classification accuracies in biomedical data. *BMC Bioinformatics*, 17(1), 359. doi:10.1186/s12859-016-1236-x

Gasimov, R. N., & Ozturk, G. (2006). Separation via polyhedral conic functions. *Optimization Methods and Software*, 21(4), 527-540. doi:10.1080/10556780600723252

Jamsandekar, S. & Mudholkar, R. R. (2014). Fuzzy classification system by self generated membership function using clustering technique. *International Journal of Information Technology*, 6(1), 697-704.

Makrehchi, M. & Kamel, M. S. (2011). An information theoretic approach to generating fuzzy hypercubes for if-then classifiers. *Journal of Intelligent & Fuzzy Systems*, 22(1), 33-52. doi:10.3233/IFS-2010-0472

Mendes, R. R. F., de Voznika, F. B., Freitas, A. A., & Nievola, J. C. (2001, September). *Discovering fuzzy classification rules with genetic programming and co-evolution.* Paper presented at Principles of Data Mining and Knowledge Discovery, PKDD 2001, Freiburg, Germany. doi:10.1007/3-540-44794-6_26

Ozturk, G., & Citfci, M. (2015). Clustering based polyhedral conic functions algorithm in classification. *Journal of Industrial and Management Optimization*, 11(3), 921-932. doi:10.3934/jimo.2015.11.921

Rapheal, A. B. & Bhattacharya, S. (2020, January). *A Study on the effect of fuzzy membership function on fuzzified RIPPER for stock market prediction*. Paper presented at 4th International Conference on Machine Learning and Soft Computing, Copenhagen, Denmark. doi:10.1145/3380688.3380716

Sanz, J., Fernandez, A., Bustince, H., & Herrera, F. (2010, July). *A genetic algorithm for tuning fuzzy rule-based classification systems with interval-valued fuzzy sets*. Paper presented at WCCI 2010 IEEE World Congress on Computational Intelligence, Barcelona, Spain. doi:10.1109/FUZZY.2010.5584097

Satı N. U. (2020). A novel semisupervised classification method via membership and polyhedral conic functions. *Turkish Journal of Electrical Engineering and Computer Sciences,* 28(1), 80-92. doi:10.3906/elk-1905-45.

Singh, H., Gupta, M. M., Meitzler, T., Hou, Z. G., Garg, K. K., Solo, A. M. G., & Zadeh, L. A. (2013). Real-life applications of fuzzy logic. *Advances in Fuzzy Systems*, 2013, 581879. doi:10.1155/2013/581879

Uylaş Satı, N. (2015). A binary classification algorithm based on polyhedral conic functions. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 3(1), 152-161. Retrieved from https://dergipark.org.tr/en/pub/dubited/issue/4809/66263.

Xie, W. B., Sang, S., Lam, H. K. & Zhang, J., (2021). A polynomial membership function approach for stability analysis of fuzzy systems. *IEEE Transactions on Fuzzy Systems*, 29(8), 2077-2087. doi:10.1109/TFUZZ.2020.2991149