Managing imprecisely observed data in Genetic Fuzzy Models by means of a fuzzy-valued fitness function

Luciano Sánchez
Dept. Informática
Universidad de Oviedo
Campus de Viesques
33271 Gijón
luciano@uniovi.es

Inés Couso
Dept. Estadística e I.O. y D.M.
Universidad de Oviedo
Campus de Viesques
33271 Gijón
couso@pinon.ccu.uniovi.es

Abstract

In previous works, we have studied the properties of Genetic Fuzzy Classifiers, when used with interval and fuzzy-valued data. For certain kind of problems, it was proved that they were inherently better than those arising from the bayesian point of view. The design of these classifiers was based on the use of a fuzzy-valued fitness function, built after a generalized definition of the density function of a fuzzy random variable.

Afterwards, we have been asked whether an equivalent definition could be derived for genetic fuzzy models, and what advantages we would expect it to have. This paper is devoted to answer this question. We have applied our definition of density of a f.r.v. to obtain a fuzzy characterization of our knowledge about the conditional expectation of the output, given a fuzzy input. Arising from this characterization, we have obtained the expression used in fuzzy least squares, and propose some different minimization criteria that can be used with Genetic Fuzzy Systems.

Keywords: Genetic Fuzzy Sytems, Fuzzy Models, Random Sets, Fuzzy Least Squares.

1 Introduction

Genetic Fuzzy Systems use fuzzy techniques to obtain linguistically understandable rules from crisp data [5]. Nevertheless, in fuzzy statistics the terms "fuzzy classifier" and "fuzzy model" are applied to extensions of discriminant analysis and regression that can cope with vague data [2]. This disparity in the use of the same terms is a clear symptom that connections between Fuzzy Statistics and Genetic Fuzzy Classifiers and Models are not frequent. Besides, some elements of fuzzy statistics can be used with GFSs in a natural manner, and we think that potentially can lead to the development of more powerful learning methods. In previous works [15], we have also suggested that the kind of problems where GFSs are inherently better than their stochastic counterparts is composed by those problems including imprecisely observed data. To justify this assert, an extended definition of the classification problem was proposed, and a fuzzy-valued fitness function, adequate for that problem, was introduced.

In this paper, we intend to apply the same procedures to the modeling problem, and to obtain its corresponding fuzzy fitness function. Besides, the modeling problem is not as closed as the classification one, and more than one definition is possible. Contrary to the classification case, where the bayesian framework provided us with a widely accepted definition of the optimal classifier, there exist many definitions of statistical regression (a survey of many of them can be found in [3].) We have decided to evaluate the fuzzy extension of the standard case first, and leave robust regression

Actas del IV Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados, MAEB2005, pp.499-506
ISBN: Vol I:84-9732-440-4 Vol II: 84-9732-467-6

techniques to be done in future works. Therefore, as we will show in the next section, in this work we will define the output of the model as the conditional expectation of the output random variable, given a certain input.

This paper is structured as follows: first, the statistical definition of modeling under stochastic noise is introduced. Then, an extension to the modeling definition is introduced. This extension copes with both stochastic noise and observation errors, and is based upon the relations between random and fuzzy sets. In section four, it is shown, by means of one example, how the fuzzy valued fitness of the extended GFS is evaluated. The work finishes with the concluding remarks and a discussion about the new opened research lines.

2 Statistical Models and GFSs

2.1 Least squared models and conditional expectation

Let us suppose we have a set Ω that contains objects ω, and let us admit also that each one of them is assigned a numerical value $Y(\omega)$. We are given a set of measurements $X(\omega) = (X_1(\omega), X_2(\omega), \ldots)$ over every object. We will say that a model is a mapping that associates every element of $X(\Omega)$ with a value $g(X)$, whose main objective is to minimize the differences between $Y(\omega)$ and $g(X(\omega))$ over Ω.

For example, let Ω be a set of people. We observe the height and the weight of a randomly selected person, and want to know its expected percentage of body fat. Suppose that someone measures and weights $X(\omega) = (180, 82)$, and has $Y(\omega) = 20\%$ of fat. We wish that the difference between the value that our model assigns to him, $g(180, 82) = 22$, and the true value $Y(\omega) = 20$ is as low as possible. If we admit that there can exist two different people that measure 180 cm. and weight 82 kg., but have a different percentage of fat because of their different body constitution, the assigment $g(\mathbf{x}) = Y(X^{-1}(\mathbf{x}))$ can not be defined. In the example at hand, this means that the model will assign the same value $g = 22$ to all people that measure 180 cm. and weight 82 kg, therefore the optimal model should be defined with respect to averaged weight differences.

To define the concept "averaged differences" we need to assume that the mappings X and Y fulfill all necessary conditions to be random variables. Let us also define a new random variable that quantifies the cost of assigning the value y to an object, when its true value is z, $\operatorname{cost}(y, z)$. For the problem stated, if the model is a mapping $g(\mathbf{x})$, and $Y(\omega)$ is the value associated to the object ω, then the merit value of the model can be numerically quantified as

$$\operatorname{err}(g) = \int_{\Omega} \operatorname{cost}(g(X(\omega)), Y(\omega))\, \mathrm{dP} \quad (1)$$

where the error function is integrated with respect to a probability measure P defined over Ω.

If we choose $\operatorname{cost}(y, z) = (y - z)^2$, the expectation of the cost function is the mean squared error, that gives rise to "least squares regression". Given that $E[(Y - g(X))^2] \geq E[(Y - E[Y|X])^2]$ for any function g, the conditional expectation $g(\mathbf{x}) = E[Y|X = \mathbf{x}]$ is then the optimal definition of model:

$$g(\mathbf{x}) = \frac{\int_{\omega: X(\omega)=\mathbf{x}} y\ dP}{\int_{\omega: X(\omega)=\mathbf{x}} dP} = \int y\, f(y|\mathbf{x})\, dy \quad (2)$$

where $f(y|\mathbf{x})$ is the (Radon-Nikodym) conditional density of the output variable conditioned to a given input variable.

2.2 GFS and fuzzy data

There are many interval and fuzzy valued extensions of the modeling problem, in both the statistics [8, 16] and the fuzzy rule learning fields [11, 12, 13, 14]. Anyway, the most widely used genetic methods for learning fuzzy models (Michigan, Pittsburgh and Iterative Learning [4]) are least-squares based.

In least-squares based learning methods, the genetic algorithm is designed to minimize the estimation of Eq. (1) over the population, using a standard experimental design (leave one out, cross validation, etc.). But, being the optimal classifier defined by Eq. (2), it is immediate that, whenever the quality of a fuzzy

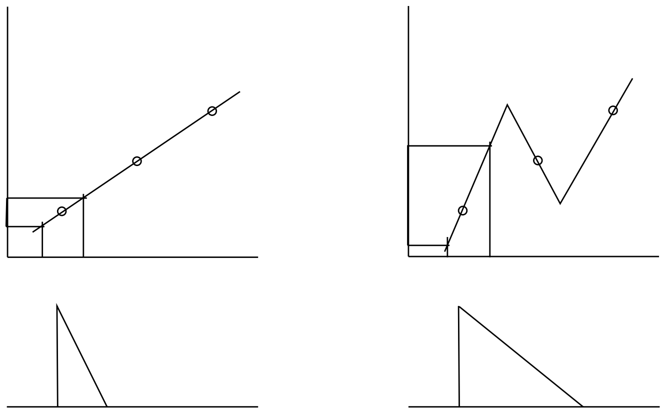

Figure 1: The specificity of the fuzzy output of a model carries information about the slope of the regression function. Upper part: Both models have the same numerical error (0) but different fuzzy error. Lower part: Fuzzy errors, according the the method explained in this paper. The right one is less specific, and that means that the slope of $\hat{y}_B(x)$ is higher in the sample points. Optimizing both the punctual error and the specificity adds a regularization term to the objective function.

model is assessed by means of its mean squared error over a sample, the best fuzzy models must be nonparametric estimators of the conditional expectation. Again, as we pointed out in [15], from a statistical point of view, the crisp problem is being solved, and not the fuzzy one, and therefore least-squared based GFS can not improve the accuracy of statistical methods over crisp problems, no matter the complexity of the genetic search. "Fuzzy" means here that the parameterising of the discriminant functions has a linguistic interpretation compatible with the fuzzy logic postulates.

When the output of a fuzzy model is *defuzzified* before it is compared to that of the other models, some information that could help us to select a good model is being discarded. In fact, we suggest that the information being discarded embodies the difference between standard (non linear least squares) statistical regression and fuzzy models. For instance, observe that the specificity of the output carries information about the slope of the model, which can be used to find regular models (see Figure 1.) Many other uses of the extra information are possible. In Figure 2 it is proposed that the α-cut of the fuzzy error that contains the value 0, in combination with the specificity of the error, can be used to search for regular models that are not further than a certain distance from all the points in the sample. This seems to us that some relations with ϵ-insensitive SVMs [6] may exist.

3 An extended definition of the modeling problem

In this section we will propose a definition of fuzzy model based on the concept of fuzzy random variable as a nested family of random sets, which in turn are defined as imprecise observations of an unknown random variable, so called the *original* random variable [10]. Consequently, it will be considered that a fuzzy valued dataset is a sample of a fuzzy random variable, as defined in [7], whose α-cuts are random sets. We will extend first the definition of modeling problem to the interval case, and then apply the results to all cuts of the fuzzy random variable sample.

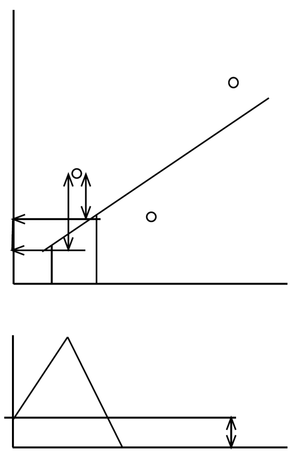

Figure 2: The value α shows the level cut of the output of the fuzzy model for which all points in the sample are covered. This value, combined with the specificity of the fuzzy output, may be used to find the most regular models that are not further than a certain distance of all the elements in the sample; the use of fuzzy fitness could be used to relate fuzzy models and epsilon-insensitive SVMs.

3.1 Interval-valued data

Recall eq. 2: to succeed, a learning algorithm should be able to estimate the values $E[Y|X = \mathbf{x}]$ from a sample of measures taken from a subset of Ω. Suppose that we are given samples from two random sets Γ and Γ_Y, that model the imprecise observations of X and Y,

$$X(\omega) \in \Gamma(\omega) \quad \omega \in \Omega \tag{3}$$

$$Y(\omega) \in \Gamma_Y(\omega) \quad \omega \in \Omega \tag{4}$$

and need to define the conditional expectation $E[Y|X]$ of the underlying, imprecisely observed random variables.

Let us suppose that Y is a discrete random variable, $Y(\omega) \in \{y_1, y_2, \ldots, y_{n_y}\}$. Then,

$$g(\mathbf{x}) = \int y\, f(y|\mathbf{x})\, dy = \sum_{i=1}^{n_y} y_i\, P(y_i|\mathbf{x}). \tag{5}$$

We need to estimate the values $P(y_i|X = \mathbf{x})$. For a given small value $h > 0$, we can try to give a couple of upper and lower bounds for the value $P(y_i|X \in (x-h, x+h))$. Following [1], the limit when h tends to 0 of these quantities is the value we need, $P(y_i|X = x)$. Applying the definition of conditional probability, we have that $P(y_i|X \in (x-h, x+h)) =$

$$\frac{P(\{y_i\} \cap \{\omega \in \Omega \mid X(\omega) \in (x-h, x+h)\})}{P(\{\omega \in \Omega \mid X(\omega) \in (x-h, x+h)\})}. \tag{6}$$

The bounds of $P(\{y_i\} \cap \{\omega \in \Omega \mid X(\omega) \in (x-h, x+h)\})$ are $\underline{P}_i(x, h) =$

$$= P(\{y_i\} \cap \{\omega \in \Omega \mid \Gamma(\omega) \subseteq (x-h, x+h)\})$$

and $\overline{P}_i(x, h) =$

$$= P(\{y_i\} \cap \{\omega \in \Omega \mid \Gamma(\omega) \cap (x-h, x+h) \neq \emptyset\})$$

and the bounds of $P(\{\omega \in \Omega \mid X(\omega) \in (x-h, x+h)\})$ are $\underline{P}(x, h) =$

$$= P(\{\omega \in \Omega \mid \Gamma(\omega) \subseteq (x-h, x+h)\})$$

and $\overline{P}(x, h) =$

$$= P(\{\omega \in \Omega \mid \Gamma(\omega) \cap (x-h, x+h) \neq \emptyset\}).$$

thus we can know that the conditional expectation $E[Y|X]$ is contained in the interval defined as follows (observe that the denominator of Eq. (6) does not depend on i)

$$[\underline{y}, \overline{y}](\Gamma, h) = \frac{\bigoplus_{i=1}^{n_y} y_i\, [\underline{P}_i(x, h), \overline{P}_i(x, h)]}{[\underline{P}(x, h), \overline{P}(x, h)]} \tag{7}$$

where $[a, b] \oplus [c, d] = \{u + v \mid u \in [a, b], v \in [c, d]\}$, and the quotient must be understood as an interval valued operation, $[a, b]/[c, d] = \{u/v \mid u \in [a, b], v \in [c, d]\}$.

In words, when the model was fed with a real input $\mathbf{x}$, its output was $g(\mathbf{x})$. Now we have fed the model with an interval Γ, and we knew that $\mathbf{x}$ was contained in Γ. Its output has been the interval $[\underline{y}, \overline{y}]$, which has been constructed to contain $g(\mathbf{x})$.

Given that $[\underline{y}, \overline{y}]$ is a set valued function, the average error of the model is not longer known (or, alternatively, we could say that the average error is a set valued statistic.) Anyway, we can find upper and lower bounds for it. Let $\mathrm{SQ}([a, b]) = \{t^2 \mid t \in [a, b]\}$ and

$[a, b] \ominus [c, d] = \{u - v \mid u \in [a, b], v \in [c, d]\}$ be the interval-valued square and substraction functions, respectively. Then, the error is contained in the interval

$$[\underline{\text{cost}}, \overline{\text{cost}}] = \text{SQ}([\underline{y}, \overline{y}](\Gamma, h) \ominus \Gamma_Y). \tag{8}$$

For example, let $[\underline{y}, \overline{y}] = [0, 4]$ and $\Gamma_Y = [2, 3]$. Then, $[\underline{\text{cost}}, \overline{\text{cost}}] = \text{SQ}([0, 4] \ominus [2, 3]) = \text{SQ}([-3, 2]) = [0, 9]$. This means that the squared error between two unknown points, if one of then is contained in $[0, 4]$ and the other one in $[2, 3]$, can be any value between 0 and 9, and we can not give a more precise indication without further assumptions.

Therefore, from Eq. (8) we can conclude that the average squared error of the model over the whole population is contained in the interval

$$\text{err}([\underline{y}, \overline{y}]) = \left[\int_\Omega \underline{\text{cost}}(\omega)\, \text{dP}, \int_\Omega \overline{\text{cost}}(\omega)\, \text{dP}\right] \tag{9}$$

3.2 Fuzzy data

If we are given a fuzzy dataset, both the output of the classifier and its expected error will be fuzzy sets, as we show in this section.

Fuzzy datasets can be regarded as samples of a fuzzy random variable $\widetilde{X} \times \widetilde{Y}$. Every instance of the variable combines two types of noise: random noise, originated in the selection of the object ("we choose a person at random") and observation error, originated in an imprecise measure ("the weight of the person is high, where 'high' is one of the values of the linguistic variable 'weight"').

The α-cuts $\widetilde{X}_\alpha$ and $\widetilde{Y}_\alpha$ are random sets (for example, the 0.5-cut of the value 'high' can be the interval [80,110]). Therefore, for every value of α we can build an interval model, as shown in the preceding section, whose output is an interval of values ("if the weight is [80,110], then the percentage of body fat is between 20 and 30"). It is intuitive to conclude that the output of the model, if presented a fuzzy input, will be a fuzzy set defined over the set of outputs ("if the weight is high, and height is low, then the body fat is high".) The same can be said about the average error of the fuzzy model; it will be a fuzzy set.

To obtain this last value, it suffices to admit that the best description we can make about the probability $P(\{y_i\} \cap \{\omega \in \Omega \mid X(\omega) \in (x-h, x+h)\})$, given that the original random variable X is contained in the fuzzy random variable $\widetilde{X}$, is a fuzzy set $\widetilde{P}_i$, whose α-cuts are intervals $[\underline{P}_i^\alpha, \overline{P}_i^\alpha]$, where $\underline{P}_i^\alpha =$

$$= P(\{y_i\} \cap \{\omega \in \Omega \mid [\widetilde{X}(\omega)]_\alpha \subseteq (x-h, x+h)\})$$

and $\overline{P}_i^\alpha$, $\underline{P}^\alpha$ and $\overline{P}^\alpha$ are defined similarly, as we did in the preceding section.

Therefore, the fuzzy output of the model will be the set $\widetilde{y}(\widetilde{X}, h)$, defined by its α-cuts:

$$[\widetilde{y}(\widetilde{X}, h)]_\alpha = \frac{\bigoplus_{i=1}^{n_y} y_i\, [\underline{P}_i^\alpha(x, h), \overline{P}_i^\alpha(x, h)]}{[\underline{P}^\alpha(x, h), \overline{P}^\alpha(x, h)]} \tag{10}$$

and its average error is another fuzzy set,

$$\widetilde{\text{err}}_\alpha = \left[\int_\Omega \underline{\text{cost}}_\alpha(\omega)\, \text{dP}, \int_\Omega \overline{\text{cost}}_\alpha(\omega)\, \text{dP}\right] \tag{11}$$

where $[\underline{\text{cost}_\alpha}, \overline{\text{cost}_\alpha}](\omega) =$

$$= \text{SQ}([\widetilde{y}(\widetilde{X}(\omega), h)]_\alpha \ominus [\widetilde{Y}(\omega)]_\alpha). \tag{12}$$

or, alternatively, we can define a f.r.v.

$$\widetilde{\text{cost}}(\omega) = \text{SQ}(\widetilde{y}(\widetilde{X}(\omega), h) \ominus \widetilde{Y}(\omega)) \tag{13}$$

where SQ and $\ominus$ are the standard square and substration fuzzy arithmetic operators [9], and define the averaged error of the fuzzy model as the expectation of this variable.

Lastly, if we are given a sample $(w_1, \ldots, w_n)$ of Ω, the sample mean of $\widetilde{\text{cost}}$ is the fuzzy set

$$\frac{1}{n} \bigoplus_{i=1}^{n} \text{SQ}(\widetilde{y}(\widetilde{X}(\omega_i), h) \ominus \widetilde{Y}(\omega_i)) \tag{14}$$

and, given the definition of the cost function, it can be proved that fulfills the properties required in [17] and that this last value converges in distribution to $\widetilde{\text{err}} = E[\widetilde{\text{cost}}]$.

We propose to use Eq. (14) as the fuzzy-valued fitness function in genetics-based fuzzy models. Contrary to the fitness that was developed to learn fuzzy classifiers [15], the fuzzy fitness of a fuzzy model has a very intuitive

	weight	fuzzy weight	fat	fuzzy fat
1	76	(74,75,76)	17	(12,17,19)
2	88	(87,88,89)	24	(21,24,27)
3	82	(81,82,83)	20	(18,20,22)
4	80	(79,80,81)	21	(18,21,22)
5	72	(71,72,73)	12	(11,12,13)

Table 1: Dataset for the example problem 'body fat'

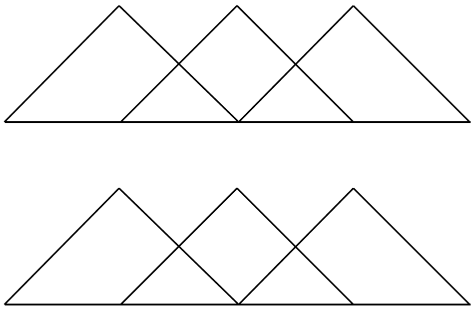

Figure 3: Definition of the linguistic variables "weight" (upper part) and "body fat" (lower part), as used in the example problem.

meaning: it consists in replacing the operations that are needed to evaluate a crisp model by their fuzzy arithmetic-based counterparts. In the next section, this expression will be made clear by means of a numerical example.

4 Example of fitness evaluation in the fuzzy model

Let us suppose that we have to guess the percentage of body fat, given the weigth of a person. To design the model, we are given a sample comprising five people, whose weights and percentages are given in table 1. Weights are triangular fuzzy numbers, designated by three numbers: leftmost, center and rightmost values.

Let us also suppose that the GFS has to evaluate the fitness of the rule base that follows:

```
if weight is small then fat is small
if weight is medium then fat is medium
if weight is high then fat is high
```

where the linguistic variable "weight" takes the values shown in figure 3. We wish to assign a fitness value to this rule base, given the mentioned dataset.

Let us evaluate first this model over the crisp dataset given by the column "weight" in table 1. We have used COG defuzzification (the output of the model is computed as a weigthed sum of the output of each rule, where the weights are the areas of the truncated memberships of the output.) The results are as follows:

X	Y	g	Cost
75	17	16.25	0.56
88	24	25.45	2.12
82	20	22.05	4.18
80	21	20	1
72	12	14.55	6.48

therefore, the cost of this model is $(0.56 + 2.12 + 4.18 + 1 + 6.48)/5 = 2.86$.

If we apply an interval input to the same model (the support of the fuzzy examples,) its output is an interval of values. Observe that, since the rule base in this example defines a monotonic continuous mapping, we just need to compute the output at the boundaries of the intervals, but this may not be true with a different rule base. The interval outputs and costs are as follows:

Sample Input	Sample Output	Model Output	Cost
$[74, 76]$	[12,19]	[15.74, 16.76]	[0,22.62]
$[87, 89]$	[21,27]	[24.81, 26.29]	[0,28.01]
$[81, 83]$	[18,22]	[21.21, 22.69]	[0,22.03]
$[79, 81]$	[18,22]	[18.79, 21.21]	[0,10.29]
$[71, 73]$	[11,13]	[13.71, 15.19]	[0.5,17.59]

therefore the cost is contained in the interval $[0.10, 20.11]$ (i.e., when data is precisely measured, we estimated that the mean squared error was 2.86. With interval-valued data, all we can say without assuming a random distribution of the observation error is that the error is contained in the interval $[0.10, 20.11]$.)

Finally, if the model is applied a fuzzy input, its outputs and costs are shown in table

Sample Input	Sample Output	Model Output	Cost
$(74, 75, 76)$	(12,17,19)	(15.74, 16.25, 16.76)	(0, 0.56, 22.62)
$(87, 88, 89)$	(21,24,27)	(24.81, 25.45, 26.29)	(0, 2.12, 28.01)
$(81, 82, 83)$	(18,20,22)	(21.21, 22.05, 22.69)	(0, 4.18, 22.03)
$(79, 80, 81)$	(18,21,22)	(18.79, 20, 21.21)	(0, 1, 10.29)
$(71, 72, 73)$	(11,12,13)	(13.71, 14.55, 15.19)	(0.5, 6.48, 17.59)

Table 2: Output of the example model when the input is a fuzzy set. The triplets (a, b, c) represent the lower limit, mode and upper limit of the corresponding fuzzy numbers. The error of the model is $(0.10, 2.86, 20.11)$, and its membership function is plotted in Figure 4.

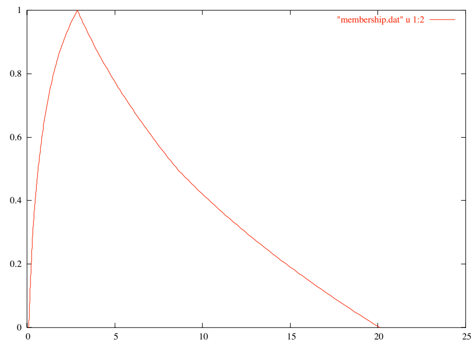

Figure 4: Fuzzy membership assigned to the example model.

2. Observe that neither the output, the costs nor the average error are triangular fuzzy numbers; the fuzzy valued fitness of this rule base is plotted in figure 4. As we mentioned in the introduction, this fuzzy fitness carries information about the punctual error of the model (the mode), the mean slope of the model (its nonspecificity) and the degree to which the examples are covered, as shown in Figure 2 (which is 0 in this case, since there is one example which does not intersect with the support of the the output of the model.) The relative importances that are given to this factor, by means of a fuzzy ranking or a multicriteria algorithm, will determine the kind of model that will be obtained.

5 Concluding remarks and open problems

This paper is a followup of [15]. In that work, we proposed a fuzzy fitness function that can be used with genetic fuzzy classifiers. Now we have completed the study with this paper, where an equivalent fitness function has been derived for Genetic Fuzzy Models. Either in stochastic classifiers or models, when data is vague, it is needed to introduce additional hypotheses, as a probability distribution (uniform, gaussian, etc.) over the measurement errors. Fuzzy algorithms do not need this extra information. Therefore, for this kind of problems, GFSs are inherently better than their stochastic counterparts. Obviously, this is not longer true if data are "defuzzified" before they are fed to the learning algorithm. In this last case, since the optimal decisions are the Bayes classifier or the conditional expectation, we can not expect GFS to outperform statistical methods, and the benefits of the fuzzy approach are restricted to the field of linguistic understandability.

Acknowledgment

The authors wish to thank Frank Hoffmann his comments about the former study about the classification problem, and for asking the question that inspired this paper.

This work was supported by the Spanish Ministry of Education and Science, under grants TIC2002-04036-C05-05 and MTM2004-01269.

References

[1] R. B. Ash. *Basic Probability Theory.* Wiley. 1970.

[2] C. Bertoluzza, M. A. Gil, D.A. Ralescu, Eds. *Statistical Modeling, Analysis and Management of Fuzzy Data* Series in Fuzziness and Soft Computing, 87, Springer. 2003

[3] Birkes, D., Dofge, Y. Alternative methods of regression. Wiley. 1993.

[4] O. Cordon , F. Herrera, F. Hoffmann, L. Magdalena *Genetic Fuzzy Systems, Evolutionary Tuning and Learning of Fuzzy Knowledge Bases.* World Scientific, 2001

[5] O. Cordon, F. A. C. Gomide, F. Herrera, F. Hoffmann, L. Magdalena. "Genetic fuzzy systems. New developments". Fuzzy Sets and Systems 141, 1, 1 pp 1-3. 2004.

[6] Cristiniani, N., Shawe-Taylor, J. An Introduction to Support Vector Machines, Cambridge University Press. 2000

[7] F. Feron. "Ensembles aleatoires flous", C.R. Acad. Sci. Paris Ser. A 282 pp. 903-906. 1975.

[8] Koenker, R., Hallock, K. Quantile Regression. Journal of Economic Perspectives 15, 143-156. 2001

[9] Kaufmann, A., Gupta, M. M. Introduction to Fuzzy Arithmetic: theory and applications. Van Nostrand Reinhold, 1985.

[10] R. Kruse, K.D. Meyer. *Statistics with Vague Data* Vol. 33. Reidel, Dordrecht, 1987.

[11] Sánchez, L. A random sets-based method for identifying fuzzy models. Fuzzy Sets and Systems 98 (3) 343-354. 1998

[12] Sánchez, L. Interval-valued GA-P algorithms. IEEE Trans. Evolutionary Computation 4(1): 64-72 (2000)

[13] Sánchez, L., Couso, I. Fuzzy Random Variables-based modeling with GA-P algorithms. In *Information, Uncertainty and Fusion.* Yager, R., Bouchon-Menier, B. (eds). Kluwer. 245-256 (2000)

[14] Sánchez, L., Casillas, J., Cordón, O., Del Jesus, M. J. Some relationships between fuzzy and random classifiers and models. International Journal on Approximate Reasoning. 29. 175-213 (2001)

[15] Sánchez, L., Couso, I. "Advocating the use of Imprecisely Observed Data in Genetic Fuzzy Systems". GFS 2005.

[16] Tanaka, H., Uejima, S. and Asai, K. Linear Regression analysis with fuzzy model. IEEE Trans. SMC 12 (6), 903-907. 1982

[17] H. C. Wu. The central limit theorems for fuzzy random variables. Information Sciences. 120, pp. 239-256. 1999