Granular modelling through regression analysis

Andrzej Bargiela

School of Computing and Informatics,
The Nottingham Trent Univ.
Nottingham NG1 4BU, U.K.
andre@doc.ntu.ac.uk

Abstract

Granular representation of experimental data D is defined as a transformation $A \rightarrow \mathcal{G}(Y)$ where A are information granules in the space of input data D and $\mathcal{G}(Y)$ is a family of information granules in the output space. In this paper, we propose an alternative to the granular representation method based on the use of possibility and necessity measure and the solution of fuzzy relational equations. The proposed method is an iterative multiple regression with fuzzy independent and dependent variables. The regression problem is posed as a gradient descent optimisation which provides both computational efficiency and generality.

Keywords: multiple regression, fuzzy data, gradient descent

1. Introductory comments

Granular modelling has emerged as one of the fundamental concerns underpinning granular computing [1-2], [12-13] that permeates a broad array of pursuits of computational intelligence. With the notion of information granules and the granularity being central to various aspects of problem abstraction and modularisation, we encounter different formal ways of expressing information granularity. While the notion of granularity is usually expressed in the language of sets (fuzzy sets, rough sets, etc.) the fundamental construct in building information granules is that of mapping. The product of granular mapping is commonly referred to as a granular model of a system. In a descriptive manner, by granular mapping we mean a transformation from some input space to output space that is characterised at the granular level; this means that it operates on information granules defined in the corresponding spaces. There are two fundamental categories of problems arising in such setting, that is: 1) analysis of granular mappings which is inherently associated with data interpretation aspects (such as in rule-based systems) and 2) design of these mappings which requires the development of the experimentally meaningful and transparent associations between the information granules.

In this paper we focus on the second aspect of granular mapping and provide an alternative to our earlier approach, based on the use of possibility and necessity measure and the solution of fuzzy relational equations [3]. The approach advocated here is that of regression analysis.

Regression analysis is one of the basic tools of scientific investigation enabling identification of functional relationship between independent and dependent variables. In the classical regression analysis both the independent and dependent variables are given as real numbers. However, in many real-life situations, where the complexity of the physical system dictates adoption of a more general viewpoint, regression variables are given as non-numerical entities such as linguistic variables. Unfortunately, such real-life situations are outside the scope of the classical regression analysis.

Following the introduction of the concept of fuzzy sets by Zadeh in 1965 [11-13] various researchers attempted extending the regression analysis from crisp to fuzzy domain. One of the first results concerning a more general form of regression analysis was contributed by Tanaka et al. [10] in which they consider crisp independent and fuzzy dependent variables. A further

generalisation of the regression model was introduced by Diamond [4, 5] who considered both independent and dependent variables as triangular fuzzy numbers. Diamond's approach has been subsequently used to develop fuzzy regression model with regression variables expressed as arbitrary fuzzy numbers, Grzegorzewski and Mrowka [6] and others [1, 7, 9]. Another generalisation of the regression model, involving the use of fuzzy random variables was suggested by Korner and Nather [8]. However, in all of the above approaches the analytical formulae quantifying the values of the parameters of the regression model have been derived only for the case of a simple linear regression, i.e. for the single independent single dependent variable system.

In this paper we re-formulate the regression problem and express it as a gradient descent optimisation. In doing so we generalise the simple regression model to multiple regression and lay foundation for a further generalisation to multiple non-linear regression with fuzzy variables.

In Section 2 we provide a boackground discussion of the classical regression analysis, fuzzy numbers and fuzzy simple linear regression. In Section 3 we extend the scope of fuzzy regression to multiple variables and provide a gradient descent optimisation algorithm that provides a practical way of calculating regression coefficients.

2. Background discussion

A. Classical regression analysis

The general task of regression analysis is defined as identification of a functional relationship between the independent variables $\mathbf{x}=[x_1, x_2, \ldots, x_n]$ and dependent variables $\mathbf{y}=[y_1, y_2, \ldots, y_m]$, where *n* is a number of independent variables in each observation and *m* is a number of dependent variables. The regression model is expressed in this case as

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon} \tag{1}$$

where, $\mathbf{f}(\mathbf{x})$ is a vector function $[f_1(\mathbf{x}),..., f_m(\mathbf{x})]$ and $\boldsymbol{\varepsilon} = [\varepsilon_1,..., \varepsilon_m]$ is a vector of random error of functional approximation. The general model (1) is frequently simplified by assuming a linear relationship between the independent and dependent variables, thus reducing the task of identification of a functional relationship $\mathbf{f}()$ to an identification of parameters of a linear function. Furthermore, the multiple dependent variables $\mathbf{y}$ are considered separately to give m independent regression models. In the simplest case of a single independent and single dependent variable the regression model is given as

$$y = a_0 + a_1 x + \varepsilon \tag{2}$$

The above is frequently referred to as a *simple linear regression model*.

Considering *k* pairs of observations of independent and dependent variables $(x^1, y^1), \ldots,(x^k, y^k)$ and assuming the least squares criterion *C*(.) for the minimisation of the discrepancies between the model and the actual observations

$$C(a_{0,} a_1) = \sum_{i=1}^{k} (y^i - (a_0 + a_1 x^i))^2 \tag{3}$$

one can derive optimal estimators for the parameters of the linear model (2) by solving $\frac{\partial C}{\partial a_0} = 0$ and $\frac{\partial C}{\partial a_1} = 0$ which gives

$$\hat{a}_0 = \frac{1}{k}\sum_{i=1}^{k} y^i - \hat{a}_1 \frac{1}{k}\sum_{i=1}^{k} x^i \tag{4}$$

$$\hat{a}_1 = \frac{\sum_{i=1}^{k} (y^i - \frac{1}{k}\sum_{i=1}^{k} y^i)(x^i - \frac{1}{k}\sum_{i=1}^{k} x^i)}{\sum_{i=1}^{k} (x^i - \frac{1}{k}\sum_{i=1}^{k} x^i)^2} \tag{5}$$

B. *Fuzzy simp-le linear regression*

In order to generalise the simple linear regression to the case of imprecise independent and dependent variables we follow the approach proposed by Diamond [5] and adopt a subfamily of fuzzy sets, called fuzzy numbers, as a formal framework for the representation of imprecise data. A fuzzy number can be formally defined as follows:

Definition 1 *A fuzzy subset A of the set of real numbers R with membership function* $\mu_A : R \to [0,1]$ *is called a fuzzy number if*

i) *A is normal, i.e. there exist an element* z_0 *such that* $\mu_A(z_0) = 1$

ii) *A is fuzzy convex, i.e.* $\forall z_1, z_2 \in R$ $\mu_A(\lambda z_1 + (1+\lambda) z_2) \geq \mu_A(z_1) \wedge \mu_A(z_2)$, $\forall \lambda \in [0,1]$;

iii) μ_A *is upper semi continuous;*

iv) $\sup(A) = \overline{\{z \in R : \mu_A(z) > 0\}}$ *is bounded.*

A fuzzy number A can be represented as a family of nonfuzzy sets called α-cuts, A_a defined as

$$A_\alpha = \{z \in R : \mu_A(z) \geq \alpha\} \tag{6}$$

giving a set representation

$$A = \{A_\alpha : \alpha \in (0,1]\} \tag{7}$$

Based on the resolution identity we get

$$\mu_A(z) = \sup\{\alpha I_{A_\alpha}(z) : \alpha \in (0,1]\} \tag{8}$$

where $I_{A_\alpha}(z)$ represents the characteristic function of A_a From the definition of the fuzzy number it is easily seen that every α-cut of a fuzzy number A is a closed interval $A_\alpha = [A^L(\alpha), A^U(\alpha)]$ where

$$A^L(\alpha) = \inf\{z \in R : \mu_A(z) \geq \alpha\} \tag{9}$$

$$A^U(\alpha) = \sup\{z \in R : \mu_A(z) \geq \alpha\} \tag{10}$$

Consequently for two fuzzy numbers A and B with α-cuts $A_\alpha = [A^L(\alpha), A^U(\alpha)]$ and $B_\alpha = [B^L(\alpha), B^U(\alpha)]$ we can define a distance between A and B as

$$d(A,B) = \sqrt{\int_0^1 (A^L(\alpha) - B^L(\alpha))^2 d\alpha + \int_0^1 (A^U(\alpha) - B^U(\alpha))^2 d\alpha} \tag{11}$$

Using the formalism of fuzzy numbers we can express the fuzzy simple linear regression problem as a problem of identification of parameters $b_0, b_1 \in R$ of a fuzzy linear model

$$Y = b_0 + b_1 X \tag{12}$$

The parameters b_0, b_1 are evaluated by minimizing the error measured as a distance between the actual observations and the estimates evaluated from the model

$$\min H(b_0, b_1) = \sum_{i=1}^{k} d^2(Y_i, b_0 + b_1 X_i) \tag{13}$$

It must be noted however that the exact form of the error function *H(.)* depends on the sign of the parameter b_1. If b_1>0 then

$$H^+(b_0, b_1) = \sum_{i=1}^{k} \int_0^1 (Y_i^L(\alpha) - b_0 - b_1 X_i^L(\alpha))^2 d\alpha + \sum_{i=1}^{k} \int_0^1 (Y_i^U(\alpha) - b_0 - b_1 X_i^U(\alpha))^2 d\alpha \tag{14}$$

and if b_1<0 then

$$H^-(b_0, b_1) = \sum_{i=1}^{k} \int_0^1 (Y_i^L(\alpha) - b_0 - b_1 X_i^U(\alpha))^2 d\alpha + \sum_{i=1}^{k} \int_0^1 (Y_i^U(\alpha) - b_0 - b_1 X_i^L(\alpha))^2 d\alpha \tag{15}$$

Having pre-determined the sign of the parameter b_1 we can calculate exact numerical values of b_0 and b_1 by solving either

$$\frac{\partial H^+(b_0, b_1)}{\partial b_0} = 0 \text{ and } \frac{\partial H^+(b_0, b_1)}{\partial b_0} = 0 \tag{16}$$

or

$$\frac{\partial H^-(b_0, b_1)}{\partial b_0} = 0 \text{ and } \frac{\partial H^-(b_0, b_1)}{\partial b_0} = 0 \tag{17}$$

In the first case we obtain

$$\hat{b}_0^+ = \tilde{Y} - \hat{b}_1^+ \tilde{X} \tag{18}$$

$$\hat{b}_1^+ = \frac{SS_{xy}^+}{SS_{xx}} \tag{19}$$

where

$$\widetilde{X} = \int_0^1 \frac{\overline{X}^L(\alpha) + \overline{X}^U(\alpha)}{2} d\alpha \quad (20)$$

$$\widetilde{Y} = \int_0^1 \frac{\overline{Y}^L(\alpha) + \overline{Y}^U(\alpha)}{2} d\alpha \quad (21)$$

$$SS_{xx} = \sum_{i=1}^{k} \int_0^1 ((X_i^L(\alpha))^2 + (X_i^U(\alpha))^2) d\alpha - 2k\widetilde{X}^2 \quad (22)$$

$$SS_{xy}^{+} = \sum_{i=1}^{k} \int_0^1 (X_i^L(\alpha) Y_i^L(\alpha) + X_i^U(\alpha) Y_i^U(\alpha)) d\alpha - 2k\widetilde{X}\widetilde{Y} \quad (23)$$

and in the second case the regression parameters are evaluated as

$$\hat{b}_0^- = \widetilde{Y} - \hat{b}_1^- \widetilde{X} \quad (24)$$

$$\hat{b}_1^- = \frac{SS_{xy}^-}{SS_{xx}} \quad (25)$$

where

$$SS_{xy}^- = \sum_{i=1}^{k} (X_i^U(\alpha) Y_i^L(\alpha) + X_i^L(\alpha) Y_i^U(\alpha)) d\alpha - 2k\widetilde{X}\widetilde{Y} \quad (26)$$

and $\widetilde{X}$, $\widetilde{Y}$, SS_{xx} are evaluated as in (20), (21), (22) respectively.

C. Gradient descent optimisation

Although with the quantification of the regression error, (14) and (15), it is possible to find analytical solution to equations (16) and (17), in the case of a large number of regression variables such a direct solution becomes rather complex. As an alternative approach one can perform iterative refinement of initial gueses of the regression model parameters b_0 and b_1 taking the partial derivatives of (14) or (15) as indicators of the local gradient of the functional *H*. The computational advantage of this approach is rooted in the fact that the calculation of the values of partial derivatives of *H* is much simpler than solving systems of simultaneous equations such as (16) or (17). Indeed, the approach can be easily applied even if the regression model is non-linear. For the simple regression model given by (12) the gradient descent optimisation can be summarised as:

[Gradient descent algorithm for simple fuzzy regression]

a) Make an initial guess of b_0 and b_1 say b_0^0 and b_1^0;

b) Set the iteration counter i=1;

c) Evaluate gradient *H* with respect of regression model parameters as per equation (14) or (15);

d) Calculate the value of the parameter update

$$\Delta b_0 = \mu_0 \frac{\partial H^+(b_0, b_1)}{\partial b_0} \text{ and } \Delta b_1 = \mu_1 \frac{\partial H^+(b_0, b_1)}{\partial b_1} \quad (27)$$

or

$$\Delta b_0 = \mu_0 \frac{\partial H^-(b_0, b_1)}{\partial b_0} \text{ and } \Delta b_1 = \mu_1 \frac{\partial H^-(b_0, b_1)}{\partial b_1} \quad (28)$$

e) Update parameter estimates

$$b_0^i = b_0^{i-1} + \Delta b_0 \text{ and } b_1^i = b_1^{i-1} + \Delta b_1 \quad (29)$$

f) If $\Delta b_0 > \varepsilon$ or $\Delta b_1 > \varepsilon$ then update iteration counter i=i+1 and repeat from c); otherwise stop.

In order to converge to the solution the algorithm requires appropriate selection of parameters μ_0 and μ_1 and the selection of the termination criteria ε. However, even with a heuristic selection of these parameters (which may require repeated runs of the algorithm) the evaluation of the regression model parameters remains computationally very appealing.

3. Multiple fuzzy linear regression

The fuzzy simple linear regression model (12) can now be extended to a fuzzy model with multiple independent variables

$$Y = b_0 + b_1 X^1 + b_2 X^2 + \ldots + b_m X^m \quad (30)$$

where $Y, X^1, X^2, ..., X^m$ are all fuzzy numbers defined on R and $b_0, b_1, b_2, ..., b_m$ are real numbers. The parameters $b_0, b_1, b_2, ..., b_m$ are evaluated by minimising the cost function *H(.)* defined as a squared distance between the fuzzy observations and the corresponding fuzzy dependent variable *Y* evaluated from (30).

$$\min H(b_0, b_1, ..., b_m) = \sum_{i=1}^{k} d^2(Y_i, b_0 + b_1 X_i^1 + b_2 X_i^2 + ... + b_m X_i^m) \quad (31)$$

with a distance function *d(.)* defined as in (11).

Using the α-cut representation of fuzzy numbers it is necessary to ensure that the minimum and maximum value of α-cut intervals are properly matched for both positive and negative values of model parameters $b_0, b_1, b_2, ..., b_m$. We can formalise this requirement by introducing the following substitution of variables

$$\hat{X}_i^{jL}(\alpha) = X_i^{jL} \text{ and } \hat{X}_i^{jU}(\alpha) = X_i^{jU} \quad \text{if } b_j \geq 0 \quad (32)$$

$$\hat{X}_i^{jL}(\alpha) = X_i^{jU} \text{ and } \hat{X}_i^{jU}(\alpha) = X_i^{jL} \quad \text{if } b_j < 0 \quad (33)$$

where *L* and *U* denote the corresponding lower and upper bounds of the α-cut intervals and *j=1,...,m*. With these substitutions the cost function *H(.)* can be written explicitly as

$$\hat{H}(b_0, b_1, ..., b_m) = \sum_{i=1}^{k} \int_0^1 (Y_i^L(\alpha) - b_0 - b_1 \hat{X}_i^{1L}(\alpha) - ... - b_m \hat{X}_i^{mL}(\alpha))^2 d\alpha + \sum_{i=1}^{k} \int_0^1 (Y_i^U(\alpha) - b_0 - b_1 \hat{X}_i^{1U}(\alpha) - ... - b_m \hat{X}_i^{mU}(\alpha))^2 d\alpha \quad (34)$$

Using the expression (34) we can calculate gradients of the cost function $\hat{H}(.)$ with respect of the regression parameters as

$$\frac{\partial \hat{H}(b_0, ..., b_m)}{\partial b_0} = -2\tilde{Y} + 4kb_0 + 2b_1 \tilde{\hat{X}}^1 + ... + 2b_m \tilde{\hat{X}}^m \quad (35)$$

and

$$\frac{\partial \hat{H}(b_0, ..., b_m)}{\partial b_j} = -2\overleftrightarrow{SS}_{X^j Y} + 2b_0 \tilde{\hat{X}}^j + 2b_1 \overleftrightarrow{SS}_{X^j X^1} + ... + 2b_m \overleftrightarrow{SS}_{X^j X^m} \quad (36)$$

where

$$\tilde{Y} = \int_0^1 (\sum_{i=1}^{k} Y_i^L(\alpha) + \sum_{i=1}^{k} Y_i^U(\alpha)) d\alpha \quad (37)$$

$$\tilde{\hat{X}}^j = \int_0^1 (\sum_{i=1}^{k} \hat{X}_i^{jL}(\alpha) + \sum_{i=1}^{k} \hat{X}_i^{jU}(\alpha)) da \quad j=1,..,m \quad (38)$$

$$\overleftrightarrow{SS}_{X^j Y} = \int_0^1 \sum_{i=1}^{k} (Y_i^L(\alpha) \hat{X}_i^{jL}(\alpha) + Y_i^U(\alpha) \hat{X}_i^{jU}(\alpha)) d\alpha \quad j = 1,..., m \quad (39)$$

$$\overleftrightarrow{SS}_{X^j X^p} = \int_0^1 \sum_{i=1}^{k} (\hat{X}_i^{pL}(\alpha) \hat{X}_i^{jL}(\alpha) + \hat{X}_i^{pU}(\alpha) \hat{X}_i^{jU}(\alpha)) d\alpha \quad j, p = 1,..., m \quad (40)$$

With the above equations (30)-(38) we can now define the gradient descent algorithm for multiple fuzzy regression.

[Gradient descent algorithm for multiple fuzzy regression]

a) Make an initial guess of $b_0, b_1, ..., b_m$: $b_0^0, b_1^0, ..., b_m^0$;

b) Set the iteration counter i=1;

c) Evaluate the α-cut intervals for individual regression variables taking into account the sign of the corresponding regression variable; (32)-(33);

d) Evaluate gradient of $\widehat{H}(.)$ with respect of regression model parameters as per equation (35) or (36);

e) Calculate the value of the regression parameters update

$$\Delta b_0 = \mu_0 \frac{\partial \widehat{H}(b_0,...,b_m)}{\partial b_0} \quad (41)$$

and

$$\Delta b_j = \mu_j \frac{\partial \widehat{H}(b_0,...,b_m)}{\partial b_j} \quad j=1,\ldots,m \quad (42)$$

where μ_j are the parameters controlling the convergence of the gradient descent optimisation.

f) Update parameter estimates

$$b_j^i = b_j^{i-1} - \Delta b_j \quad j=0,\ldots,m \quad (43)$$

g) If $\exists_{j=0,1,...,m} \Delta b_j > \varepsilon$ then update iteration counter i=i+1 and repeat from c); otherwise stop.

4. Computational complexity

Computational complexity of the proposed fuzzy regression algorithm is a product of the number of iterations and the computational complexity of a single iteration. Since the cost function (34) is essentially a regular quadratic function the convergence properties of the iterative scheme are veru good and the number of iterations does not depend significantly on the number of regression variables. In most cases it is possible to achieve good solution with less than 10 iterations. But, of course, this is conditional on the prior normalisation of the individual regression variables so that the shape of the quadratic form is not unduly distorted. Alternatively, one can require that the individual convergence parameters are set by some heuristic procedure so as to achieve a balanced augmentation of the optimal solution in all dimensions.

The computational complexity of a single iteration is essentially determined by the complexity of evaluation of the gradient in equations (35) and (36). It is clear from (37)-(40) that the calculation of the gradient is proportional to the square of the number of regression variables *m.*

Consequently the whole algorithm can be described as having complexity $O(m^2kc)$, where k is the number of data items processed by the algorithm and c is a constant defined by the convergence accuracy requirement.

The improvement offered by the proposed algorithm can be appreciated by comparing its computational complexity to the complexity of the direct analytical solution. (3)-(5). Since $m=1$ for the simple linear regression model the complexity of the proposed algorithm is $O(kc)$ while the complexity of the direct analytical solution is $O(k^2c)$, a significantly higher value.

For the multiple regression models the advantage of the proposed gradient descent based approach is even more pronounced since the analytical methods are likely to be excessively complex.

Acknowledgements

The support from the Engineering and Physical Sciences Research Council (EPSRC, UK is gratefully acknowledged.

References

[1] Bargiela, A., Pedrycz, W., (2002), *Granular Computing,* Kluwer.

[2] Bargiela, A., Pedrycz, W., (2003), Recursive information granulation, *IEEE Transactions SMC-B,* 33, 1, 96-112.

[3] Bargiela, A., Pedrycz, W., (2005), Granular mappings, *IEEE Trans. SMC,* A, 35, 2, 292-297.

[4] Diamond, P., (1988), Fuzzy least squares, *Inf. Sci.,* 46, 141-157.

[5] Diamond, P., (1992), Least squares and maximum likelihood regression for fuzzy linear models, in.: *Fuzzy Regression Analysis,* Kacprzyk J, Fedrizzi M (eds), 137-151.

[6] Grzegorzewski, P., Mrowka, E., (2004), Regression analysis with fuzzy data, in *Soft Computing: Tools, Techniques and*

Applications, Grzegorzewski, Krawczak, Zadrozny (eds.).

[7] Kacprzyk J., Fedrizzi M., (1992), *Fuzzy regression analysis*, Omnitech Press, Warsaw and Physica-Verlag, Heildelberg.

[8] Korner R., Nather W., (1998), Linear regression with random fuzzy variables, *Inf. Sci.* 109, 95-118.

[9] Pedrycz, W., F. Gomide (1998), *An Introduction to Fuzzy Sets: Analysis and Design*, MIT Press, Cambridge, MA.

[10] Tanaka H., Uegima S., Asai K., (1982), Linear regression analysis with fuzzy model, *IEEE Trans. Systems Man Cybern.,* 12, 903-907.

[11] Zadeh L.A., (1965), Fuzzy sets, *Inform. And Control,* 8, 338-353.

[12] Zadeh, L.A. (1979), Fuzzy sets and information granularity, In: M.M. Gupta, R.K. Ragade, R.R. Yager, eds., *Advances in Fuzzy Set Theory and Applications*, North Holland, Amsterdam, 3-18.

[13] Zadeh, L.A. (1996), Fuzzy logic = computing with words, *IEEE Trans. on Fuzzy Systems*, vol. 4, 2, 103-111.