Data-driven Induction of Shadowed Sets Based on Grade of Fuzziness

Dario Malchiodi[0000−0002−7574−697X] and Anna Maria Zanaboni[0000−0003−0362−5527]

Università degli Studi di Milano, Milano, Italy
{malchiodi, zanaboni}@di.unimi.it

Abstract. We propose a procedure devoted to the induction of a shadowed set through the post-processing of a fuzzy set, which in turn is learned from labeled data. More precisely, the fuzzy set is inferred using a modified support vector clustering algorithm, enriched in order to optimize the fuzziness grade. Finally, the fuzzy set is transformed into a shadowed set through application of an optimal alpha-cut. The procedure is tested on synthetic and real-world datasets.

Keywords: Shadowed sets · Fuzzy set induction · Machine learning · Support vector clustering

1 Introduction

In the last decades fuzzy sets have been proved to be a powerful means for knowledge representation, reasoning and decision making in uncertain contexts. However, as their usage becomes widespread, the trade-off between the detailed nature of the fuzzy membership function and its symbolic interpretation is getting undisguised. A possible way to address uncertainity trying to manage this trade-off is to identify three regions of the universe of discourse, namely a belongingness region, an exclusion region and a "grey" region where genuine uncertainity holds. A lot of work as been done in this direction in different research areas (rough sets [22], fuzzy sets [4, 8, 23], three-valued logic [3], type-2 fuzzy logic [21], see [15, 24] for a survey). We will focus on the construct of shadowed sets introduced by Pedrycz [18,19], and used in different learning contexts [5, 9, 10, 13, 25, 26]. Given a shadowed set A, the domain of discourse is split into three regions, called the *core*, the *exclusion* and the *shadowed* region, where membership value to A is 1, 0 and unknown, respectively. The shadowed set is induced by a fuzzy set because its shadowed regions' position and width are determined by the constraint of preserving the amount of fuzziness of the originating fuzzy set. More precisely, the induced shadowed set is completely determined by an *α-cut*, namely a value $0 \leq \alpha \leq 1/2$ used to cut the codomain [0,1] of the fuzzy membership function into the zones $[0, \alpha]$, $(\alpha, 1 - \alpha)$, $[1 - \alpha, 1]$ where full belongingness, uncertainity and full exclusion are, respectively, assigned. This gives rise to a membership function $S_A : X \mapsto \{0, [0, 1], 1\}$, where

X is the universe of discourse and 1, $[0, 1]$, 0 are associated to the three mentioned zones.

The search for the α-cut is an optimization problem; moreover, the particular definition of fuzziness of a fuzzy set obviously affects the resulting procedure. In [20] analytical formulas are provided to calculate the optimal α-cut using the gradual grade of fuzziness, and a comparison with other fuzziness measures is discussed.

We describe a data-driven procedure for the induction of shadowed sets based on the post-processing of a fuzzy set learned from labeled data. The procedure exploits a support vector clustering [1] algorithm in which the inference is done starting from a set of objects in X, labeled with their membership degrees to A. As a next step, a sphere S in a space H is found so that the images of objects through a function $\Phi : X \mapsto H$ are positioned w.r.t. S in function of the membership degrees. More precisely, in case of unitary membership the image of an object will belong to S, otherwise it will fall farther from the border of S as its membership to A decreases from 1 to 0 [16]. This learning algorithm is further enriched with the optimization of the fuzziness grade of A.

The paper is organized as follows: Sect. 2 is devoted to the derivation of the fuzziness degree of a piecewise linear membership function and its optimal α-cut, Sect. 3 is devoted to (a) the description of the modified support vector clustering optimization problem for learning a membership function and (b) to its enrichment with a term accounting for fuzziness degree minimization of the inferred shadowed set. In Sect. 4 we discuss experimental results on a synthetic and two real-world benchmarks. Some concluding remarks end the paper.

2 Gradual grade of fuzziness of a fuzzy set

The fuzziness grade of a fuzzy set measures the *vagueness* of the set itself. Such concept captures the amount of entropy inherently contained in a fuzzy set: indeed crisp sets have a null fuzziness grade, while on the other hand the maximal grade is attained by a fuzzy set with membership function constantly equal to 1/2. As a rule of thumb, the sharper the boundaries of a fuzzy set, the smaller the related fuzziness. Among the proposed measures quantifying the fuzziness grade of a fuzzy set (see for instance [12, 14]), we consider the one introduced in [11] quantifying the fuzziness grade of a continuous, measurable fuzzy set A whose membership function is μ_A as

$$\varphi(A) = \int_X (1 - |2\mu_A(x) - 1|)\mathrm{d}x \ .$$

The notion of fuzziness grade is linked to the search of an optimal α-cut transforming a fuzzy set into a shadowed set [20]. Namely, denoted by ω_1, ω_2 and ω_3 the definite exclusion, the definite belongingness and the uncertainty regions mentioned in the Introduction and restricting to them the fuzziness degree computation, the optimal α is such that

$$\varphi(\omega_1) + \varphi(\omega_2) = \varphi(\omega_3) \qquad (1)$$

holds. In this way, the overall fuzziness of A is equally balanced between the shadowed (ω_3) and unshadowed ($\omega_1 \cup \omega_2$) regions.

In the rest of this paper we will focus on the family of piecewise linear functions whose general member has the following form

$$f_{R^2,M}(x) = \begin{cases} 1 & \text{if } x \leq R^2 \quad , \\ 1 - \frac{x-R^2}{M-R^2} & \text{if } R^2 < x \leq M \\ 0 & \text{otherwise} \quad , \end{cases} \quad ,$$

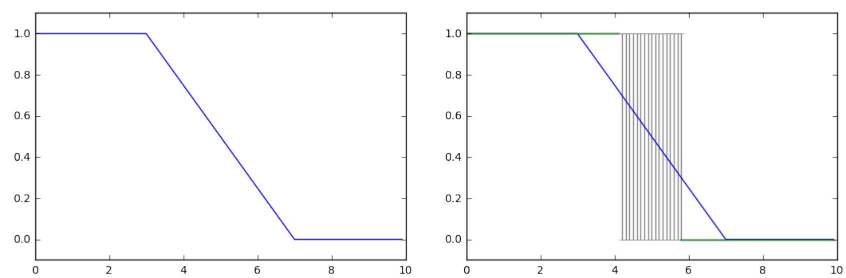
where $R^2 > 0$ and $M > R^2$ denote the boundaries of the crisp regions of the fuzzy set (see Fig. 1(a))[1]. It is easy to show that the fuzziness degree of a fuzzy set A whose membership function has the form $\mu_A(x) = f_{R^2,M}(x)$ is

$$\varphi(A) = \frac{M - R^2}{2} \tag{2}$$

while, for fixed α

$$\begin{aligned} \varphi(\omega_1) &= \alpha^2 \left(M - R^2\right) \quad , \\ \varphi(\omega_2) &= \alpha^2 \left(M - R^2\right) \quad , \\ \varphi(\omega_3) &= 2(1-\alpha^2) \left(M - R^2\right) \quad , \end{aligned}$$

thus the optimal cut condition (1) reads $\alpha^2 = 1 - \alpha^2$, which corresponds to $\alpha = \sqrt{2}/2$.



(a) (b)

Fig. 1. Graph of (a) a membership function μ_A to a fuzzy set in the considered family, and (b) the membership function S_A to a shadowed set obtained from (a) after an optimal α-cut. Blue curve: graph of μ_A; green segments: crisp values of S_A; gray area: uncertainty zone of S_A.

[1] The choice of R^2 and M as names for these symbols is linked to a special role they will play in Sect. 3.

3 Shadowed set induction

The proposed procedure learns a shadowed set in two phases: the first one infers a fuzzy set starting from a set of labeled objects, while the second phase performs on this set the α-cut described in the previous section.

Focusing on the first phase, consider a universe of discourse X, fix $n \in \mathbb{N}$ and denote by $\{x_1, \dots, x_n\} \in X^n$ a sample of objects. Given also a set of labels $\{\mu_1, \dots, \mu_n\} \in [0, 1]^n$ whose values are the membership degrees of objects to an unknown fuzzy set A, the membership function μ_A can be learned using the approach proposed in [16] optimizing the square radius R^2 of a sphere S belonging to a space H and centered in a. Namely, the images of objects through a function $\Phi : X \mapsto H$ are such that $\Phi(x_i) \in S$ when $\mu_i = 1$, and $\Phi(x_i)$ tends to fall farther from the border of S as μ_i decreases from 1 to 0. This amounts to modifying the support vector clustering algorithm proposed in [1] as follows:

$$\min R^2 + C \sum_{i=1}^{n} (\xi_i + \tau_i) \tag{3}$$

$$\mu_i ||\Phi(x_i) - a||^2 \leq \mu_i R^2 + \xi_i \ \forall i = 1, \dots, n \ , \tag{4}$$

$$(1 - \mu_i) ||\Phi(x_i) - a||^2 \geq (1 - \mu_i) R^2 - \tau_i \ \forall i = 1, \dots, n \ , \tag{5}$$

$$\xi_i \geq 0, \tau_i \geq 0 \quad \forall i = 1, \dots, n \ . \tag{6}$$

In this formulation $C > 0$ measures the relative importance of the two components in the objective function, while ξ_i and τ_i are slack variables relaxing the constraints dealing with the positioning of points inside and outside S, respectively. Once S has been learned, the membership function μ_A is obtained by mapping its argument x to 1 if $\Phi(x) \in S$, and to a value belonging to $[0, 1)$ otherwise. This value is computed applying a suitable function f to the squared distance $r^2(x) = ||\Phi(x) - a||^2$. We will choose f within the family described in Sect. 2, dropping subscripts for sake of conciseness. According to (2), the problem (3–6) can be easily modified in order to take into account also the minimization of the fuzziness degree of the inferred set as follows:

$$\min R^2 + C \sum_{i=1}^{n} (\xi_i + \tau_i) + D(M - R^2) \tag{7}$$

$$\mu_i ||\Phi(x_i) - a||^2 \leq \mu_i R^2 + \xi_i \ \forall i = 1, \dots, n \ , \tag{8}$$

$$(1 - \mu_i) ||\Phi(x_i) - a||^2 \geq (1 - \mu_i) R^2 - \tau_i \ \forall i = 1, \dots, n \ , \tag{9}$$

$$||\Phi(x_i) - a||^2 \leq M\psi \ \forall i = 1, \dots, n \ , \tag{10}$$

$$\xi_i \geq 0, \tau_i \geq 0 \quad \forall i = 1, \dots, n \ . \tag{11}$$

In this new formulation, $D > 0$ is a new hyperparameter jointly ruling with C the relative importance of the components in (7), namely devoted to the optimization of radius, slack variables, and fuzziness degree. Analogously, M is introduced as a new variable, bounded in (10) to be higher than the distance between a and any of the images $\Phi(x_i)$. Jointly considering this constraint and the objective function (7) amounts to requiring M to equal the maximum of such distances. Actually, an additional hyperparameter ψ in (10) allows to fine tune this requirement: $\psi > 1$ promotes higher values for M, and *vice versa*.

Letting $E = (1 - D(1 - 1/\psi))$ and denoting with k the kernel associated to Φ (that is, $k(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$), the Wolfe dual of (7–11) corresponds to the maximization of

$$\sum_{i=1}^{n}(\epsilon_i + \beta_i)k(x_i, x_i) - E^{-1} \sum_{i,j=1}^{n}(\epsilon_i + \beta_i)(\epsilon_j + \beta_j)k(x_i, x_j) \tag{12}$$

subject to the constraints

$$\sum_{i=1}^{n} \epsilon_i = 1 - D \ , \tag{13}$$

$$\sum_{i=1}^{n} \beta_i = D/\psi \ , \tag{14}$$

$$-(1 - \mu_i)C \leq \epsilon_i \leq \mu_i C \ \forall i = 1, \dots, n \ , \tag{15}$$

$$\beta_i \geq 0 \ \forall i = 1, \dots, n \ . \tag{16}$$

It it easy to see that β_i is the generic Lagrangian multiplier associated to (10), while $\epsilon_i = \mu_i \gamma_i - (1 - \mu_i)\hat{\gamma}_i$, being γ_i and $\hat{\gamma}_i$ the multipliers for (8) and (9). In order to be solvable, the dual problem requires $D \neq \psi/(\psi - 1)$, otherwise the objective function would not be computable.

In the experiments described later on, we will consider two kinds of kernel: the *linear* kernel defined by $k(x_i, x_j) = x_i \cdot x_j$, and the family of *Gaussian* kernels defined by $k(x_i, x_j) = \exp\left(-||x_i - x_j||^2/\sigma^2\right)$, where $\sigma > 0$ is an additional hyperparameter to be considered. This use of the so-called *kernel trick* allows to consider a universe of discourse whose members are not necessarily numbers or numerical vectors. For instance, [17] uses a similar technique in order to solve the problem of detecting a set of reliable axioms in the context of semantic Web.

Dealing with the KKT conditions [7] is a bit tricky, because these are expressed in terms of γ_i, $\hat{\gamma}_i$, β_i, and the remaining Lagrange multipliers. For sake of conciseness, we just list the salient relations linking primal and dual variables when we consider the optimal solution:

$$0 < \gamma_i < C \to R^2 = r^2(x_i) \ , \tag{17}$$

$$0 < \hat{\gamma}_i < C \to R^2 = r^2(x_i) \ , \tag{18}$$

$$\beta_i > 0 \to M = \psi^{-1} r^2(x_i) \ , \tag{19}$$

where $r^2(x) = ||\Phi(x) - a||^2$ can be obtained as

$$r^2(x) = k(x, x) - 2E^{-1} \sum_{i=1}^{n}(\epsilon_i + \beta_i)k(x, x_i) + \\ + E^{-2} \sum_{i,j=1}^{n}(\epsilon_i + \beta_i)(\epsilon_j + \beta_j)k(x_i, x_j) \ .$$

Table 1. Relations between the dual variables γ_i, $\hat{\gamma}_i$, and β_i.

	$\hat{\gamma}_i = 0$	$0 < \hat{\gamma}_i < C$	$\hat{\gamma}_i = C$
$\gamma_i = 0$	$\epsilon_i = 0$	$-C(1-\mu_i) < \epsilon_i < 0$	$\epsilon_i = -C(1-\mu_i)$
$0 < \gamma_i < C$	$0 < \epsilon_i < C\mu_i$	$-(1-\mu_i)C < \epsilon_i < \mu_i C$	$-C(1-\mu_i) < \epsilon_i < C(2\mu_i - 1)$
$\gamma_i = C$	$\epsilon_i = C$	$C(2\mu_i - 1) < \epsilon_i < C\mu_i$	$\epsilon_i = C(2\mu_i - 1)$

The problem here is that (7–11) explicitly depend only on ϵ_i and β_i, thus only (19) is directly exploitable. By analyzing all combinations between the critical values of γ_i and $\hat{\gamma}_i$ and computing the corresponding values for ϵ_i (see Table 1), it is easy to check that

$$0 < \epsilon_i < C\mu_i \rightarrow 0 < \gamma_i < C \ , \tag{20}$$

$$-C(1-\mu_i) < \epsilon_i < 0 \rightarrow 0 < \hat{\gamma}_i < C \ . \tag{21}$$

Jointly considering (17–21) it is therefore possible to link the optimal values of dual and primal variables:

$$0 < \epsilon_i < C\mu_i \rightarrow R^2 = r^2(x_i) \ , \tag{22}$$

$$-C(1-\mu_i) < \epsilon_i < 0 \rightarrow R^2 = r^2(x_i) \ , \tag{23}$$

$$\beta_i > 0 \rightarrow M = \psi^{-1} r^2(x_i) \ . \tag{24}$$

Once R^2 and M have been found, a shadowed set can be obtained from the corresponding fuzzy set through application of the optimal α-cut described in Sect. 2.

4 Experiments

As a first set of experiments, we tested the sensitivity of the overall learning procedure to hyperparameters[2]. Focusing on C, we considered a synthetic dataset composed by seven points whose crisp membership[3] has been fixed according to an interval, thus suggesting a unimodal membership both to a fuzzy and a shadowed set. Using a linear kernel and fixing $D = 0.3$ and $\psi = 1$, Fig. 2 shows that rising C has the effect of sharpening the boundaries of the fuzzy set. In other words, as C grows the membership increases from 0 to 1 (and decreases from 1 to 0) in a more linear fashion.

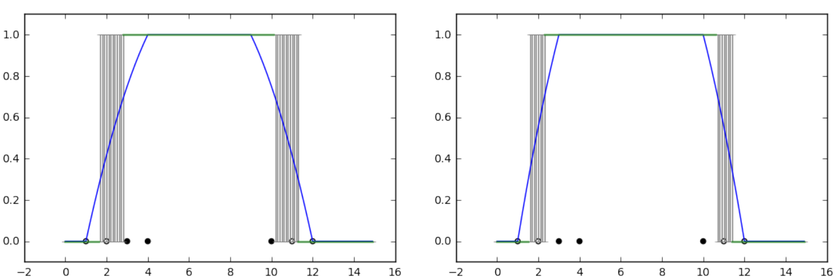
Figures 3 and 4 describe analogous experiments focusing on the role of D and ψ which we can summarize as follows:

[2] Code and data to replicate experiments are available at `https://github.com/dariomalchiodi/WILF2018`.

[3] It is worth highlighting that the learning algorithm of Sect. 3 can in principle be run on objects labeled using more generic membership grades (that is, values belonging to $[0, 1]$). However, as such a rich information is normally not available in public datasets, all reported experiments rely on crisp membership labels.

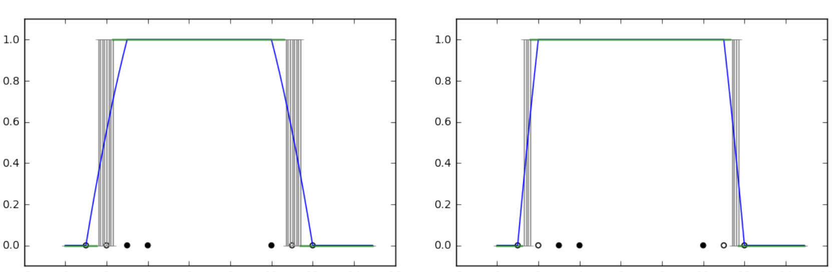
– D is directly related to the amount of uncertainty of the inferred shadowed set (the higher its value, the lower the fuzziness degree of the set);
– ψ primarily influences the *localization* of the inferred set, although it also affects the optimal value of M, thus it has an impact on the uncertainty described in the previous point, too.

Finally, the kernel choice obviously affects the general form of μ_A, and thus also S_A. For instance, Fig. 5 shows the effect of decreasing the parameter of a Gaussian kernel when learning S_A on the same dataset of Fig. 2 and 3. The augmented plasticity in the considered class of functions allows the procedure to find bimodal memberships concentrating around the positive points as σ decreases.



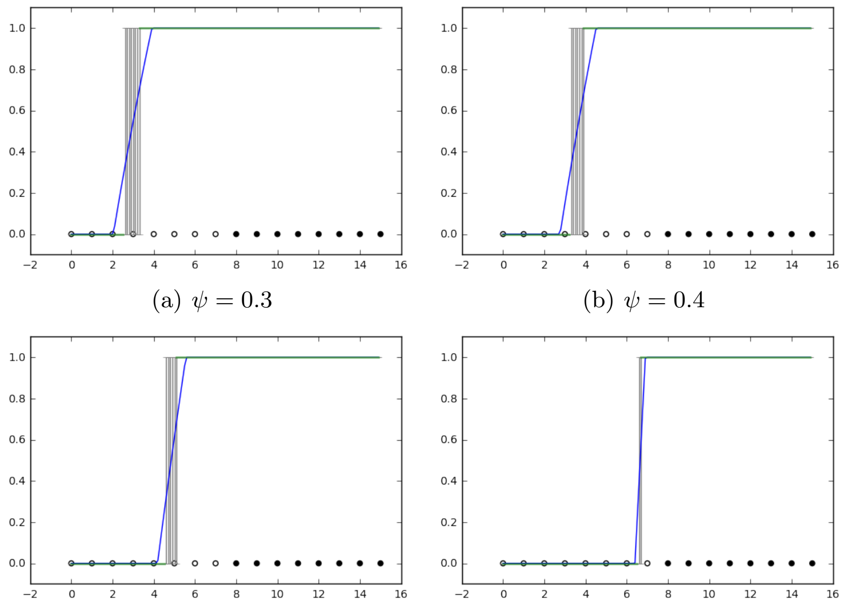
(a) $C = 0.3$ (b) $C = 6$

Fig. 2. Effect of changes of C on the membership functions learned using a linear kernel and setting $D = 0.3$ and $\psi = 1$. The fuzzy and shadowed membership functions were plotted using the same notation of Fig. 1. For each sample point, a bullet on the X-axis is drawn using black and white color when $\mu_i = 1$ and $\mu_i = 0$.



(a) $D = 0.1$ (b) $D = 2$

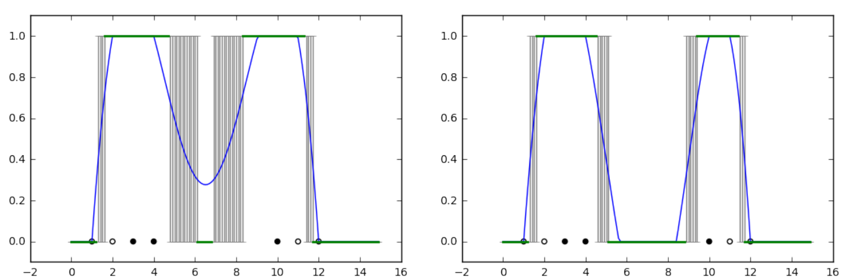
Fig. 3. Effect of changes in the parameter D on the learned unimodal membership function to a shadowed set, letting $C = 10$, $\psi = 1$ and using a linear kernel. Same notation as in Fig. 2.



(a) $\psi = 0.3$ (b) $\psi = 0.4$

(c) $\psi = 0.6$ (d) $\psi = 1$

Fig. 4. Effect of changes in the parameter ψ on the membership function to a shadowed set learned with a linear kernel and setting $C = 1$ and $D = 4$. Same notation as in Fig. 2.



(a) $\sigma = 5$ (b) $\sigma = 3$

Fig. 5. Effect of changes in the parameter σ of the used Gaussian kernel on the bimodal membership function to a shadowed set learned when $C = 30$, $\psi = 1$ and $D = 0.8$. Same notation as in Fig. 2.

Switching to a non-synthetic dataset, Fig. 6 shows the result of the proposed technique for a sample from the veterinary domain, in which each observation is the measurement of the level of kidney function (namely, the rate of glomerular filtration, measured in mL/min/Kg), in a set of 37 dogs, each one labeled either

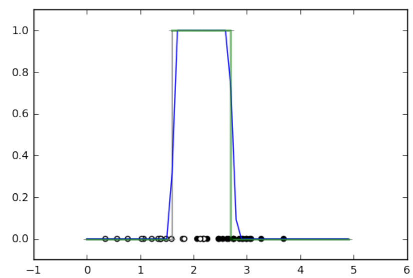


Fig. 6. Membership function for the fuzzy and shadowed sets capturing the concept of "ill dog" expressed by a real-world dataset. Same notation as in Fig. 2.

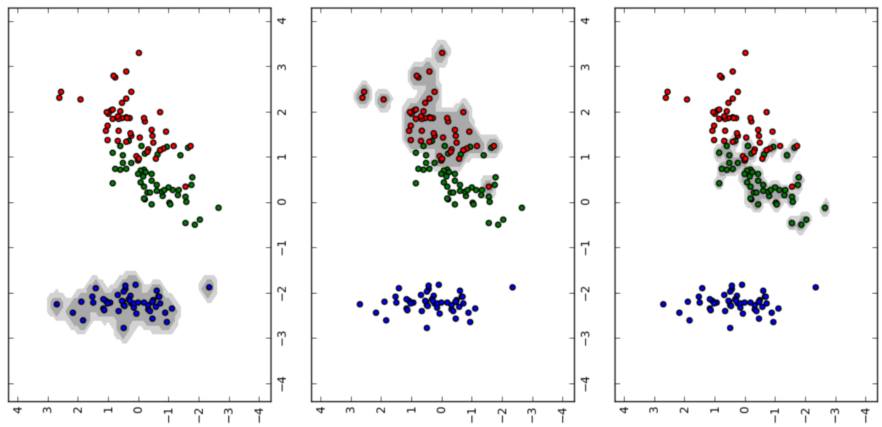
as "ill" or "healthy". In this case the procedure relied on a Gaussian kernel with parameter $\sigma = 0.3$, fixing $C = 0.5$, $D = 0.45$ and $\psi = 1$.

As the described learning algorithm can handle objects of arbitrary dimension, we also considered the Iris dataset [6], gathering the observations of 150 iris plants, expressed as a 4-dimensional vector (sepal length, sepal width, petal length, petal width), with length and width measured in centimeters. Each observation belongs to exactly one of the classes *Setosa*, *Virginica*, and *Versicolor* (where the first one is linearly separable from the remaining two, and the latter are linked by a more complex relationship). For sake of visualization, we extracted the first two principal components from the observations and performed the shadowed set inference for all the available classes, each time labeling with $\mu_i = 1$ the observations referring to the target class and with $\mu_i = 0$ the remaining observations. Figure 7 shows the obtained results when the whole dataset is considered, using a Gaussian kernel (the related value for σ, as well as those of the remaining hyperparameters have been chosen after an exploratory procedure). In the figure, bullets are superimposed to a visualization of the membersip functions where a dark gray and white background respectively refer to positive and negative values, while a light gray background shows the uncertain areas.

In order to get quantitative results, we performed a more accurate experiment in which the following holdout scheme was iterated one hundred times.

- We randomly shuffled data and subsequently performed a stratified sampling in order to get a training and a test set gathering respectively 80% and 20% of the available items. Stratification ensured training and test sets to be balanced (in the sense that the three classes are equally represented).
- We applied the inference procedure to the training set, obtaining three membership functions to shadowed sets, each linked to a specific Iris class.
- We assessed the joint performance of these three shadowed sets by assigning each object to the class maximizing the membership function value (using the obvious order $0 < [0, 1] < 1$) and comparing the result with the target label[4].

[4] Ties were resolved in favor of the correct class, when possible.



(a) S_{Setosa} (b) $S_{Virginica}$ (c) $S_{Versicolor}$

Fig. 7. Inferred shadowed sets for the Iris dataset. Bullets show the two principal components of each data item, colored in blue, red, and green respectively for the *Setosa*, *Virginica*, and *Versicolor* classes. Dark gray, light gray and white background correspond to the positive, uncertain, and negative values for the membership function.

Table 2. Results of one hundred holdout iterations of a joint shadowed set learning procedure on the Iris dataset. Each row shows average, median, and standard deviation of test error, in function of the number of principal components (# PC) extracted from the original sample.

# PC	Average error	Median error	Error std
2	0.07	0.07	0.04
3	0.03	0.03	0.02
4	0	0	0

The experiments were repeated extracting two, three, and four principal components from the dataset. Table 2 summarizes the obtained results: the proposed approach definitely learns the Iris dataset, outperforming similar techniques based on the sole induction of fuzzy sets [2, 16].

5 Conclusions

Reducing the complexity of structures described in terms of fuzzy sets has the desirable effect of allowing an easier interpretation of models induced from data. With this aim, we propose a learning algorithm for shadowed sets, which are sets endowed with a three-valued membership function defining full membership, full exclusion and genuine uncertainty w.r.t. candidate points. This algoritm identifies the shadowed set according to an optimal α-cut performed on a fuzzy set, in turn inferred from data using a modified support vector clustering approach also optimizing the fuzziness degree. A preliminary set of experiments on synthetic data allowed us to gain better insights on the role of hyperparameters; we also tested the procedure on real-world datasets, getting improvements with respect to a previous approach solely based on fuzzy sets. Besides a deeper experimentation phase, we plan to extend the technique considering different families of membership functions, as well as the jointly learning of several shadowed sets.

References

1. Ben-Hur, A., Horn, D., Siegelmann, H.T., Vapnik, V.: Support Vector Clustering. Journal of Machine Learning Research **2**(Dec), 125–137 (2001)
2. Cermenati, L., Malchiodi, D., Zanaboni, A.M.: Simultaneous learning of fuzzy sets. In: Proceedings of the 2018 Italian Workshop on Neural Networks. Springer (2018), submitted
3. Ciucci, D., Dubois, D., Lawry, J.: Borderline vs. unknown: Comparing three-valued representations of imperfect information. International Journal of Approximate Reasoning **55**(9), 1866–1889 (2014)
4. Deng, X., Yao, Y.: Decision-theoretic three-way approximations of fuzzy sets. Information Sciences **279**, 702–715 (2014)
5. El-Hawy, M., Wassif, K., Hefny, H., Hassan, H.: Hybrid multi-attribute decision making based on shadowed fuzzy numbers. In: 2015 IEEE 7th International Conference on Intelligent Computing and Information Systems, ICICIS 2015. pp. 514–521 (2016)
6. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Annals of Eugenics **7**(2), 179–188 (1936)
7. Fletcher, R.: Practical Methods of Optimization. John Wiley & Sons (2013)
8. Grzegorzewski, P.: Fuzzy number approximation via shadowed sets. Information Sciences **225**, 35–46 (2013)
9. Hryniewicz, O.: Possibilistic Analysis of Bayesian Estimators when Imprecise Prior Information is Described by Shadowed Sets. Advances in Intelligent Systems and Computing **642**, 238–247 (2018)
10. Jian, C.: Improved Shadowed Sets Data Selection Method in Extension Neural Network. Journal of Networks **8**(12), 2728–2735 (2013)
11. Klir, G.J., St. Clair, U.H.S., Yuan, B.: Fuzzy Set Theory: Foundations and Applications. Prentice Hall (1997)
12. Knopfmacher, J.: On measures of fuzziness. Journal of Mathematical Analysis and Applications **49**(3), 529–534 (1975)
13. Li, X., Geng, P., Qiu, B.: A cluster boundary detection algorithm based on shadowed set. Intelligent Data Analysis **20**(1), 29–45 (2016)
14. Liu, B.: Uncertainty theory. In: Uncertainty Theory, pp. 1–79. Springer (2010)
15. Lu, J., Han, J., Hu, Y., Zhang, G.: Multilevel decision-making: A survey. Information Sciences **346-347**, 463–487 (2016)
16. Malchiodi, D., Pedrycz, W.: Learning Membership Functions for Fuzzy Sets through Modified Support Vector Clustering. In: Masulli, F., Pasi, G., Yager, R.R. (eds.) Fuzzy Logic and Applications - 10th International Workshop, WILF 2013, Genoa, Italy, November 19-22, 2013. Proceedings. vol. LNCS 8256, pp. 52–59. Springer (2013)
17. Malchiodi, D., Tettamanzi, A.G.B.: Predicting the Possibilistic Score of OWL Axioms through Modified Support Vector Clustering. In: SAC 2018: Symposium on Applied Computing, April 9–13, 2018, Pau, France. ACM, New York, NY, USA (2018). https://doi.org/10.1145/3167132.3167345

18. Pedrycz, W.: Shadowed sets: Representing and processing fuzzy sets. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics **28**(1), 103–109 (1998)
19. Pedrycz, W.: From Numeric to Granular Description and Interpretation of Information Granules. Fundamenta Informaticae **127**(1-4), 399–412 (2013)
20. Tahayori, H., Sadeghian, A., Pedrycz, W.: Induction of Shadowed Sets Based on the Gradual Grade of Fuzziness. IEEE Transactions on Fuzzy Systems **21**(5), 937–949 (2013)
21. Wijayasekara, D., Linda, O., Manic, M.: Shadowed Type-2 Fuzzy Logic Systems. In: Proceedings of the 2013 IEEE Symposium on Advances in Type-2 Fuzzy Logic Systems, T2FUZZ 2013 - 2013 IEEE Symposium Series on Computational Intelligence, SSCI 2013. pp. 15–22 (2013)
22. Yao, Y.: Rough Sets and Three-Way Decisions. In: Proceedings of the International Conference on Rough Sets and Knowledge Discovery. LNCS, vol. 9436, pp. 62–73. Springer (2015)
23. Yao, Y., Wang, S., Deng, X.: Constructing shadowed sets and three-way approximations of fuzzy sets. Information Sciences **412-413**, 132–153 (2017)
24. Zhang, G., Lu, J., Gao, Y.: Multi-Level Decision Making, Models, Methods and Application. Springer, Berlin (2014)
25. Zhang, J., Shen, L.: An Improved Fuzzy c-Means Clustering Algorithm Based on Shadowed Sets and PSO. Computational Intelligence and Neuroscience **2014** (2014)
26. Zhou, Y., Tian, L., Liu, L.: Improved Extension Neural Network and its Applications. Mathematical Problems in Engineering **2014** (2014)