FUZZY CLASSIFICATION MODELS BASED ON TANAKA'S FUZZY LINEAR REGRESSION APPROACH AND NONPARAMETRIC IMPROVED FUZZY CLASSIFIER FUNCTIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

GİZEM ÖZER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
INDUSTRIAL ENGINEERING

JULY 2009

Approval of the thesis:

FUZZY CLASSIFICATION MODELS BASED ON TANAKA'S FUZZY LINEAR REGRESSION APPROACH AND NONPARAMETRIC IMPROVED FUZZY CLASSII FUNCTIONS

submitted by **GİZEM ÖZER** in partial fulfillment of the requirements for the degree of **Master of Science in Industrial Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** __________

Prof. Dr. Nur Evin Özdemirel
Head of Department, **Industrial Engineering** __________

Prof. Dr. Gülser Köksal
Supervisor, **Industrial Engineering Dept., METU** __________

Assoc. Prof. Dr. İnci Batmaz
Co-Supervisor, **Statistics Department, METU** __________

Examining Committee Members:

Prof. Dr. Murat Köksalan
Industrial Engineering Dept., METU __________

Prof. Dr. Gülser Köksal
Industrial Engineering Dept., METU __________

Assoc. Prof. Dr. İnci Batmaz
Statistics Department, METU __________

Prof. Dr. Levent Kandiller
Industrial Engineering Dept., Çankaya University __________

Assoc. Prof. Dr. Canan Sepil
Industrial Engineering Dept., METU __________

Date: 24.07.2009

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: GİZEM ÖZER

Signature:

ABSTRACT

FUZZY CLASSIFICATION MODELS BASED ON TANAKA'S FUZZY LINEAR REGRESSION APPROACH AND NONPARAMETRIC IMPROVED FUZZY CLASSIFIER FUNCTIONS

Özer, Gizem

M.S., Department of Industrial Engineering

Supervisor : Prof. Dr. Gülser Köksal

Co-Supervisor: Assoc. Prof. Dr. İnci Batmaz

July 2009, 97 pages

In some classification problems where human judgments, qualitative and imprecise data exist, uncertainty comes from fuzziness rather than randomness. Limited number of fuzzy classification approaches is available for use for these classification problems to capture the effect of fuzzy uncertainty imbedded in data. The scope of this study mainly comprises two parts: new fuzzy classification approaches based on Tanaka's Fuzzy Linear Regression (FLR) approach, and an improvement of an existing one, Improved Fuzzy Classifier Functions (IFCF). Tanaka's FLR approach is a well known fuzzy regression technique used for the prediction problems including fuzzy type of uncertainty. In the first part of the study, three alternative approaches are presented, which utilize the FLR approach for a particular customer satisfaction classification problem. A comparison of their performances and their applicability in other cases are discussed. In the second part of the study, the improved IFCF method, Nonparametric Improved Fuzzy Classifier Functions (NIFCF), is presented, which proposes to use a nonparametric method, Multivariate Adaptive Regression

Splines (MARS), in clustering phase of the IFCF method. NIFCF method is applied on three data sets, and compared with Fuzzy Classifier Function (FCF) and Logistic Regression (LR) methods.

Keywords: Fuzziness, Fuzzy Classification, Fuzzy Classifier Function, Improved Fuzzy Classifier Function, Fuzzy Linear Regression, Customer Satisfaction.

ÖZ

TANAKA'NIN BULANIK DOĞRUSAL REGRESYON YAKLAŞIMINA DAYALI BULANIK SINIFLANDIRMA MODELLERİ VE PARAMETRİK OLMAYAN İYİLEŞTİRİLMİŞ BULANIK SINIFLANDIRMA FONKSİYONLARI

Özer, Gizem

Yüksek Lisans, Endüstri Mühendisliği Bölümü

Tez Yöneticisi : Prof. Dr. Gülser Köksal

Ortak Tez Yöneticisi: Doç. Dr. İnci Batmaz

Temmuz 2009, 97 sayfa

İnsan değerlendirmeleri, niteliksel ve kesin olmayan verilerin yer aldığı bazı sınıflandırma problemlerinde, belirsizlik, rastgelelikten ziyade bulanıklıktan kaynaklanmaktadır. Böyle sınıflandırma problemlerinde veri içine gömülmüş bulanık belirsizliğin etkisini yansıtmak için sınırlı sayıda bulanık sınıflandırma yaklaşımı mevcuttur. Bu çalışmanın kapsamı temel olarak iki bölümden oluşmaktadır: Tanaka'nın Bulanık Doğrusal Regresyon (BDR) yaklaşımına dayalı yeni bulanık sınıflandırma yaklaşımları ve var olan İyileştirilmiş Bulanık Sınıflandırma Fonksiyonları (İBSF) yaklaşımının daha da iyileştirilmesi. Tanaka'nın BDR yaklaşımı bulanık yapıda belirsizlik içeren tahmin problemleri için kullanılan tanınmış bir bulanık regresyon yöntemidir. Çalışmanın ilk bölümünde, belirli bir müşteri memnuniyeti sınıflandırma problemi için BDR yaklaşımından yararlanan üç alternatif yaklaşım sunulmuştur. Bu yaklaşımların performanslarının karşılaştırması ve farklı

durumlarda uygulanabilirliği tartışılmıştır. Çalışmanın ikinci bölümünde ise, İBSF yönteminin kümeleme aşamasında, parametrik olmayan bir yöntem olan Çok Değişkenli Uyarlanabilir Regresyon Eğrilerini (ÇDURE) kullanmayı öneren iyilştirilmiş İBSF yöntemi, Parametrik Olmayan İyileştirilmiş Bulanık Sınıflandırma Fonksiyonları (POİBSF) yöntemi sunulmuştur. POİBSF yöntemi üç veri setine uygulanmış ve Bulanık Sınıflandırma Fonksiyonu (BSF) ve Lojistik Regresyon (LR) yöntemleri ile karşılaştırılmıştır.

Anahtar Kelimeler: Bulanıklık, Bulanık Sınıflandırma, Bulanık Sınıflandırma Fonksiyonu, İyileştirilmiş Bulanık Sınıflandırma Fonksiyonu, Bulanık Doğrusal Regresyon, Müşteri Memnuniyeti.

To my family and my fiancé

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor Prof. Dr. Gülser Köksal and my co-supervisor Assoc. Prof. Dr. İnci Batmaz for their invaluable support, insight and encouragement throughout this study.

I would like to thank Dr. Özlem Türker Bayrak for her detailed and constructive support during the study. I also wish to thank Elçin Kartal and Tuna Kılıç not only for their support but also for their friendship and for enjoyable time we spent during the study.

I wish to thank Prof. Dr. Burhan Türkşen for valuable insights he provided in this research.

I would like to present my thanks to all my friends and my colleagues for their encouragement and moral support they provided during my study.

I would also like to acknowledge the Scientific and Technological Research Council of Turkey (TÜBİTAK) for the support they provided during my graduate study.

I would like to thank my grandfather, Salih Özer, my grandmother Gül Alıcı and my aunt, Aysel Alıcı. I always feel their love and belief with me even if they are no longer with us. Every moment that we shared is invaluable for me.

Most special thanks go to my family and my fiancé, Necati Şekkeli for their everlasting love, moral support and valuable guidance. I would like to thank them for always being there whenever I need. I would never succeed without their love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ADI : Alternative Dunn's Index
ANFIS : Adaptive Neuro Fuzzy Inference Systems
AUC : Area Under ROC Curve
CE : Classification Entropy
DI : Dunn's Index
DT : Decision Trees
FC : Fuzzy Classifier
FCF : Fuzzy Classifier Functions
FCM : Fuzzy C-Means
FF : Fuzzy Functions
FLR : Fuzzy Linear Regression
FLSR : Fuzzy Least Squares Regression
FR : Fuzzy Regression
FRB : Fuzzy Rule Base
FRC : Fuzzy Relational Classifier
GCV : Generalized Cross Validation
IFC : Improved Fuzzy Clustering
IFCF : Improved Fuzzy Classifier Functions
IFF : Improved Fuzzy Functions
LR : Logistic Regression
LSR : Least Squares Regression
MARS : Multivariate Adaptive Regression Splines
MCDA : Multicriteria Decision Aid
MCR : Misclassification Rate
NIFC : Nonparametric Improved Fuzzy Clustering

NIFCF : Nonparametric Improved Fuzzy Classifier Functions

NN : Neural Networks

PC : Partition Coefficient

PCC : Percentage of Correctly Classified

S : Separation Index

SC : Partition Index

SVM : Support Vector Machines

XB : Xie and Beni’s Index

CHAPTER 1

INTRODUCTION

For a very limited number of systems, the information content available can be known as certain with no imprecision, no vagueness and so on. Uncertainty can exist in various forms, which may be resulted from ignorance, from lack of knowledge, from various classes of randomness, from inability to perform adequate measurements or from vagueness (Ross, 2004).

Most widely used conceptual basis for handling uncertainty is probability theory, which is concerned with the random type of uncertainty. Roots of probability theory dates back to 16th century when the rules of probability were recognized in the games of chance by Gerolamo Cardano (Ross, 2004). From the late 19th century to the late 20th century, statistical methods based on probability theory dominated the methods formulated for handling uncertainty (Ross, 2004).

Numerous statistical methods for classification exist in the literature such as Logistic Regression (LR) and Discriminant Analysis. These are the methods widely used for modeling the systems where the uncertainty comes from randomness. In these conventional statistical methods, deviations between observed and estimated output variables are supposed to be resulted from sampling errors and measurement errors (non-sampling errors). While sampling errors arise from the use of a sample to estimate a population characteristic instead of using the entire population, measurement errors refer to the errors generally resulted from the manner in which the observations are taken such as inaccurate measurements due to poor processes or imperfect measuring device. Statistical methods handle crisp data, the members of which are just single valued real numbers.

Handling uncertainty using probability theory was challenged by the introduction of the concept of fuzziness by Zadeh in 1965 (Ross, 2004). Zadeh introduced a different kind of uncertainty than randomness, called fuzziness. In systems in which human judgments, qualitative or imprecise data exist, fuzziness is the source of uncertainty rather than randomness. In these systems, deviations are supposed to be due to the indefiniteness of the system structure (Tanaka et al., 1982). Unlike statistical methods, fuzzy methods may work with fuzzy data as well as crisp data. Fuzzy data, the members of which are the fuzzy numbers, can be thought of as interval numbers, values within which have varying degrees of memberships.

In order to model the systems having fuzzy type of uncertainty, several fuzzy classification methods have been developed. These include Fuzzy Classifier Functions (FCF), Improved Fuzzy Classifier Functions (IFCF), Adaptive Neuro Fuzzy Inference Systems (ANFIS), and Fuzzy Relational Classifier (FRC). These methods deal with fuzzy uncertainty imbedded in data in different ways. For example, FRC method deals with fuzzy type of uncertainty by modeling the relationship between cluster membership and class membership values. Methods based on fuzzy functions such as FCF and IFCF on the other hand, reflect the fuzziness of the system by adding crisp membership values of observations to the linear regression model as additional independent variables.

In this study, different fuzzy classification approaches are developed and IFCF approach is improved. For these developments, we basically utilize two different fuzzy approaches; Tanaka’s FLR approach, which is a widely used fuzzy regression approach, and IFCF, which is a recently developed fuzzy classification approach.

We develop three alternative fuzzy classification approaches that utilize Tanaka's FLR method for a particular case of building customer satisfaction classification models. In all of these approaches, the discrete dependent variable is converted into an equivalent continuous variable, first, and then, Tanaka's FLR modeling approach is applied on the latter. The methods differ only in the way this conversion takes place. For the conversion, alternative ways reflecting random and fuzzy types of uncertainties in the data are considered. Performances of the methods are compared and possible reasons behind their low and high performance results are discussed. Possible use of these approaches in other cases is also discussed.

In the second main part of the study, IFCF approach developed by Çelikyılmaz (2008) is improved further. In order to overcome fitting problems encountered in clustering phase of the IFCF, we propose to use a nonparametric method, Multivariate Adaptive Regression Splines (MARS), in the clustering phase of the algorithm. We call this Nonparametric Improved Fuzzy Classifier Function (NIFCF) approach. The performance of the NIFCF is compared with another fuzzy classification approach, FCF and a statistical classification method, LR using two real life data sets, customer satisfaction and casting, collected for quality improvement; and one data set from physical sciences, ionosphere.

This thesis is organized into six chapters. In the second chapter, some background information about classification, fuzziness, fuzzy modeling approaches and fuzzy clustering are given. In the third chapter, data sets used in the fuzzy classification applications are described. Three alternative classification models based on Tanaka's FLR approach are presented in chapter four. In the fifth chapter, NIFCF approach is presented and its performance is compared with those of the FCF and LR methods. Conclusions and possible future works are mentioned in the last chapter.

CHAPTER 2

LITERATURE SURVEY AND BACKGROUND

2.1. Handling Uncertainty: Randomness versus Fuzziness

Complexity of a system, the ground of the many of our problems today, increases with both the amount of information available and the amount of uncertainty allowed (Klir and Folger, 1988). Klir and Folger (1988) illustrate this issue with an example of driving a car. Driving a car with a standart transmission is more complex than driving a car with an automatic transmission since more information is needed when driving a car with a standart transmission. Also, driving a car in heavy traffic or on unfamiliar roads are more complex since we are more uncertain about when we will stop or swerve to avoid an obstacle. It is achieved by satisfactory trade-off between the amount of information available and the amount of uncertainty allowed to cope better with complexity (Klir and Folger, 1988). In many of the systems, more precision results in higher costs and less tractability of a problem (Ross, 2004). Hence, it is reasonable to increase the amount of uncertainty by sacrificing some amount of precision. Uncertainty can exist in various forms, which is resulted from ignorance, from lack of knowledge, from various classes of randomness, from inability to perform adequate measurements or from vagueness (Ross, 2004).

For handling uncertainty, several approaches have been developed. Among these approaches, probability has been widely accepted by far, which provides a conceptual basis for handling random type of uncertainty. The history of probability dates back to 16th century when rules of probability were firstly recognized in the games of chance by Gerolamo Cardano (Ross, 2004). Probability measures how likely an event occurs. For example, when it is stated that the probability it will rain tomorrow is 0.6, it means that there is %60 chance of rain tomorrow.

Handling uncertainty using probability theory was challenged by the introduction of the concept of fuzziness by Zadeh in 1965 (Ross, 2004). Fuzziness is a type of uncertainty as randomness. However, fuzziness describes the uncertainty resulted from the lack of abrupt distinction of an event (Ross, 2004), which means that the compatibility of an event with the given concept is vague. In other words, an event may not be expressed as totally compatible or totally incompatible with the given concept. It may be compatible with the given concept to some degree. This vagueness is generally resulted from the use of linguistic terms, the meanings of which vary from person to person. Hence, fuzzy logic measures how compatible an event is with the given concept while probability measures how likely an event occurs.

Klir and Folger (1988) illustrate fuzzy type of uncertainty using examples about description of weather and description of travel directions. They state that it is more useful to describe weather as sunny than giving exact percentage of cloud cover since it does not cause any loss in the meaning even if it is less precise. Also, it is more useful to describe travel directions using city blocks instead of giving exact inches. The uncertainty in these examples is resulted from the vagueness due to the use of linguistic terms. For example, the vagueness in the use of adjective ‘sunny’ is resulted from the lack of abrupt distinction between the types of weather, which can be described as sunny or not, according to particular amount of cloud cover. When the weather with %25 cloud cover is described as sunny, is it reasonable to describe weather with %26 cloud cover as not sunny? It is difficult to draw exact distinctions between these two terms; sunny and not sunny. As can be seen, there should be a gradual transition between the levels of sunny and not sunny type of weather instead of abrupt distinctions. Thus, it is more reasonable to describe weather with %26 percent cloud cover as sunny but it is less compatible with the term sunny compared to weather with %25 cloud cover. This transition between the levels is achieved by the use of membership functions. In this example, membership function measures the degree to which weather can be described as sunny according to the percentage of cloud cover. In general, membership function describes the degree of compatibility of an individual with the given concept represented by a fuzzy set (Klir and Folger, 1988). It is denoted by $\mu_A(X)$ where X denotes a universal set and A is a fuzzy set. $\mu_A(x)$ represents the membership degree of an element x, which is from the set X, to the fuzzy set A. Membership values take values between zero and one. That is,

$\mu_A: X \rightarrow [0, 1]$.

While zero degree of membership represents nonmembership, one degree of membership represents full membership. Membership values between zero and one express the partial membership.

As can be seen, there is not any abrupt distinction between the members and nonmembers in the fuzzy sets and their membership values change between values zero and one. However, there is an abrupt distinction between the members and nonmembers in the crisp sets, in which only zero and one membership values exist representing nonmembership and membership, respectively. Thus, fuzzy sets can be seen as a more general form of the crisp sets, which allow intermediate membership levels between zero and one. While crisp sets provide a mathematical basis for the probability theory, fuzzy sets provide basis for the possibility theory (Zadeh, 1978).

According to Zadeh (1978), imprecision resulted from linguistic terms is possibilistic rather than probabilistic since the concern is the meaning of information not the measure of information, which is the concern of probability theory. While probability theory copes with the uncertainties resulted from randomness, possibility theory copes with the uncertainties resulted from fuzziness. While a random variable represents a probability distribution, fuzzy variable, the distribution of which is given by a membership function, represents possibility distribution. Zadeh (1978) illustrates the distinction between probability and possibility by an example about the number of eggs that Hans ate. The possibility distribution, *π(u)* expresses the degree of ease that Hans can eat u number of eggs. However, probability distribution, *P(u)* gives information about occurrence of the event that Hans can eat u number of eggs. The probability and possibility distributions are given in Table 2.2 below (Zadeh, 1978).

Table 2.1: The Probability and Possibility Distributions for *u*, the number of eggs

u	1	2	3	4	5	6	7	8
π (u)	1	1	1	1	0.8	0.6	0.4	0.2
P(u)	0.1	0.8	0.1	0	0	0	0	0

As can be seen from the table, while Hans has the ability to eat 3 eggs in a day, which is given by the possibility value of 1, it may not be so probable for him to eat such a high number of eggs in a day, which is given by 0.1 probability value. Hence, it can be inferred that high degree of possibility does not result in high degree of probability. However, there is a connection between possibility and probability. If it is impossible to eat 10 number of eggs for Hans, it is also improbable. Zadeh (1978) named this connection between probability and possibility as *possibility/probability consistency principle*.

2.2. Classification Methods and Performance Measures

2.2.1. Classification Methods

Classification problem refers to classifying objects into given classes. Thus, classification methods aim to construct models to predict class labels while prediction methods build models to predict continuous valued dependent variable. Many methods exist in the literature particularly for quality improvement, to be used for classification problems such as Decision Trees (DT), Support Vector Machines (SVM), Multivariate Adaptive Regression Splines (MARS) and Logistic Regression (LR) (Köksal et al., 2008). DT represents the models by tree like structures composed of root nodes, internal nodes, arcs and leaf nodes. These models are easy to understand and interpret. They do not need many classical assumptions as in other classification methods, thus they are widely used for many prediction and classification problems. MARS is a nonparametric regression method, which automatically models nonlinearities and interactions in the data using piecewise linear regression models. It is a flexible modeling technique that can be used for both high dimensional classification and prediction problems. Similar to DT, MARS does not need many classical assumptions as in the other classification methods and is easy to understand and interpret. Another classification method, LR models the frequency of an event. Unlike DT and MARS, this method requires the validation of assumptions such as independency of error terms. SVM is another classification method that classifies objects by using a hyperplane achieving maximum distance between the data sets belonging to different classes.

Moreover, some nonparametric classification approaches based on Multicriteria Decision Aid (MCDA) have been developed over the last three decades (Zopounidis, 2002). The studies in this area can be divided into 2 groups: criteria aggregation models and model development techniques (Zopounidis and Doumpos, 2002). Outranking relation and utility functions are the most widely used criteria aggregation models. Outranking relation is used to estimate the outranking degree of an object over another object. If an object outranks the other object, it means that it is at least as good as that object. In this method, objects are classified by assessing their outranking degree over the reference profile, r_k, which distinguishes classes C_k and C_{k+1}. On the other hand, utility functions give overall performance measure of an object. After calculating utilities of each object, the objects are classified according to predefined utility threshold values. In model development techniques, which constitute the second group of the classification methods based on MCDA, optimal model parameters are specified by mathematical programming techniques if the model has a quantitative form. These techniques date back to 1950's when they were used to develop regression analysis and multiple criteria ranking models. In 1960's, these techniques were started to be used for classification problems and gained popularity with the development of LP models used to develop discriminant functions proposed by Hand (1981) and Freed and Glover (1981) (Zopounidis and Doumpos, 2002).

Apart from the classification methods based on the probability theory such as LR, fuzzy classification methods exist, which depend on the possibility theory. In fuzzy classification methods, uncertainty is supposed to be resulted from fuzziness of the system structure instead of randomness. These methods are explained in detail in Section 2.3.

In this section, general information about LR and MARS, which are the statistical classification methods used in this study, is given.

2.2.1.1. Logistic Regression (LR)

LR is a parametric modeling approach used for classification problems where the dependent variables are qualitative rather than continuous. It has similar general principles with linear regression analysis (Hosmer and Lemeshow, 2000). However, LR models the relationship between the independent variables and the probability of occurrence of an event, p(Y=1|**X**) while linear regression analysis models the relationship between independent variables and expected value of the dependent variable, E(Y|**X**),

where

X: Input Matrix,

Y: Output Vector.

Since change in the conditional mean gets smaller while getting closer to the values of 0 and 1, the distribution of conditional mean E(Y|**X**) for binary data resembles an S-shaped curve, which is called logistic function (Hosmer and Lemeshow, 2000). Logistic curve can be seen in Figure 2.1. Thus, a link function is used to connect the independent variables with the qualitative dependent variable, the mean value of which has a logistic distribution. One of the widely used link function in LR is logit transformation given by

$$g(\mathbf{X}) = ln\left[\frac{\pi(\mathbf{X})}{1-\pi(\mathbf{X})}\right] = \beta_0 + \boldsymbol{\beta}\mathbf{X}\,, \tag{2.1}$$

where

$\pi(\mathbf{X})$: the conditional mean of Y for a given input matrix, **X,** when the logistic distribution is used,

β_0 : intercept of the regression function,

$\boldsymbol{\beta}$: coefficient vector of the regression function.

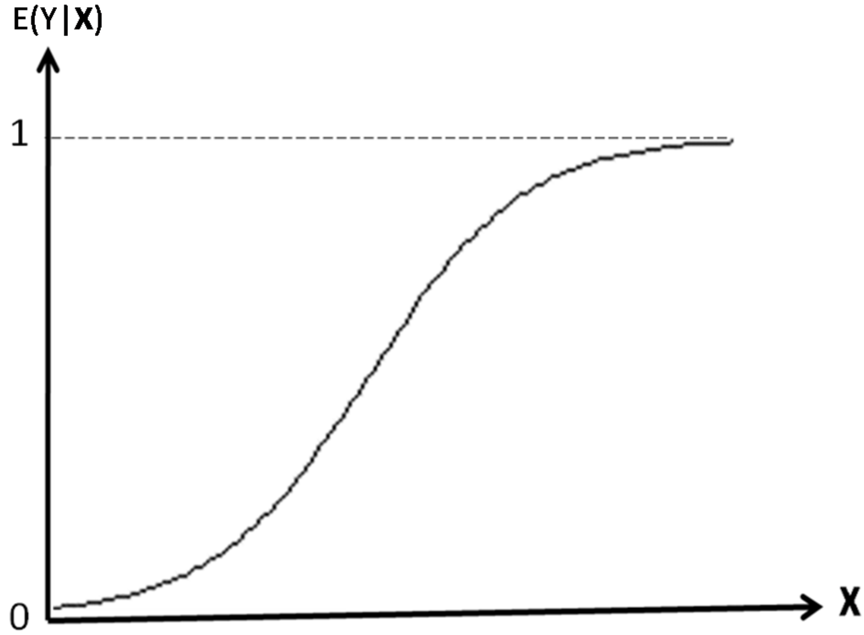

Figure 2.1: Logistic Curve

For the estimation of the logistic regression coefficients, maximum likelihood estimation method, which aims to maximize the probability of obtaining the observed values using a likelihood function,

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{N} \pi(\boldsymbol{X_i})^{y_i}[1 - \pi(\boldsymbol{X_i})]^{1-y_i}, \tag{2.2}$$

is used (Hosmer and Lemeshow, 2000).

The log-likelihood function, which is the logarithmic transformation of likelihood function, given below

$$L(\boldsymbol{\beta}) = ln[l(\boldsymbol{\beta})] = \sum_{i=1}^{N}\{y_i ln[\pi(\boldsymbol{X_i})] + (1 - y_i) ln[1 - \pi(\boldsymbol{X_i})]\} \tag{2.3}$$

provides an easier mathematical equation to work on.

Maximum likelihood estimators for the logistic regression parameters, $\hat{\boldsymbol{\beta}}$ are determined as the values that maximize the log-likelihood function.

2.2.1.2. Multivariate Adaptive Regression Splines (MARS)

MARS is a nonparametric flexible regression modeling approach developed by Friedman (1991). It automates the selection of variables, variable transformations and interactions between variables while constructing a model. Thus, it is a suitable approach for modeling high-dimensional relations and expected to show high performance for fitting nonlinear multivariate functions (Taylan et al., 2008). It can be used for both prediction and classification problems (Yerlikaya, 2008).

MARS constructs relationship between dependent and independent variables by fitting piecewise linear regression functions, in which each pieces are named as *basis functions*. Then, the model, $f(\mathbf{X})$, constructed by MARS for given input matrix, **X,** as the following:

$$y = f(\mathbf{X}) = \sum_{m=1}^{M} \boldsymbol{a}_m \boldsymbol{B}_m(\mathbf{X}), \tag{2.4}$$

where

$\boldsymbol{a}_m$: coefficient vector for the m^{th} basis function,

$\boldsymbol{B}_m(\mathbf{X})$: the m^{th} basis function,

M: the number of basis functions.

MARS builds a model using an algorithm which has two phases: the forward and the backward stepwise algorithm, which are performed only once.

The Forward Stepwise Algorithm:

The forward stepwise algorithm starts with the formation of constant basis function consisting only intercept term and then continues with the forward stepwise search to choose the basis function, which gives maximum reduction in the sum of squared errors. This process continues until the maximum number of terms is reached, which is initially determined by the user.

The Backward Stepwise Algorithm:

In this phase, it is aimed to prevent over-fitting of the model constructed in the forward stepwise algorithm. Thus, a new model with better generalization ability is constructed by removing the terms that result in smallest increase in the sum of squared errors at each step. This process continues until the best model is selected according to the Generalized Cross Validation (GCV), measure whose formula is given below

$$\text{GCV}= \frac{1}{N} \sum_{i=1}^{N} \left(y_i - f_i(\mathbf{X})\right)^2 / (1-(u+dK)/N)^2, \tag{2.5}$$

where

y_i: actual value of the i^{th} dependent variable,

$f_i(\mathbf{X})$: predicted value of i^{th} dependent variable for a given input matrix, **X,**

N: number of observations,

u: number of independent basis functions,

d: cost of optimal basis,

K: number of knots selected by forward stepwise algorithm.

Finally, the model that has the minimum GCV value is selected as the best model.

2.2.2. Performance Measures

Several classification performance measures exist in the literature to be used for evaluating the performances of classification methods. Performance measures mentioned in the studies of Weiss and Zhang (2003) and Ayhan (2009) are explained in this section.

The measures explained below use the inputs of confusion matrix, which illustrates the number of positive and negative observations classified correctly and incorrectly (see Table 2.1). True Positives (TP) and True Negatives (TN) represent the number of correctly classified actual positive and negative observations, respectively, while False Positives (FP) and False Negatives (FN) represent the number of positive and negative observations misclassified, respectively.

Table 2.2: Confusion Matrix

		Actual Class	
		Positive	Negative
Predicted Class	Positive	True Positives (TP)	False Positives (FP)
	Negative	False Negatives (FN)	True Negatives (TN)

Misclassification Rate (MCR):

Misclassification rate (MCR) is the proportion of the misclassified observations in total number of observations, *N*.

$$MCR = (FP + FN)/N \tag{2.6}$$

Percentage of Correctly Classified (PCC):

Percentage of correctly classified (PCC) is the proportion of the correctly classified observations in total number of observations.

$$PCC = (TP + TN)/N = 1 - MCR \tag{2.7}$$

Kappa:

Kappa gives the chance-corrected proportion of the correctly classified observations, in which the probability of chance agreement is removed.

$$Kappa = (\theta_1 - \theta_2)/(1 - \theta_2) \tag{2.8}$$

where θ_1 and θ_2 denote the observed and chance agreement calculated by the Equations (2.9) and (2.10), respectively.

$$\theta_1 = (TP + TN)/N \tag{2.9}$$

$$\theta_2 = \frac{[(TP + FN)/2]\,[(TP + FP)/2] + [(FP + TN)/2]\,[(FN + TN)/2]}{N^2} \tag{2.10}$$

Precision:

Precision is the proportion of the actual positive observations classified correctly in the total number of positive observations.

$$Precision = \frac{TP}{TP+FP} \tag{2.11}$$

Recall:

Recall, which is also called sensitivity, gives the proportion of the correctly classified positive observations in the total number of correctly classified positive observations and misclassified negative observations.

$$Recall = \frac{TP}{TP+FN} \tag{2.12}$$

Specificity:

Specificity is the proportion of correctly classified negative observations in the total number of correctly classified negative observations and misclassified positive observations.

$$Specificity = \frac{TN}{TN+FP} \tag{2.13}$$

F Measure:

F measure is the weighted harmonic mean of the precision and recall. Since there is a tradeoff between recall and precision, F measure gives more valuable information about test's accuracy by considering both recall and precision. $F_{0.5}$, F_1 and F_2 are widely used F measures, which are calculated by:

$$F_\beta = \frac{(1+\beta^2)(Precision)(Recall)}{\beta^2 Precision + Recall} \tag{2.14}$$

Log-Odds Ratio:

Log-Odds ratio is the natural logarithm of the odds ratio between the correctly classified and misclassified observations.

$$LogOdds\ Ratio = log\left(\frac{(TP)(TN)}{(FP)(FN)}\right) \tag{2.15}$$

Stability:

A classification model is said to be stable if its performance for testing data is close to its performance for training data.

$$Stability == \frac{PCC_{TR} - PCC_{TE}}{PCC_{TR} + PCC_{TE}} \tag{2.16}$$

where PCC_{TR} and PCC_{TE} denote the percentage of the correctly classified values for the training and testing data sets, respectively.

While the measure is getting closer to 0, the classification model is said to be more stable.

Area under ROC Curve (AUC):

It measures the area under the ROC curve, which is a plot of the sensitivity versus $(1 - specificity)$.

2.3. Fuzzy Modeling Approaches

In general, fuzzy modeling methods can be grouped into three: Fuzzy Rule Based (FRB) approaches, Fuzzy Regression (FR) approaches and approaches based on fuzzy functions.

2.3.1. Fuzzy Rule Based (FRB) Approaches

FRB approach was firstly developed by Zadeh (1965 and 1975, as cited in Türkşen and Çelikyılmaz, 2006) and applied by Mamdani (1981, as cited in Türkşen and Çelikyılmaz, 2006). These models have been applied and improved by many researchers. Takagi and Sugeno (1985) have developed a mathematical tool to model a system by using fuzzy implications and reasoning. Thus, natural languages used in daily life can be added in the models by fuzzy reasoning and applications and input-output relations can be built. Sugeno and Yasukawa (1993) have developed a method to build a qualitative model, which is divided into two parts called *fuzzy modeling* and *linguistic approximation*.

In addition, several other approaches have been developed to be used for building classification models based on FRB. These approaches mainly include methods based on fuzzy clustering and Adaptive Neuro Fuzzy Inference Systems (ANFIS). Fuzzy Classifier (FC) developed by Abe and Thawonmas (1997) and Fuzzy Relational Classifier (FRC) developed by Setnes and Babuşka (1999) are among the FRB methods depending on fuzzy clustering. In FC method, fuzzy clusters are determined for each class and fuzzy rules are developed for each cluster determined. In the FRC method, the membership values are calculated by any clustering algorithm such as fuzzy c-means (FCM) and the fuzzy relation between cluster membership values and class membership values is built. Huang et al. (2007) have developed a classifier depending on ANFIS. This method is composed of two stages: feature extraction and ANFIS. In the stage of feature extraction, the inputs are selected by using orthogonal vectors and then in the final step ANFIS method is applied.

2.3.2. Fuzzy Regression (FR) Approaches

In this section, fuzzy regression approaches used for prediction problems are explained in three groups, which are possibilistic approaches, Fuzzy Least Squares Regression (FLSR) approaches and other approaches.

2.3.2.1. Possibilistic Approaches

Fuzzy Linear Regression (FLR) analysis was firstly developed by Tanaka et al. (1982) and generally named as “possibilistic regression”. In this regression approach, deviations between observed and estimated output variables are assumed to be resulted from the fuzziness of the system structure, not the measurement errors as in the conventional statistical methods (Tanaka et al., 1982). In order to be able to apply this method, observed independent variables must be crisp numbers. However, an observed dependent variable may be crisp or symmetrical triangular fuzzy number whose spread is represented by e_i taking the value zero for crisp number and positive value for the fuzzy number. Spread is a meaure of dispersion of a fuzzy number. In symmetrical triangular fuzzy numbers, spread is calculated as the half width of the fuzzy interval.

This approach uses a Linear Programming (LP) model, in which total fuzziness of the regression coefficients are minimized when predicted intervals include observed intervals at a certain degree of fit. The LP model is shown below:

$$\text{Min } J = \sum_{i=1}^{N} \sum_{j=0}^{M} c_j \quad (2.17)$$

s.t.

$$\sum_{j=0}^{M} m_j \, x_{ij} + (1 - H) \sum_{j=0}^{M} c_j \left|x_{ij}\right| \geq y_i + (1 - H) e_i \quad \text{for } i = 1, \dots, N \quad (2.18)$$

$$\sum_{j=0}^{M} m_j \, x_{ij} - (1 - H) \sum_{j=0}^{M} c_j \left|x_{ij}\right| \leq y_i - (1 - H) e_i \quad \text{for } i = 1, \dots, N \quad (2.19)$$

$$c_j \geq 0, \ m_j \text{ free} \quad \text{for } j = 0, \dots, M \quad (2.20)$$

where the variables are

x_{ij} : value of the j^{th} independent variable for the i^{th} observation,

y_i : value of the dependent variable for the i^{th} observation,

and the parameters are

e_i : spread of the dependent variable for the i^{th} observation,

H : target degree of belief,

m_j : midpoint of the j^{th} regression coefficient,

c_j : spread of the j^{th} regression coefficient,

M : number of independent variables,

N : number of observations.

Fuzzy regression coefficient parameters are estimated by solving this LP problem that minimizes the total fuzziness of the system. Total fuzziness of the system is expressed by the total half widths (spreads), c_j; of the regression coefficients as shown in Equation (2.17). In this LP model, predicted intervals should include observed intervals at H level degree of fit, which is ensured by the constraints given in Equations (2.18) and (2.19).

H level, which is determined by the user, is called the target degree of belief (Chang and Ayyub, 2001). It can be considered as a measure of goodness of fit for FLR method, which shows the compatibility between the FLR model and the data (Chang and Ayyub, 2001). According to the LP model, membership value of an observed dependent variable to its estimated fuzzy dependent variable, h_i, must be at least H (Tanaka et al, 1982). h_i value for crisp and fuzzy observed dependent variables is illustrated in Figure 2.2. As can be seen from the Figure 2.2, predicted fuzzy interval, $\widehat{\mathrm{Y}}_i$, contain observed fuzzy interval, $\widetilde{\mathrm{Y}}_{\mathrm{i}}$ or observed crisp number, Y_{i} for the membersip values equal to or below h_i. Thus, h_i is the maximum membership degree that predicted fuzzy interval contain observed fuzzy interval or crisp number.

For a symmetrical triangular fuzzy number, h_i is obtained by the following equation:

$$h_i = 1 - \frac{\left|y_i - \sum_{j=0}^{M} m_j x_{ij}\right|}{\sum_{j=0}^{M} c_j \left|x_{ij}\right| - e_i} \tag{2.21}$$

The LP model given above aims to find fuzzy regression coefficients under the constraints (2.18) and (2.19), which are derived from the inequality $h_i \geq H$ for $i = 1, \ldots, N$. According to the LP model, midpoints of the predicted fuzzy regression coefficients are not affected by the H value, however, spread values of the predicted fuzzy regression coefficients increase with the increase in H value (Tanaka and Watada, 1988 as cited in Kim, Moskowitz and Köksalan, 1996).

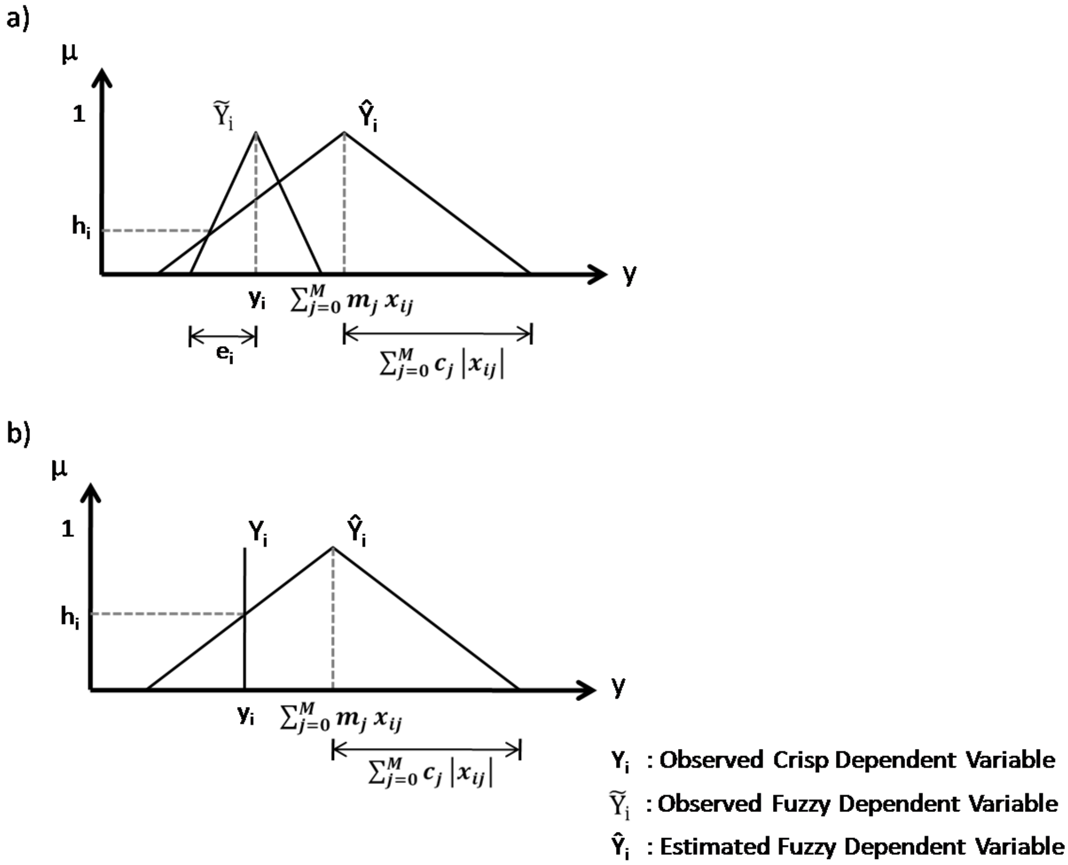

Figure 2.2: Illustration of h value for a) fuzzy dependent variable and b) crisp dependent variable

Since h_i value increases when the midpoints of predicted and observed dependent variable get closer, H level can be seen as the level of credibility or level of confidence desired (Kim, Moskowitz and Köksalan, 1996). Since it is determined by the user, proper selection of H level is important for the fuzzy regression model (Wang and Tsaur, 2000). Tanaka and Watada (1988) suggested to determine H value according to the sufficiency of the data (Wang and Tsaur, 2000). If the data set collected is sufficiently large, then H level should be determined as 0 and it should be increased with the decreasing volume of the data set.

This method has been criticized for having input scale dependencies and often having zero coefficient spread (Jozsef, 1992 as cited in Hojati et al., 2005). To overcome these problems, Tanaka improved his method by replacing the objective function, representing total fuzziness of the coefficients, by the total fuzziness of the predicted values under the same constraints (Tanaka et al., 1989). In this case the LP model becomes as follows:

$$\text{Min } J = \sum_{i=1}^{N} \sum_{j=0}^{M} c_j \left| x_{ij} \right| \quad (2.22)$$

s.t.

$$\sum_{j=0}^{M} m_j x_{ij} + (1-H) \sum_{j=0}^{M} c_j \left| x_{ij} \right| \geq y_i + (1-H) e_i \text{ for } i = 1, \dots, N \quad (2.23)$$

$$\sum_{j=0}^{M} m_j x_{ij} - (1-H) \sum_{j=0}^{M} c_j \left| x_{ij} \right| \leq y_i - (1-H) e_i \text{ for } i = 1, \dots, N \quad (2.24)$$

$$c_j \geq 0 \text{ , } m_j \text{ free} \qquad \text{for } j = 0, \dots, M \quad (2.25)$$

As a solution of this LP problem, symmetrical triangular fuzzy regression coefficients are obtained, which are denoted by $\tilde{A}_j = \left(m_j , c_j \right)$ (see Figure 2.3).

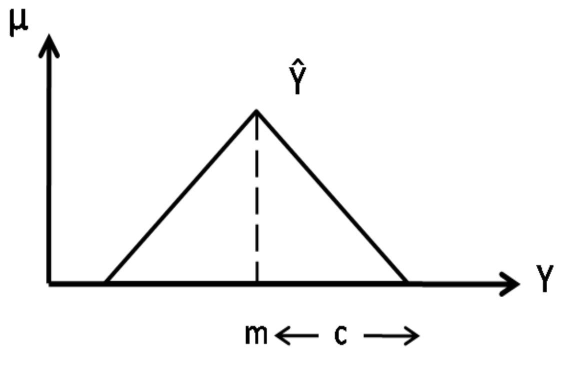

Figure 2.3: Triangular fuzzy number with mean m and spread c

In this method, the predicted dependent variables are triangular fuzzy numbers as a result of the fuzzy regression model including symmetrical triangular fuzzy regression coefficients and crisp independent variables. Illustration of a simple FLR model can be seen from Figure 2.4 below.

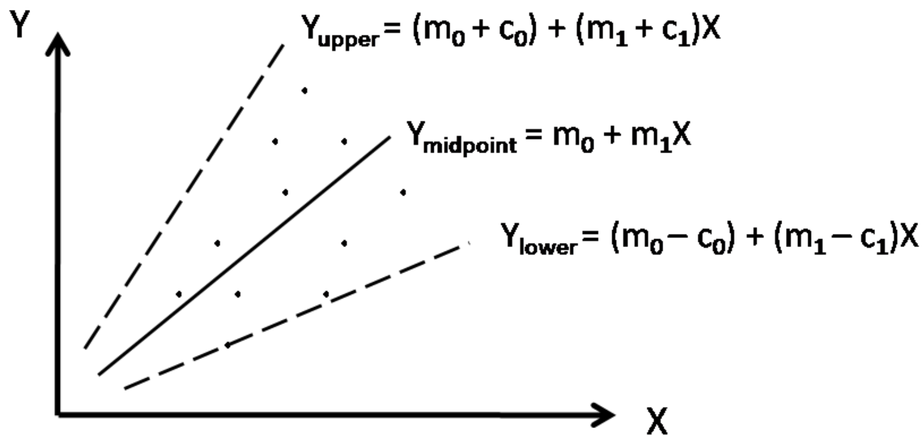

Figure 2.4: Graph of Simple FLR Model

Tanaka's method was criticized for its sensitivity to outliers (Peters, 1994), requirement of crisp independent variables and linearity assumption between dependent and independent variables (Sakawa and Yano, 1992), lack of interpretation of fuzzy intervals (Wang and Tsaur, 2000), uncertainty in the field of forecasting (Savic and Pedrycz, 1991 as cited in Wang and Tsaur, 2000) and existence of multicollinearity (Wang and Tsaur, 2000) and increasing spreads for estimated outputs (Nasrabadi and Nasrabadi, 2004) with increased number of independent variables. In addition, increasing number of constraints with the increase in the number of data results in problems about the capacity of softwares used to solve LP problems (Chang and Ayyub, 2001).

Tanaka's model has been improved several times by researchers to overcome these problems. Wang and Tsaur (2000) propose several improvements addressing some of these criticisms; they prove that the midpoints ($Y^{h=1}$) are the points that best represent the given fuzzy interval when the fuzzy regression coefficients are symmetrical triangular fuzzy numbers (Wang and Tsaur, 2000). Savic and Pedrycz (1991) (as cited in Chung and Ayyub, 2001) have developed a method, in which regression coefficient spreads are predicted by Tanaka's method while midpoints are calculated using the least-squares regression method. Sakawa and Yano (1992) have developed an iterative algorithm for the case where both dependent and independent variables are fuzzy. Peters (1994) has developed a model, in which predicted intervals are allowed to intersect observed intervals rather than including them in order to reduce the sensitivity of the FLR method to outliers. Hung and Yang (2006) have developed an approach that proposes to omit outliers for Tanaka's FLR method to overcome the problems resulted from the existence of outliers. In this approach, the outliers are tried to be determined by examining the effect of omitted variables on the objective function value and by using several visual statistical graphs such as box plots. Kim and Bishu (1998) have revised the objective function by minimizing the difference between membership values of the observed and predicted fuzzy numbers. Ozelkan and Duckstein (2000) have developed a method that minimizes the difference between the observed and predicted intervals' upper and lower bounds when the predicted intervals intersect the observed intervals. Hojati et al. (2005) have developed a method similar to Ozelkan and Duckstein (2000) but their method tries to obtain narrower intervals by minimizing the difference between the upper and lower bounds of the observed and predicted dependent variable values whether the predicted intervals intersect the observed intervals or not. Kao and Chyu (2002) have developed a method depending on two phases. In the first phase, fuzzy observations are converted to crisp numbers by defuzzification and regression coefficients are calculated by the classical least squares regression method. In the second phase, the error term of the model is determined by using the calculated regression coefficients. Since regression coefficients are crisp numbers, problems resulting from increasing spreads because of the increase in the number of independent variables as in many other fuzzy regression methods are not encountered.

Several variable selection algorithms have been proposed in order to overcome the multicollineraity problems resulted from the increased number of independent variables. Wang and Tsaur (2000) have developed a variable selection algorithm based on minimizing the sum of squared errors. However, this algorithm does not guarantee the optimal solution. D'urso and Santoro (2006) have developed variable selection algorithms that depend on coefficient of determination (R^2), adjusted coefficient of determination (Adjusted-R^2) and Mallows C_p statistics.

Another possibilistic regression approach is interval regression analysis. Ishibuchi (1992) has developed interval regression analysis by assuming that fuzzy data and fuzzy coefficients behave like interval numbers having no membership function. In this model, regression coefficients, which are interval numbers, are tried to be determined by an LP minimizing total predicted interval lengths while predicted intervals include observed intervals, as in the Tanaka's method (1982 and 1989).

2.3.2.2. Fuzzy Least Squares Regression (FLSR) Approaches

Another fuzzy regression approach is Fuzzy Least Squares Regression (FLSR) approach developed by Diamond (1988). Celmins (1987) has applied FLSR approach with using conic dependent membership functions. Wang and Tsaur (2000) have developed a new FLSR method used for crisp independent variables and fuzzy dependent variables and compared this method with Tanaka's FLR method and conventional statistical least squares regression method. They conclude that their proposed method performs better than Tanaka's method in terms of prediction power and more efficient than statistical least squares regression method in terms of computational efficiency. D'Urso and Gastaldi (2000) have proposed *doubly linear adaptive fuzzy regression model.* In this method, two different models are constructed, which are called *core regression model* and *spread regression model* for explaining centers and spreads of fuzzy numbers, respectively. In this approach, models are formed in such a way to consider the relationship between centers and spreads. D'Urso (2003) has improved the method developed by D'Urso and Gastaldi (2000), which can be applied for only crisp independent and fuzzy dependent variables, in order to be used for every combination of crisp/fuzzy independent and dependent variables.

2.3.2.3. Other Approaches

In addition to the approaches mentioned above, different approaches have also been developed to capture the effect of fuzzy type of uncertainty in the data. Hathaway and Bezdek (1993) have developed fuzzy c-regression approach. They have developed an algorithm in which both problems of clustering of dataset and determination of regression coefficients are tried to be solved simultaneously. In this algorithm, sum of the squared errors between observed and predicted dependent variables weighted with membership values is minimized.

Another fuzzy regression approach has been developed by Bolotin (2005). He proposes to replace indicator variables by membership values in linear regression models with indicator variables and predict crisp regression coefficients by least squares regression method. Thus, in this method, fuzziness in the data is captured by using the membership values replacing the indicator variables while the fuzziness is reflected by the fuzzy regression coefficients in the possibilistic FLR models. By the use of this model, common problems aroused from the use of possibilistic FLR models are aimed to be accomplished.

2.3.3. Approaches Based on Fuzzy Functions

In this section, approaches based on fuzzy functions developed for both prediction and classification problems are explained.

2.3.3.1. Fuzzy Functions (FF)

Türkşen (2008) has developed Fuzzy Functions (FF) approach, the conceptual origin of which is based on the studies of Demirci (1998, 2003 and 2004, as cited in Türkşen, 2008). FF approaches are proposed to be determined by Least Squares Regression (LSR) and Support Vector Machines (SVM). This method depends on construction of one fuzzy function for each cluster after partitioning the data with fuzzy c-means (FCM) clustering algorithm. Membership values of the observations for each cluster obtained from a clustering algorithm and their possible transformations are taken as new input variables in addition to the original input space and functions are constructed to explain input-output relationship for each cluster using the new input space, which are called “Fuzzy Functions” (Çelikyılmaz, 2008). The final estimate for output variable is obtained by weighted average of the estimates obtained for each cluster with related membership values.

Türkşen and Çelikyılmaz (2006) compare the performance of the FF method with the FRB approaches of Sugeno and Yasukawa (1993) and Takagi and Sugeno (1985). The comparison results indicate that FF methods show better performance than other methods for most of the data sets used. Moreover, they state that the proposed approach is more suitable for the analysts, who are familiar with the applications of conventional statistical regression but do not master all the aspects of the fuzzy theory since it only requires basic understanding of membership functions and the use of fuzzy clustering algorithms.

2.3.3.2. Fuzzy Classifier Functions (FCF)

Fuzzy Classifier Functions (FCF) approach is the adaptation of FF approach to classification problems (Çelikyılmaz et al., 2007). This method is very similar to FF method, but a classification method is used for building a model for each cluster rather than a prediction method as in FF approach. LR or SVM classification methods are proposed to be used for building linear or nonlinear fuzzy classifier functions for each cluster. The training and testing algorithms of the FCF approach are given below.

Steps of the training algorithm for FCF:

1. Set initial parameter α, which is the level used for eliminating the points farther away from the cluster centers.

2. Calculate cluster centers for input-output variables using the FCM algorithm for *c* number of clusters and *m* degree of fuzziness.

$$\boldsymbol{v}(\boldsymbol{X}Y)_{\boldsymbol{i}} = \left\{v(x_1)_i, \dots, v\left(x_p\right)_i, v(y)_i\right\}$$

where,

$v\left(x_J\right)_i$: cluster center of the *j*th independent variable for the *i*th cluster,

$v(y)_i$: cluster center of the dependent variable for the i^{th} cluster.

3. For each cluster $i = 1, \dots, n$

    3.1. For each observation number $k = 1, \dots, N$

Using cluster centers for input space, $\boldsymbol{v}(\boldsymbol{X})_{\boldsymbol{i}} = \left\{ v(x_1)_i, \ldots, v\left(x_p\right)_i \right\}$

3.1.1. Calculate membership values for input space, u_{ik}.

$$u_{ik} = \left(\sum_{j=1}^{n} \left[\frac{\|\boldsymbol{X_k} - \boldsymbol{v}(\boldsymbol{X})_{\boldsymbol{i}}\|}{\|\boldsymbol{X_k} - \boldsymbol{v}(\boldsymbol{X})_{\boldsymbol{j}}\|} \right]^{\frac{2}{m-1}} \right)^{-1} \tag{2.27}$$

3.1.2. Calculate alpha-cut membership values, μ_{ik}.

$$\mu_{ik} = \{u_{ik} \geq \alpha\} \tag{2.28}$$

3.1.3. Calculate normalized membership values, γ_{ik} .

$$\gamma_{ik} = \frac{\mu_{ik}}{\sum_{j=1}^{n} \mu_{jk}} \tag{2.29}$$

3.2. Determine the new augmented input matrix for each cluster *i*, $\mathbf{\Phi_i}$, using observations selected according to α-cut level. $\mathbf{\Phi_i}$ matrix is composed of input variable matrix, $\mathbf{X_i^{\alpha}}$, vector of normalized membership values for the cluster *i*, $\boldsymbol{\gamma_i}$, and the matrix composed of their selected transformations, $\boldsymbol{\gamma_i}'$, such as $\boldsymbol{\gamma_i}^2, \boldsymbol{\gamma_i}^3, \boldsymbol{\gamma_i}^m$, exp($\boldsymbol{\gamma_i}$), log(($\mathbf{1}$-$\boldsymbol{\gamma_i}$)/ $\boldsymbol{\gamma_i}$).

$$\mathbf{\Phi_i}(\mathbf{X}, \boldsymbol{\gamma_i}) = [\mathbf{X_i^{\alpha}} \quad \boldsymbol{\gamma_i} \quad \boldsymbol{\gamma_i'}]$$

where,

$$\mathbf{X_i^{\alpha}} = \{\boldsymbol{x_k} \in \mathbf{X} \,|u_{ik}(\boldsymbol{x_k}) \geq \alpha, k = 1, \ldots, N\}$$

3.3. Using LR or SVM as a classifier, calculate a *local fuzzy function* using new augmented matrix $\mathbf{\Phi_i}(\mathbf{X}, \boldsymbol{\gamma_i})$ **.**

3.3.1. For each observation $k = 1, \ldots, N$

3.3.1.1. Using the local fuzzy classifier function constructed at step 3.3, calculate posterior probabilities, $\hat{p}_{ik}\,(y_k^{\alpha} = 1/\mathbf{\Phi_i}(\boldsymbol{x}, \boldsymbol{\gamma_i}))$.

4. For each observation $k = 1, \dots, N$

4.1. Calculate a single probability output $\hat{p}_k$, weighting the posterior probabilities, $\hat{p}_{ik}$, with their corresponding membership values, γ_{ik}.

$$\hat{p}_k = \frac{\sum_{i=1}^{n} \gamma_{ik}\,\hat{p}_{ik}\,(y_k^{\alpha} = 1/\boldsymbol{\Phi}_i(\boldsymbol{X}, \boldsymbol{\gamma_i}))}{\sum_{i=1}^{n} \gamma_{ik}} \tag{2.30}$$

Steps of testing algorithm for FCF:

1. Standardize testing data.

2. For each observation $r = 1, \dots, N^{test}$

2.1. For each cluster $i = 1, \dots, n$

2.1.1. Calculate improved membership values, u_{ir}^{test}.

$$u_{ir}^{test} = \left(\sum_{j=1}^{n} \left[\frac{\|\boldsymbol{x_r}^{test} - \boldsymbol{v}(\boldsymbol{X})_i\|}{\|\boldsymbol{x_r}^{test} - \boldsymbol{v}(\boldsymbol{X})_j\|} \right]^{2/(m-1)} \right)^{-1} \tag{2.31}$$

where,

$\boldsymbol{x_r}^{test}$: testing data input vector for the r^{th} observation,

$\boldsymbol{v}(\boldsymbol{X})_i$: the i^{th} cluster centers for input variables calculated using training data set at the FCM algorithm.

2.1.2. Calculate alpha-cut membership values, μ_{ir}^{test}.

$$\mu_{ir}^{test} = \{u_{ir}^{test} \geq \alpha\} \tag{2.32}$$

2.1.3. Calculate normalized membership values, γ_{ir}^{test}.

$$\gamma_{ir}^{test} = \frac{\mu_{ir}^{test}}{\sum_{q=1}^{n} \mu_{qr}^{test}} \tag{2.33}$$

2.1.4. Determine the new augmented input vector, $\mathbf{\Phi}_{\mathbf{ir}}^{\mathbf{test}}$ which is composed of testing data input vector for the r^{th} observation, $\boldsymbol{x_r}^{test}$, normalized membership value of the r^{th} observation for the i^{th} cluster, γ_{ir}^{test}, and the vector composed of their transformations, $[\boldsymbol{\gamma_{ir}^{test}}]'$ used at the 3.2. step of the training data algorithm of FCF.

$$\mathbf{\Phi}_{\mathbf{ir}}^{\mathbf{test}}(\boldsymbol{x_r}^{test}, \gamma_{ir}^{test}) = \left[\mathbf{x}_{\mathbf{r}}^{test} \quad \gamma_{ir}^{test} \quad [\boldsymbol{\gamma_{ir}^{test}}]'\right]$$

2.1.5. Using the fuzzy classifier function constructed at step 3.3 of FCF training data algorithm, calculate posterior probabilities, $\hat{p}_{ir}^{test}(y_r^{test} = 1 / \boldsymbol{\Phi_{ir}^{test}})$.

2.2. Calculate a single probability output $\hat{p}_r^{test}$, weighting the posterior probabilities, $\hat{p}_{ir}^{test}$, with their corresponding membership values, γ_{ir}^{test}.

$$\hat{p}_r^{test} = \frac{\sum_{r=1}^{n} \gamma_{ir}^{test}\, \hat{p}_{ir}^{test}\ (y_r^{test} = 1 / \mathbf{\Phi}_{\mathbf{ir}}^{\mathrm{test}})}{\sum_{i=1}^{n} \gamma_{ir}^{test}} \tag{2.34}$$

The use of the transformations of the membership values as new input variables in both FF and FCF methods ensures that data points closer to the cluster center have more impact on the model constructed for this cluster since they have greater membership values than the others that are farther away from the related cluster center (Çelikyılmaz et al., 2007).

2.3.3.3. Improved Fuzzy Functions (IFF)

Improved Fuzzy Functions (IFF) approach has been developed by Çelikyılmaz (2008). This approach proposes to use Improved Fuzzy Clustering (IFC) algorithm developed by Çelikyılmaz (2008), which is explained in detail in Section 2.4.2, in the clustering phase of the FF approach. As mentioned above, the membership values obtained from the FCM algorithm are used as new predictors to estimate output variable for constructing fuzzy functions. However, they may not be optimum membership values to be used as predictors since they are calculated by the FCM algorithm which considers only data vectors' distances as similarity measure while partitioning data. Thus, a new clustering algorithm, IFC has been proposed by Çelikyılmaz (2008) in order to obtain optimum membership values to be used for FF approaches, which are used as new predictors. By using IFC, it is aimed that the prediction error is minimized by improving prediction power of membership functions. The prediction power of membership values are tried to be increased by considering also the relationship between actual output values and membership values of related cluster and their transformations without including original input variables. The model constructed to estimate output value using only membership values and their transformations is called *interim fuzzy function.* The squared error term between the actual output and the estimated output of the *interim fuzzy function* is added to the objective function of the FCM algorithm as seen below

$$J^{IFC} = \underbrace{\sum_{i=1}^{n} \sum_{k=1}^{N} \mu_{ik}^{m} \, d_{ik}^{2}}_{\text{FCM}} + \sum_{i=1}^{n} \sum_{k=1}^{N} \mu_{ik}^{m} \left(y_k - f(\boldsymbol{\tau}_{ik})\right)^2. \tag{2.35}$$

After optimum membership values are obtained by the IFC algorithm, FF method is applied. The training algorithm of IFF is exactly the same as the training algorithm of FF method except that it uses the membership values calculated by the IFC algorithm rather than the FCM algorithm. However, in the testing algorithm of IFF, the actual value of output variable is needed in order to calculate the squared error term. Thus, in the testing algorithm of IFF, k-nearest neighbor algorithm is used to find an estimate for squared error term using the actual values of k-nearest neighbors from the training data set.

Çelikyılmaz and Türkşen (2008) propose to use LSR as a linear function estimation method and SVM as a nonlinear function estimation method for estimating *interim fuzzy functions* in the IFC algorithm.

Çelikyılmaz and Türkşen (2008) compare the IFF approach with other fuzzy modeling approaches; FRB, FF and ANFIS and non-fuzzy approaches, SVM and Neural Networks (NN) using three data sets. The results of the experiments indicate that the IFF method gives better performance results.

2.3.3.4. Improved Fuzzy Classifier Functions (IFCF)

As FCF, Improved Fuzzy Classifier Functions (IFCF) approach is the extension of IFF method, which is used for prediction problems, to be used for classification problems. In the IFCF algorithm, classifier functions are used for constructing both *interim fuzzy functions* in the clustering phase of the algorithm and *local fuzzy functions* instead of regression functions. SVM and LR classification methods are proposed to be used for the construction of *local fuzzy classifier functions.*

2.4. Fuzzy Clustering

Clustering means grouping of objects into subclasses, which are called clusters. The members of the clusters bear more mathematical similarity among each other than other members of the clusters (Ross, 2004). Similarity between objects in a cluster is generally defined by a distance measure. Clustering analysis can be performed by two types of clustering methods: hard clustering and fuzzy clustering. While hard clustering methods partition data with hard boundaries, the boundaries between clusters determined by fuzzy clustering methods are vague. Thus, an object may belong to several clusters with different membership values in the fuzzy clustering methods, not fully belong to only one cluster as in the hard clustering methods.

In this section, two fuzzy clustering methods, FCM and IFC, which are used to find sub-grouping of objects for the methods applied in this study and the validity measures used to find optimum parameters for these clustering algorithms are explained.

2.4.1. Fuzzy c-Means (FCM) Clustering

As indicated above, FCM is among the fuzzy clustering methods, which provides fuzzy partitioning of data by assigning membership degrees to each object describing their belongings to the related clusters. FCM algorithm has been developed by Bezdek (1981) (as cited in Ross, 2004).

FCM clustering algorithm aims to find fuzzy partitions that minimize the objective function $J(\boldsymbol{X};\boldsymbol{U},\boldsymbol{V})$ given by

$$J(\boldsymbol{X};\boldsymbol{U},\boldsymbol{V}) = \sum_{i=1}^{n}\sum_{k=1}^{N}\mu_{ik}^{m}\,\|\boldsymbol{x_k} - \boldsymbol{v_i}\|^2, \tag{2.36}$$

subject to the constraints

$$\sum_{i=1}^{n}\mu_{ik} = 1$$

$$0 < \sum_{k=1}^{N}\mu_{ik} < N$$

$$\mu_{ik} \in [0,1]$$

where

n: number of clusters,

m: degree of fuzziness,

N: number of observations,

$\boldsymbol{v_i}$: center of the *i*^th^ cluster,

$\boldsymbol{x_k}$: the k^{th} observation vector,

μ_{ik}: membership value of the k^{th} observation for the i^{th} cluster.

The algorithm of FCM provided by user guide of *Fuzzy Clustering Toolbox* is given below.

Steps of FCM algorithm:

1. For a given data set **X**, determine the number of clusters, n, degree of fuzziness, m and termination tolerance, ε. Initialize the partition matrix $U^{(0)} = [\mu_{ik}]$ where μ_{ik} denotes the membership value of the k^{th} object for the i^{th} cluster.

2. Compute the cluster centers, $v_i^{(l)}$ for the i^{th} cluster.

$$v_i^{(l)} = \frac{\sum_{k=1}^{N}\left(\mu_{ik}^{(l-1)}\right)^m \boldsymbol{x}_k}{\sum_{k=1}^{N}\left(\mu_{ik}^{(l-1)}\right)^m} \quad \text{for } \forall i = 1, \dots, c \tag{2.37}$$

3. Calculate distances d_{ik} of the k^{th} observation for the i^{th} cluster.

$$d_{ik} = \|\boldsymbol{x_k} - \boldsymbol{v_i}\|^2 \quad \text{for } \forall i = 1, \dots, n \quad \text{for } \forall k = 1, \dots, N \tag{2.38}$$

4. Update the partition matrix $U^{(l)} = \left[\mu_{ik}^{(l)}\right]$.

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{n}(d_{ik}/d_{jk})^{1/(m-1)}} \tag{2.39}$$

until $\left\| U^{(l)} - U^{(l-1)} \right\| < \varepsilon$

2.4.2. Improved Fuzzy Clustering (IFC)

IFC approach has been developed by Çelikyılmaz (2008). This proposed algorithm aims to transform membership values into powerful predictors to be used for approaches based on fuzzy functions. Therefore, while partitioning data, the relationship between input and output variables is considered in addition to the similarity based on distance measures in order to minimize the modeling error of fuzzy functions.

The prediction power of membership values are tried to be increased by using a function called *interim fuzzy function*, which is constructed to estimate output variable by using only membership values and their transformations. LSE and SVM methods are proposed to be used for construction of *interim fuzzy functions* (Çelikyılmaz, 2008). The squared error between the predicted output of *interim fuzzy functions,* $f(\boldsymbol{\tau}_{ik})$, and actual output, y_k, is considered as additional similarity measure and added to the objective function of the FCM clustering algorithm as given below:

$$J_m^{IFC} = \sum_{i=1}^{n}\sum_{k=1}^{N}\mu_{ik}^{m}\left\|\boldsymbol{x_k} - \boldsymbol{v}_i\right\|^2 + \sum_{i=1}^{n}\sum_{k=1}^{N}\mu_{ik}^{m}\left(y_k - f(\boldsymbol{\tau}_{ik})\right)^2 , \tag{2.40}$$

where $\boldsymbol{\tau}_{ik}$ is the row vector composed of the membership value of the k^{th} observation for the i^{th} cluster and its transformations. Improved membership values are calculated by minimizing this objective function under the same constraints with FCM.

By the use of proposed objective function, it is aimed both to calculate membership values that are good predictors to be used for *local fuzzy functions* and to ensure proper partitioning of data according to distance similarity measure (Çelikyılmaz, 2008).

The algorithm of IFC is very similar to the FCM algorithm. In the IFC algorithm, the initial partition matrix is calculated using the FCM or any other clustering algorithm. Then, the cluster centers are computed using the same equation in the second step of the FCM algorithm. In the IFC algorithm, the distance function is changed to include squared error term between actual output and predicted output value as below:

$$d_{ik}^{IFC} = \underbrace{\|\boldsymbol{x_k} - \boldsymbol{v_i}\|^2}_{d_{ik}^{FCM}} + (y_k - f(\boldsymbol{\tau}_{ik}))^2 . \tag{2.41}$$

Thus, the membership values, μ_{ik},

$$\mu_{ik} = \sum_{j=1}^{n} \left[\frac{d_{ik}^{IFC}}{d_{jk}^{IFC}}\right]^{1/(1-m)} = \sum_{j=1}^{n} \left[\frac{d_{ik}^{FCM} + (y_k - f(\boldsymbol{\tau}_{ik}))^2}{d_{jk}^{FCM} + (y_k - f(\boldsymbol{\tau}_{ik}))^2}\right]^{1/(1-m)} , \tag{2.42}$$

are calculated using the distances calculated in Equation (2.41).

To the IFC algorithm for classification problems, the procedure is the same except that the *interim fuzzy functions* are constructed by using a classifier function instead of a regression function. When Logistic Regression (LR) is used, the distance function becomes

$$d_{ik}^{IFC} = \|\boldsymbol{x_k} - \boldsymbol{v_i}\|^2 + (y_k - \hat{p}_{ik}(y_k = 1/\boldsymbol{\tau}_{ik}))^2 , \tag{2.43}$$

where $\hat{p}_{ik}(y_k = 1/\boldsymbol{\tau}_{ik})$ denotes the posterior probability of the k^{th} observation for a given $\boldsymbol{\tau}_{ik}$ vector, which is composed of membership value of the k^{th} observation for the i^{th} cluster and its transformations.

LR, SVM and NN are proposed to be used as a classifier for constructing *interim fuzzy functions* for classification problems (Çelikyılmaz, 2008).

Çelikyılmaz (2008) compares the results of the FCM and IFC algorithms using an artificial data set. The prediction power of membership values calculated by the FCM and IFC algorithms are compared by evaluating the significance of fuzzy functions, which are constructed by these membership values and their log-odds transformations, using F-value and p-value measures. The comparison results indicate that membership values calculated by the IFC algorithm are better predictors of the output variable than the membership values calculated by the standard FCM algorithm.

2.4.3. Validity Indices

Before the application of a clustering algorithm, initial parameters such as degree of fuzziness, the number of clusters, must be determined. Several validity indices are proposed to be used for the determination of optimum parameter values in the literature. Some of the widely used validity indices are given in the handbook of *Fuzzy Clustering and Data Analysis Toolbox* of *MATLAB* for *N* number of observations, which are listed in Table 2.3. Since none of the indices is reliable only by itself, the handbook proposes to choose optimum parameter value by evaluating several validity indices.

Table 2.3: Validity Indices

Validity Index	**Equation**	**Objective**
Partition Coefficient (PC)	$\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{N}\mu_{ij}^{2}$	Max
Classification Entropy (CE)	$-\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{N}\mu_{ij}\,log(\mu_{ij})$	Min
Partition Index (SC)	$\sum_{i=1}^{n}\frac{\sum_{j=1}^{N}\mu_{ij}^{m}\left\lVert x_j - v_i\right\rVert^2}{N\sum_{k=1}^{n}\lVert v_k - v_i\rVert^2}$	Min
Separation Index (S)	$\frac{\sum_{i=1}^{n}\sum_{j=1}^{N}\mu_{ij}^{2}\left\lVert x_j - v_i\right\rVert^2}{N\min_{i,k}\lVert v_k - v_i\rVert^2}$	Min
Xie and Beni's Index (XB)	$\frac{\sum_{i=1}^{n}\sum_{j=1}^{N}\mu_{ij}^{m}\left\lVert x_j - v_i\right\rVert^2}{N\min_{i,j}\left\lVert x_j - v_i\right\rVert^2}$	Min
Dunn's Index (DI)	$\min_{i\in c}\left\{\min_{j\in c, i\neq j}\left\{\frac{\min_{x\in C_i, y\in C_j} d(x,y)}{\max_{k\in c}\{\max_{x,y\in C} d(x,y)\}}\right\}\right\}$	Max
Alternative Dunn's Index (ADI)	$\min_{i\in c}\left\{\min_{j\in c, i\neq j}\left\{\frac{\min_{x_i\in C_i, x_j\in C_j}\left\lvert d(y,v_j) - d(x_i,v_j)\right\rvert}{\max_{k\in c}\{\max_{x,y\in C} d(x,y)\}}\right\}\right\}$	Max

CHAPTER 3

DESCRIPTIONS OF DATA SETS USED IN THE FUZZY CLASSIFICATION APPLICATIONS

In this study, three different data sets named as *customer satisfaction*, *casting* and *ionosphere* data sets are selected to be used in fuzzy classification applications. Descriptions of these data sets are presented in the following sections.

3.1. Customer Satisfaction Data Set

One of the data sets that we use for fuzzy classification applications is the customer satisfaction data set, which is a real life data set from the field of quality improvement. This data set was collected by Çabuk (2008), which included information about satisfaction levels of customers from a driver seat of a particular vehicle. The data set was preprocessed and then analyzed in the study of Çabuk (2008). For the purpose of improving the design of the driver seat, a "customer satisfaction classification model" was built using LR to predict how much a particular customer is likely to be satisfied with the driver seat.

In collecting the data, a survey was applied on 80 customers from predetermined customer segments. Customer segments used in the survey were determined by considering survey results conducted in 2006 by marketing department of the company, experts' advice, information from the after sale service of the company and studies about anthropometric measurements in literature (Çabuk, 2008). Since it was important to observe customers while using the vehicle, the survey was applied on the customers using the same seat from a particular vehicle in a selling and service center with a high operation volume, where customers come to buy a vehicle or to use advantages of service opportunities (Çabuk, 2008). The survey includes questions about the satisfaction levels of customers from the driver seat, demographic features, vehicle usage and anthropometric measurements.

In the study, overall satisfaction grade of customers was used as the dependent variable of the model. It was initially measured on a Likert scale with 7 levels, which can be seen in Table 3.1.

Table 3.1: Scale for Overall Satisfaction Grade

Very Bad	Bad	Somewhat Bad	Neutral	Somewhat Good	Good	Very Good
1	2	3	4	5	6	7

In the data preprocessing stage of the study, these seven levels were converted to two levels because of zero or low frequencies of levels 1, 2, 3 and 5, as seen from Figure 3.1. As a result of this conversion, somewhat bad, neutral and somewhat good levels were represented by "somewhat satisfied"; and good and very good levels were represented by "highly satisfied" as it is seen from Table 3.2.

Table 3.2: Previous and New Levels for Overall Satisfaction Grade

Previous Levels	**New Levels**	**Explanation**
1,2	Removed	No one selected
3,4,5	1	Somewhat satisfied
6,7	2	Highly satisfied

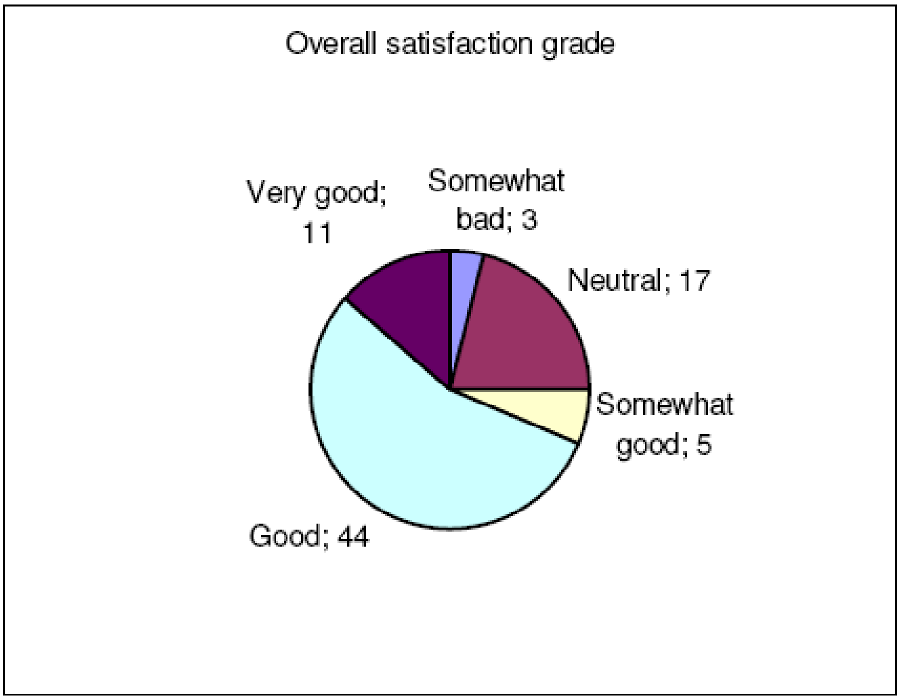

Figure 3.1: Frequency of the Answers

Finally, nine significant variables were determined as a result of a factor analysis followed by the stepwise procedure of LR (Çabuk, 2008). In our study, we have used these variables in developing the classification models.

The dependent variable of the data set, overall satisfaction level of a customer, is measured using linguistic terms such as “good” and “bad”. Because of this qualitative assessment, which may vary from person to person, fuzzy type of uncertainty is supposed to exist in the data. Fuzziness resulted from the data involving human judgments leads us to use fuzzy classification approaches to build a classification model for this data set as well.

3.2. Casting Data Set

Casting data set is also a real life data set provided by a manufacturing company from the metal casting industry, which was studied by Bakır (2007). The study aims to improve the quality of a product by determining the influential process variables that can cause defects and to optimize them by modeling the relationship between process variables and defect types in order to reduce the percentage of defective items. In the study, the data for a particular product, the cylinder head, was studied, which is seen as an important part because of its effect on the performance of another part, the internal combustion engine. After preprocessing of data, 36 process variables and 7 quality variables including continuous variables such as frequencies of total defective items or frequencies of defective items having certain defect types in a batch and binary variables such as existence of defective items or existence of defective items having certain defect types in a batch, for 92 cases chosen to form basic data set. In our study, we use the existence of nitrogen gas cavity defect types in a batch as dependent variable.

The process variables were taken as the average of the values measured in the day the observations are taken. Thus, the upper and lower values of the chosen value are also possible values for that process variable. Since, the chosen values of process variables are not the only possible value and since the possible values can take numerous values around this value, the fuzzy type of uncertainty is supposed to exist in the data. In addition, defective items are determined by a personal judgment in inspection process, which brings additional fuzzy type of uncertainty. For that reason, fuzzy classification methods are considered to be used for modeling the relationship between process variables and existence of defect types in a batch.

Since the usage of high number of independent variables in fuzzy methods causes some shortcomings such as increasing spreads for estimated outputs (Nasrabadi and Nasrabadi, 2004) , collinearity between variables and increase in calculation time (Wang and Tsaur, 2000), 36 process variables are decided to be reduced by using a variable selection procedure. However, the variable selection procedures developed for fuzzy methods is very limited in number and they are proposed for only specific fuzzy methods. Moreover, the data set is used for FCF and NIFCF methods, which employs conventional statistical methods after adding membership values as new independent variables to the model. Hence, variable selection procedure of a statistical method, LR is decided to be used for the selection of significant variables. By using LR forward stepwise procedure of *SPSS 9.0* with 0.15 entrance and 0.20 removal levels, 36 process variables are reduced to eight process variables. One can refer to Appendix A.1. for the implementation details of the selection procedure applied.

3.3. Ionosphere Data Set

Ionosphere data set from the area of physical sciences is taken from the UCI Machine Learning Data Repository web-site. Our desire to increase the reliability of the performance tests conducted for our developed NIFCF method leads us to use an additional data set from a different field. This data set aims to classify the radar returns from the ionosphere as either “good” or “bad”. Since the levels “good” and “bad” are the qualitative terms expressing human judgments, the data is thought to include fuzzy type of uncertainty as it is the case in customer satisfaction data.

The data is composed of 34 continuous independent variables and a binary dependent variable with an observation number of 351. Because of the same reasons mentioned in Section 3.2, 34 independent variables are reduced to 14 independent significant variables by using LR forward stepwise procedure of *SPSS 9.0* with 0.15 entrance and 0.20 removal levels. The *SPSS 9.0* output is given in Appendix A.2.

After variable selection procedure is carried out, all data sets described above are partitioned using a 3-fold and 3-replicate cross validation method in order to compare the performances of the methods applied in this study. According to this approach, the data is randomly (with stratification) divided into three parts three different times (replicates). For each replicate, classification models are developed, each time using two different parts (folds) of the data, and the models are tested on the third part using several classification measures.

CHAPTER 4

TANAKA BASED FUZZY CLASSIFICATION MODELS FOR CUSTOMER SATISFACTION DATA

In this chapter, we propose three alternative approaches for building classification models for the customer satisfaction survey data, based on Tanaka's FLR approach.

4.1. Alternative Approaches

As stated in Section 2.3.2.1, Tanaka's FLR approach is used for the prediction problems having fuzzy type of uncertainty. Thus, it is not appropriate to use this method directly for classification problems, which include discrete dependent variables. We show on a case problem that this widely used fuzzy regression approach can be used for classification purposes after the discrete dependent variable is converted into an equivalent continuous fuzzy variable.

The extensive literature on this subject shows that uncertainty imbedded in data can be interpreted in several ways and reflected in models via different ways as stated in Chapter 1. Three alternative approaches are developed by considering different types of uncertainties in data and interpreting and reflecting these uncertainties in data in different ways (Özer et al., 2009). These approaches are used for building classification models based on the customer satisfaction data set as explained in detail in Section 3.1 in order to predict the general satisfaction level of a particular customer. Based on this information, it is aimed to determine the factors that affect satisfaction levels of customers and as a result to improve the design of the driver seat to increase the general satisfaction levels of customers.

In order to compare the alternative methods developed, a 3-fold and 3-replicate cross validation approach is used. Then, some classification performance measures are calculated using the observed dependent variables and midpoints of predicted dependent variables of testing data sets, since the midpoints of symmetrical fuzzy triangular numbers are the points that best represent the given fuzzy interval (Wang and Tsaur, 2000).

4.1.1. Alternative 1 – Fuzzy Classification Model Based on The Dependent Variable with 7 Levels

This alternative approach proposes to apply Tanaka's FLR method using the overall satisfaction grade based on seven levels as the fuzzy dependent variable.

In this approach, we assume that the uncertainty comes from fuzziness. Since the overall satisfaction grade depends on human perception, there exists a fuzzy type of uncertainty (Zadeh, 2000). For example, while a person expresses his/her satisfaction for the driver seat as "somewhat good", another person having exactly the same sensation and experience from the seat, who is more perfectionist and demanding, can express his/her satisfaction as "neutral". Hence, the dependent variable can be considered as a fuzzy one.

As stated in Section 3.1, overall satisfaction grades expressed in linguistic terms such as "very bad" and "good" are represented by a numeric scale from one to seven. In this method, this discrete numeric scale is converted to a continuous scale, on which membership functions of overall satisfaction grades are constructed. It can be considered as reasonable because, a person, who chooses an expression indicating her/his overall satisfaction level, may be undecided between its former or latter levels and chooses the level which expresses her/his feelings best. However, if s/he were given an additional level between the levels s/he is undecided about, s/he would choose that one. For example, a person, who is undecided whether to express her/his overall satisfaction level with "somewhat good" or "good", can choose overall satisfaction the level, "somewhat good". However, this person may have neither "somewhat good" nor "good" satisfaction level. S/he may, instead, have a satisfaction level in between them, but s/he must choose one of these qualitative discrete levels, because s/he is asked to do so. If s/he were allowed to express her/his satisfaction grade with a continuous numeric scale from one to seven, s/he might express his/her overall satisfaction level by 5.5, which is between the "somewhat good" and "good" levels represented by five and six, respectively. As can be seen, oversimplification of data by the use of linguistic terms can cause some information loss and this loss can be compensated by the use of fuzzy membership functions (Chang and Ayyub, 2001). Thus, the membership functions of these satisfaction grades composed of linguistic terms are constructed on a continuous numeric scale from one to seven (see Figure 4.1).

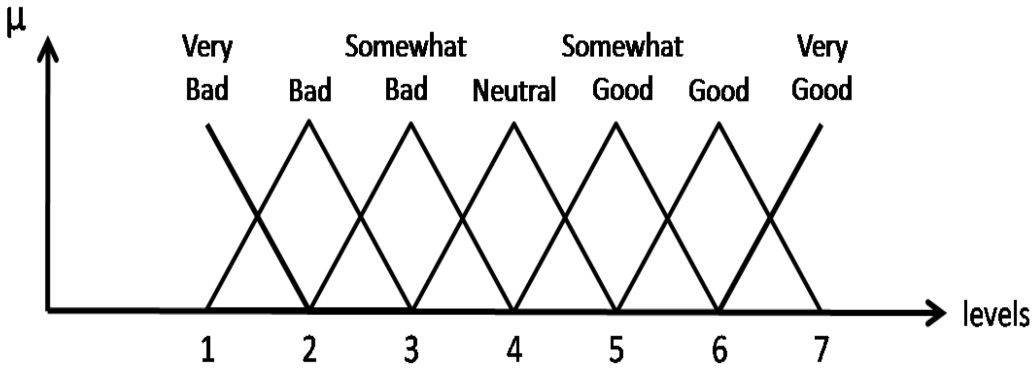

Figure 4.1: Membership Function of the Dependent Variable Measured Using 7 Levels

As can be seen from Figure 4.1, the membership function of an overall satisfaction level except for “very bad” and “very good” is symmetrical triangular fuzzy number bounded by its former and latter levels. However, membership functions representing “very good” and “very bad” levels are not symmetrical triangular fuzzy numbers, which is a requirement of Tanaka’s FLR method. Since the frequency of “very bad” level is zero and the frequency of “very good” level is low, they are also thought of as symmetrical triangular fuzzy numbers with spread one and the same LP model is used in order to overcome this problem without making it more complicated.

Since the discrete numerical scale used for representing qualitative overall satisfaction levels are converted to a continuous scale, the Tanaka’s FLR method can be used for building a classification model for the customer satisfaction data, which is given in Section 2.3.2.1.

Finally, the predicted class is determined for each observation by using the midpoint of the fuzzy predicted value obtained from the fuzzy linear regression model. The predicted overall satisfaction level is determined as the level, which has the highest membership value for the midpoint of the predicted continuous dependent variable. For example, an observation for which the midpoint of the predicted continuous dependent variable is 6.3 should be classified as “good” since its membership value for the level “good” is the greatest among the others. Thus, 1.5 and lower predicted values are expressed by “very bad” level, values between 1.5 and 2.5 are expressed by “bad” level, values between 2.5 and 3.5 are expressed by “somewhat bad” level and so on.

In order to be able to compare this classification model with the other classification models explained in the following sections, the predicted classes are converted to two levels, since the other models are based on a binary dependent variable. For this conversion, “very bad”, “bad”, “somewhat bad”, “neutral” and “somewhat good" overall satisfaction levels are represented by "somewhat satisfied" and "good" and "very good" overall satisfaction levels are represented by "highly satisfied", which is consistent with the conversion explained in Section 3.1.

4.1.2. Alternative 2 – Fuzzy Classification Model Based on Logistic Regression

This alternative approach proposes to apply Tanaka's FLR method using the posterior probabilities obtained from Logistic Regression (LR) as crisp dependent variables.

In this approach, we assume that the uncertainty comes from both fuzziness and randomness. As stated in the previous section, the fuzzy type of uncertainty imbedded in the customer satisfaction data set is resulted from the linguistic terms reflecting human judgments. In addition to fuzziness, another type of uncertainty, randomness is supposed to exist in the data since there can be factors that randomly vary in the process, which have some influences on the overall satisfaction level of the customer. For example, a person, who has skeletal problems such as hunchback or neck hernia, may have a low level of overall satisfaction, while another person having exactly the same demographic features, vehicle usage and anthropometric measurements but healthy skeletal structure may express her/his satisfaction level with "very good".

In order to take these types of uncertainties resulted from randomness into consideration, we benefit from the methods of probability theory. Hence, Tanaka's FLR approach is used to model the frequencies of customers choosing the level "highly satisfied" instead of modeling directly overall satisfaction levels as in Alternative 1, which is expected to capture the effect of randomness. While, frequencies are modeled by maximum likelihood estimation method in LR, which is a statistical method that depends on probability theory, this approach proposes to model frequencies by using a fuzzy regression method that depends on the possibility theory. Since our classification approach cannot handle repetitive measures, the frequencies of customers who choose the level “highly satisfied” are estimated by the posterior probabilities obtained from LR. After obtaining the posterior probabilities, Tanaka’s FLR method is applied using them as crisp dependent variable values. The LP model is as shown below:

$$\text{Min} \;\; J = \sum_{i=1}^{N} \sum_{j=0}^{M} c_j \left| x_{ij} \right| \tag{4.1}$$

s.t.

$$\sum_{j=0}^{M} m_j \, x_{ij} + (1-H) \sum_{j=0}^{M} c_j \left| x_{ij} \right| \geq \hat{p}_i(y_i = 1|\mathbf{x_i}) \;\; \text{for} \;\; i = 1, \ldots, N \tag{4.2}$$

$$\sum_{j=0}^{M} m_j \, x_{ij} - (1-H) \sum_{j=0}^{M} c_j \left| x_{ij} \right| \leq \hat{p}_i(y_i = 1|\mathbf{x_i}) \;\; \text{for} \;\; i = 1, \ldots, N \tag{4.3}$$

$$c_j \geq 0, \;\; m_j \text{ free} \qquad\qquad \text{for } j = 0, \ldots, M \tag{4.4}$$

where the variable

$\hat{p}_i(y_i = 1|\mathbf{x_i})$: LR estimate for the probability that the i^{th} customer is highly satisfied with the driver seat for the given input vector, $\mathbf{x_i}$, about him/her.

As a result, while the fuzzy uncertainty in the data is reflected in the model with fuzzy regression coefficients, the uncertainty resulted from randomness is reflected by the use of posterior probabilities.

In this alternative, two classes, “highly satisfied” and “somewhat satisfied”, are used for both of the training and testing data sets. The predicted classes are determined by using the midpoints of the predicted fuzzy dependent variable values. If the midpoint of the prediction is greater than 0.5, then the observation is classified as “highly satisfied”, otherwise as “somewhat satisfied”.

4.1.3. Alternative 3 - Fuzzy Classification Model Based on FCM

This alternative proposes to apply Tanaka's FLR method using the membership values obtained from FCM clustering algorithm as crisp dependent variables.

In this approach, we assume that the uncertainty comes from both fuzziness and randomness as in Alternative 2. It evaluates the type of uncertainty in the data in a way that is similar to Alternative 2, in which, fuzzy uncertainty is considered to be resulted from the human judgments and randomness is resulted from the factors randomly varying in the process. Recall that Alternative 2 proposes to apply Tanaka's FLR approach using the posterior probabilities obtained from LR in order to cope with these uncertainties. However, it may be inappropriate to use these posterior probabilities directly as dependent variable values in Tanaka's FLR method. Since these posterior probabilities are estimated using maximum likelihood estimation, they may not be as accurate and precise as estimates obtained directly from a large number of repetitive observations at the same levels of independent variables. Hence, we think that this imprecision brings additional uncertainty to the data, which should be handled. Although one may choose to handle it using probability theory (distribution of the estimation error), we treat it as a fuzzy type of uncertainty.

For this purpose, an approach similar to that of Bolotin (2006) is adopted. Fuzziness in data is captured by membership values replacing the indicator variables in the study of Bolotin (2005). Similarly, we propose to use membership values of posterior probabilities obtained from LR as crisp dependent variable values in Tanaka's FLR method. So, this alternative can be seen as the mixture of two methods, Tanaka's FLR method and Bolotin's method. While Bolotin's method uses the least squares regression to estimate the membership values of indicator variables, our approach uses Tanaka's FLR method. The membership values of posterior probabilities for "somewhat satisfied" and "highly satisfied" levels are determined by using the FCM algorithm. In applying the FCM algorithm, number of clusters is taken as 2, which are assumed to represent "highly satisfied" and "somewhat satisfied" levels, and degrees of fuzziness are determined by using several validity measures such as separation index, Xie-Beni index and so on (see Section 2.4.3). As a result of the application of the FCM algorithm, the membership functions of "somewhat satisfied" and "highly satisfied" levels are obtained, which looks similar to Figure 4.2.

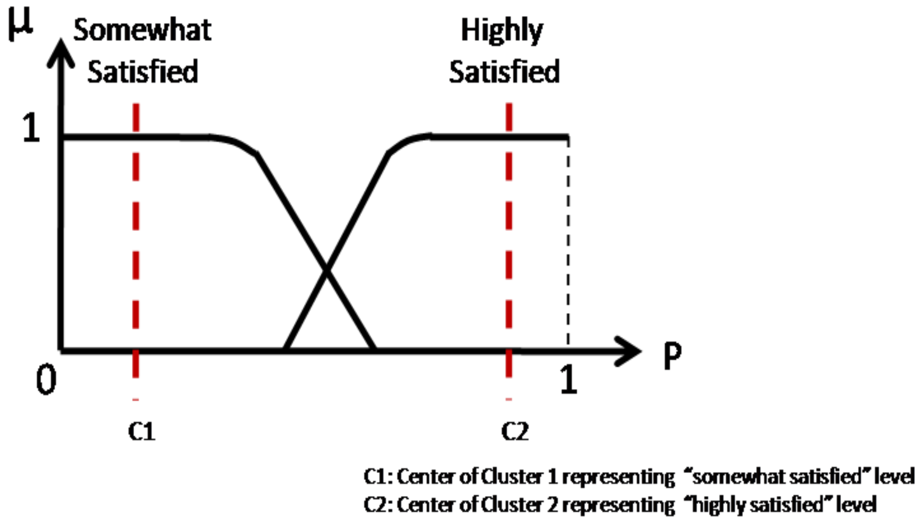

Figure 4.2: Membership functions of "somewhat satisfied" and "highly satisfied" levels

Then, Tanaka's FLR method is applied using the membership values of "highly satisfied" level as dependent variable values, since the membership values of "highly satisfied" and "somewhat satisfied" levels sum up to one. Hence, in this alternative, possibility of overall satisfaction is tried to be modeled rather than the probability of satisfaction as in Alternative 2. The LP model of this approach is given below:

$$\text{Min } J = \sum_{i=1}^{N}\sum_{j=0}^{M} c_j \left|x_{ij}\right| \tag{4.5}$$

s.t.

$$\sum_{j=0}^{M} m_j\, x_{ij} + (1-H)\sum_{j=0}^{M} c_j \left|x_{ij}\right| \geq \mu\{\hat{p}_i(y_i = 1|\mathbf{x_i})\} \text{ for } i = 1, \dots, N \tag{4.6}$$

$$\sum_{j=0}^{M} m_j\, x_{ij} - (1-H)\sum_{j=0}^{M} c_j \left|x_{ij}\right| \leq \mu\{\hat{p}_i(y_i = 1|\mathbf{x_i})\} \text{ for } i = 1, \dots, N \tag{4.7}$$

$$c_j \geq 0,\ m_j \text{ free} \qquad \text{for } j = 0, \dots, M \tag{4.8}$$

where the variable

$\mu\{\hat{p}_i(y_i = 1|\mathbf{x_i})\}$: membership value of the posterior probability $\hat{p}_i(y_i = 1|\mathbf{x_i})$ obtained from LR for the i^{th} observation.

Finally, the predicted membership values are obtained as fuzzy numbers with midpoints and spreads. If the midpoint is greater than 0.5, the observation is classified as “highly satisfied”, otherwise as “somewhat satisfied”. As can be seen, in this alternative, we classify the observations according to membership values rather than probabilities.

4.2. Discussion and Performance Analysis

Three alternatives described in the previous sections approach the classification problem from different ways by considering the types of uncertainties and interpreting them in the data in alternative ways. While in Alternative 1, the uncertainty is assumed to be resulted from fuzziness, other alternatives consider the type of uncertainties in the data as both fuzziness and randomness as stated in previous sections. When we compare Alternative 1 with the other alternatives, Alternatives 2 and 3 are expected to outperform Alternative 1, since they also take the other dimension of uncertainty, randomness, into account, which is included in the data inherently as fuzziness.

Alternatives 2 and 3 are similar approaches, but Alternative 3 takes into account the fuzziness, which is brought by the imprecision resulted from the use of estimated frequencies in the model, in addition to other types of uncertainties that are common with those of Alternative 2. Since Alternative 3 copes with all the uncertainties in the data, better results can be expected from Alternative 3. However, when we look at the posterior probabilities obtained from LR and membership values calculated by the FCM algorithm, we see that they are very close to each other, which is considered as expected, since membership values take values between zero and one as probabilities and their values are expected to increase with the increase in probabilities. So, there may not be any significant difference between the performances of Alternative 2 and Alternative 3.

All LP models are coded in *MATLAB 7.3.0* using functions from its optimization toolbox. The midpoints and spreads of fuzzy regression coefficients are calculated for each training data set at H=0.1 target degree of belief. H level is determined as 0.1 according to the sufficiency of the customer satisfaction survey data. Recall that if the data set collected is sufficiently large, H level should be determined as 0 and it should be increased with the decreasing volume of the data set (Tanaka and Watada, 1988 as cited in Wang and Tsaur, 2000). Since 80 observations exist for 9 selected independent variables, the volume of data set can be seen as almost sufficient. Thus, a small value for H level is decided to be selected. In fact, as mentioned in Section 2.3.2.1, midpoints, which are used for the determination of predicted classes, are not affected by the H value. However, spread values increase with the increase in H value (Tanaka and Watada, 1988 as cited in Kim, Moskowitz and Köksalan, 1996). The spread values increase by inversely proportional to (1-H) value. By determining 0.1 target degree of belief, it is expected to have smaller spreads with respect to the solutions obtained using larger target degree of belief values.

The performances of these alternative approaches are tested using several classification performance measures such as MCR, KAPPA and AUC for each testing data set. (see Section 2.2.2). The average values of the classification performance measures for the nine testing data sets can be seen in Table 4.1.

Table 4.1: Average Classification Measures of the Alternative Approaches

Measure (Avg.)	Alternative 1	Alternative 2	Alternative 3
MCR	0.3804	0.2094	0.2179
PCC	0.6196	0.7906	0.7821
KAPPA	0.6002	0.7819	0.7717
Precision	0.4444	0.6844	0.6572
Recall	0.8750	0.5972	0.6667
Specificity	0.5062	0.8765	0.8333
$F_{0.5}$	0.4918	0.6567	0.6505
F_1	0.5866	0.6264	0.6489
F_2	0.7296	0.6062	0.6564
Log-Odds Ratio	1.7141	2.3521	2.3333
Stability	0.0022	0.0471	0.0480
AUC	0.6772	0.7255	0.7368

It can be seen from Table 4.1 that the average application results of Alternatives 2 and 3 are very similar to each other and they outperform Alternative 1 according to many of the performance measures such as MCR, PCC, AUC, KAPPA and precision, as expected.

These 3 alternatives are statistically compared by using one-way ANOVA in order to see whether there is a statistically significant difference between these alternatives using *Minitab 15*. Firstly, residual plots are controlled to check the assumptions of ANOVA. For the performance measures that the assumptions of ANOVA are not satisfied, logarithmic transformations of them are used. The one-way ANOVA results for each classification measure can be seen in Table 4.2, where DF, SS, MS, F and P stand for degrees of freedom, sum of squared error, mean squared error, F value and P value, respectively.

Since the sum of MCR ad PCC is one as indicated in Section 2.2.2, MCR and PCC performance measures result in same ANOVA results. As can be seen from Table 4.2, three alternative approaches significantly differ from each other with respect to the measures MCR, PCC, KAPPA, precision, recall, specificity and $F_{0.5}$ at the significance level of α=0.01.

Table 4.2: One-Way ANOVA Results for Each Classification Measure

Measure	Source	DF	SS	MS	F	P
MCR, PCC	Method	2	0.1670	0.0835	21.4700	0.0000
	Error	24	0.0934	0.0039		
	Total	26	0.2603			
KAPPA	Method	2	0.1876	0.0938	22.2100	0.0000
	Error	24	0.1014	0.0042		
	Total	26	0.2890			
Precision	Method	2	5.9872	2.9936	12.6800	0.0000
	Error	24	5.6645	0.2360		
	Total	26	11.6517			
Recall	Method	2	11.3336	5.6668	6.2300	0.0070
	Error	24	21.8334	0.9097		
	Total	26	33.1670			
Specificity	Method	2	0.7382	0.3691	54.7300	0.0000
	Error	24	0.1618	0.0067		
	Total	26	0.9000			
$F_{0.5}$	Method	2	2.9624	1.4812	8.6000	0.0020
	Error	24	4.1315	0.1721		
	Total	26	7.0939			
F_1	Method	2	0.3925	0.1963	0.9100	0.4140
	Error	24	5.1484	0.2145		
	Total	26	5.5409			
F_2	Method	2	1.0170	0.5085	1.0400	0.3690
	Error	24	11.7527	0.4897		
	Total	26	12.7696			
Log-Odds Ratio	Method	2	0.1780	0.0890	0.0500	0.9540
	Error	24	45.5080	1.8960		
	Total	26	45.6860			
Stability	Method	2	0.0124	0.0062	1.7900	0.1880
	Error	24	0.0827	0.0034		
	Total	26	0.0951			
AUC	Method	2	0.0181	0.0090	1.6800	0.2070
	Error	24	0.1288	0.0054		
	Total	26	0.1469			

A multiple comparison analysis is applied in order to see which alternative method performs better than the others. The alternative approaches are compared using Tukey's multiple comparison test. The test is conducted using "ANOVA with General Linear Model (GLM)" tool of *Minitab 15* with the family error rate of 0.05. The p-values of Tukey's multiple comparison test for the classification measures MCR, PCC, KAPPA, precision, recall, specificity, $F_{0.5}$ and AUC, in which methods significantly differ, are given in Table 4.3. Notations, ">" and "<" indicate that the method in the column list shows higher and lower performance than the method in the row list, respectively.

Table 4.3: p-values of Tukey's Multiple Comparison Test

Measures	Alternative Approaches	Alt 2	Alt 3
MCR, PCC	Alt 1	0.0000 (>)	0.0000 (>)
	Alt 2		0.9547
KAPPA	Alt 1	0.0000 (>)	0.0000 (>)
	Alt 2		0.9416
Precision	Alt 1	0.0003 (>)	0.0015 (>)
	Alt 2		0.7929
Recall	Alt 1	0.0077 (<)	0.0329 (<)
	Alt 2		0.8046
Specificity	Alt 1	0.0000 (>)	0.0000 (>)
	Alt 2		0.5138
$F_{0.5}$	Alt 1	0.0029 (>)	0.0059 (>)
	Alt 2		0.9522

According to Table 4.3, Alternatives 2 and 3 significantly outperform Alternative 1 according to measures MCR, PCC, KAPPA, precision, specificity and $F_{0.5}$. The only measure that Alternative 1 has better performance than the others is recall, which measures the proportion of the customers highly satisfied with the driver seat, which are correctly identified. However, it can be neglected since Alternatives 2 and 3 give significantly better performance results than Alternative 1 according to F_1measure, which weights precision twice as much as recall. As stated in Section 2.1.2, when there is a tradeoff between recall and precision as in this case, F measure gives more valuable information about test's accuracy, since it considers both recall and precision by calculating their weighted harmonic mean.

However, there is not any statistically significant difference found between Alternatives 2 and 3, as expected. The use of membership values obtained by the FCM algorithm in Alternative 3 rather than probabilities as in Alternative 2 does not bring any significant improvement to the classification performance. It may be because of that the use of membership values in order to cope with the additional fuzziness resulted from the use of posterior probabilities predicted by LR may not be necessary, since the fuzziness included in posterior probabilities might be negligible. In other words, Tanaka's FLR approach can be sufficient on its own to overcome both types of fuzzy uncertainties in the data and the additional operations may not be needed to cope with fuzzy uncertainties.

According to Hojati et al. (2005), the fuzzy number spreads should be both narrow enough for the ease of use and wide enough to contain maximum number of observations. Thus, Alternatives 2 and 3 are also decided to be compared using their spread values in addition to the classification performance measures mentioned above, according to which they do not differ significantly. As it is seen from the LP models used for Tanaka's FLR analysis given in Section 2.3.2.1, the objective is to decrease total fuzziness, which is described by sum of half widths of the predicted fuzzy intervals or sum of half widths of the fuzzy regression coefficients while giving fuzzy intervals wide enough to contain all of the observations. Thus, in order to evaluate the spread values of the alternatives, we use the sum of half widths of the predicted fuzzy intervals and sum of half widths of the fuzzy regression coefficients as stated in the study of Hojati et al. (2005). The average values of these measures for nine testing data sets are given in Table 4.4.

Table 4.4: Total Fuzziness Measures of Alternative 2 and Alternative 3

Measures	Alt 2	Alt 3
Sum of half-widths of the fuzzy predicted intervals	5.5	17.20
Sum of half-widths of the fuzzy regression coefficients	0.48	0.78

As it is seen from Table 4.4, Alternative 2 gives FLR models that have less total fuzziness than Alternative 3. In other words, Alternative 2 results in narrower spreads for both fuzzy predicted intervals and fuzzy regression coefficients as desired. Thus, it can be inferred that the use of membership values obtained by the FCM algorithm in Alternative 3 rather than probabilities as in Alternative 2 increases the total fuzziness of the model. As a result, it can be said that Alternative 2 has better performance than Alternative 3 according to the total fuzziness measures while its performance is very similar to Alternative 3 according to classification performance measures.

These three alternative approaches developed for a particular case of building customer satisfaction classification models can also be used for other similar classification problems. In these problems, it is required that the independent variables are crisp and the dependent variable can be treated as a fuzzy one. For some classification problems, uncertainty in the dependent variable is only resulted from the use of linguistic terms, qualitative, and fuzzy data, and there is not any type of uncertainty resulted from the factors randomly varying in the process. For these classification problems, Alternative 1 can be used after the proper conversion of crisp dependent variable to a continuous fuzzy dependent variable in a similar manner explained in the above case. However, in some cases, it may not be preferred to use the dependent variable as a fuzzy number, for example, when the observed dependent variable values are crisp numbers and they are not preferred to be fuzzified. In this case, the dependent variable can be treated as a crisp one the left and right hand side spreads of which are zero.

In many cases, fuzzy and random types of uncertainties coexist in the data. Thus, both types of uncertainties, fuzziness and randomness, should be handled for proper modeling of these classification problems. Alternative 2 and Alternative 3 propose an appropriate framework for these classification problems since they both utilize the tools of probability theory and fuzzy modeling approach, Tanaka's FLR approach to handle both types of uncertainties imbedded in the data. For the cases in which frequencies of occurrence of an event are obtained directly from large numbers of repetitive observations at the same levels of the independent variables, Alternative 2 can be used without the use of LR. In this case, the frequencies are used as crisp dependent variable values and Tanaka's method for crisp dependent variables is applied. However, when these frequencies do not exist, Alternative 2 and Alternative 3 can be applied after predicting frequencies from LR. Alternative 3 aims to handle the additional type of uncertainty resulted from the use of predicted probabilities, which is treated as fuzziness in our case. However, performance analysis results of these alternative approaches show that the use of Alternative 3 to handle this additional fuzzy type of uncertainty does not bring any significant improvement for our case. However, in other cases Alternative 3 might be worth trying.

CHAPTER 5

NONPARAMETRIC IMPROVED FUZZY CLASSIFIER FUNCTIONS

In this chapter, a method, called Nonparametric Improved Fuzzy Classifier Function (NIFCF) is proposed. It presents an improvement of the IFCF approach. Its performance is compared with another fuzzy classification method, FCF and a statistical classification method, LR.

5.1. Motivation

As stated in Section 2.3.3.4, the IFCF method is the improved version of FCF method, which uses the IFC algorithm rather than the FCM algorithm. In the IFC clustering algorithm, the prediction power of the membership values are tried to be increased by adding the term, squared deviation between the actual output and the estimated output of the model constructed by using only the membership values and their transformations as input variables, to the objective function of the FCM algorithm.

$$J_m^{IFC} = \sum_{i=1}^{n} \sum_{k=1}^{N} \mu_{ik}^{m}\, d_{ik}^{2} + \sum_{i=1}^{n} \sum_{k=1}^{N} \mu_{ik}^{m}\, (y_k - f(\boldsymbol{\tau_{ik}}))^2 \tag{5.1}$$

where

μ_{ik} : membership value of the k^{th} observation for the i^{th} cluster,

d_{ik} : distance value of the k^{th} observation for the i^{th} cluster,

y_k : actual observed class of the k^{th} observation,

$f(\boldsymbol{\tau_{ik}})$: predicted value calculated for the k^{th} observation and the i^{th} cluster for a given input vector, $\boldsymbol{\tau}_{ik}$,

$\boldsymbol{\tau_{ik}}$: row vector, the elements of which are the membership values of the k^{th} observation for the i^{th} cluster and their chosen transformations,

m: the degree of fuzziness

n: number of clusters

N: number of observations

Using an algorithm, this function is minimized to specify variables, μ_{ik} subject to the constraints

$\sum_{i=1}^{n} \mu_{ik} = 1$

$0 < \sum_{k=1}^{N} \mu_{ik} < N$

$\mu_{ik} \in [0, 1]$

According to this algorithm explained in Section 2.4.2, in order to be able to partition data into clusters using the IFC algorithm, a model should be fitted at each iteration in a loop to calculate the distance values. If a model cannot be formed at any iteration of the loop, the algorithm is terminated and the data cannot be clustered. However, it may not always be possible to fit a model using a parametric classification method. A parametric method may not able to find a significant relationship between the response and predictor variables and so, fit a model, since the output variable is tried to be modeled by using only membership values and their transformations rather than actual predictors of the data in the clustering phase of IFC.

When we try to apply the IFC algorithm using LR as a classifier, we face with fitting problems at the clustering phase of the algorithm as mentioned above. After several trials of IFC application using different membership transformation matrices, number of clusters and degree of fuzziness levels, we can achieve clustering for a very limited number of training data sets. In addition, the determination of optimized values of parameters by assessing the values of validity indices at different levels of parameters cannot be performed since the data cannot be clustered for each value of parameters.

In order to overcome all of these fitting problems encountered, we propose to use a non-parametric method, MARS, in the clustering phase of the IFC method, which automates the model formation and selection of transformations of predictors as well as the selection of variables to find a best model fit. By the use of MARS, it is intended to achieve clustering of data for every level of parameters in order to be able to perform both selections of optimum model parameters and achieve the method application, which needs the outputs of the IFC clustering algorithm as input data. The clustering method, which proposes to use a nonparametric method, MARS, as a classifier, and the fuzzy classifier method, which proposes to use this method as a clustering algorithm are called Nonparametric Improved Fuzzy Clustering (NIFC) and Nonparametric Improved Fuzzy Classifier Function (NIFCF), respectively (Özer et al., 2009).

5.2. The Method

As a starting phase of the NIFCF algorithm, clustering should be performed using the NIFC clustering algorithm, the steps of which are given in detail below.

Steps of the NIFC algorithm:

1. Set the initial parameters,

   n : number of clusters,

   m : degree of fuzziness,

   ε : termination constant,

   n_{max}: maximum number of iterations.

2. Standardize the data.

3. Apply FCM algorithm to find initial membership values $\boldsymbol{f_i}^{(0)}$ and cluster centers $\boldsymbol{v_i}^{(0)}$ for the i^{th} cluster.

4. For each iteration $t = 1, \dots, n_{max}$

   4.1. For each cluster $i = 1, \dots, n$

   4.1.1. Construct interim fuzzy classifier function, $f(\boldsymbol{\tau_{ik}})$, using actual outputs, y_k as dependent variables and membership values of the k^{th} observation for the i^{th} cluster, $f_{ik}^{(t-1)}$ calculated from previous iteration, *(t-1)* as independent variables with a nonparametric method, MARS.

   4.1.2. For each observation $k = 1, \dots, N$

   4.1.2.1. Calculate squared error term $SE_{ik}^{(t)}$ using the interior fuzzy function calculated at step 3.1.1.

$$SE_{ik}^{(t)} = [y_k - f(\boldsymbol{\tau_{ik}})]^2 \tag{5.2}$$

where

y_k : actual observed class of the k^{th} observation,

$f(\boldsymbol{\tau_{ik}})$: predicted value calculated for the k^{th} observation and the i^{th} cluster for a given input vector, $\boldsymbol{\tau_{ik}}$,

$\boldsymbol{\tau_{ik}}$: row vector the elements of which are the membership values of the k^{th} observation for the i^{th} cluster, $f_{ik}^{(t-1)}$ calculated from the previous stage, *(t-1)* and their transformations selected by MARS.

4.1.2.2. Calculate the distance values, $d_{ik}^{(t)}$

$$d_{ik}^{(t)} = \|(\boldsymbol{x_k}, y_k) - \boldsymbol{v_i^{(t-1)}}(\mathbf{X}\mathrm{Y})\|^2 + SE_{ik}^{(t)} \tag{5.3}$$

where

$\boldsymbol{v_i^{(t-1)}}(\mathbf{X}\mathrm{Y})$: the i^{th} cluster center of the variables calculated from previous iteration *(t-1)* for input-output matrix, **X**Y,

$(\boldsymbol{x_k}, y_k)$: row vector of input-output values for k^{th} observation.

4.2. Calculate the membership values, $f_{ik}^{(t)}$

$$f_{ik}^{(t)} = \left(\sum_{j=1}^{n}\left[\frac{d_{ik}^{(t)}}{d_{jk}^{(t)}}\right]^{1/(m-1)}\right)^{-1} \tag{5.4}$$

4.3. For each variable $j = 1, \dots, p$

4.3.1. For each cluster $i = 1, \dots, n$

4.3.1.1. Calculate the i^{th} cluster center, $v_{ij}^{(t)}$ for the j^{th} variable

$$v_{ij}^{(t)} = \frac{\sum_{k=1}^{N} f_{ik}^{m} x_{jk} y_k}{\sum_{k=1}^{N} f_{ik}^{m}} \tag{5.5}$$

where

x_{jk} : value of the j^{th} variable for the k^{th} observation,

y_k : actual observed class of the k^{th} observation,

4.4. Terminate if the maximum change between the membership values calculated at iteration, t and previous iteration *(t-1)* does not exceed termination constant, ε:

$$max\left(f_{ik}^{(t)} - f_{ik}^{(t-1)}\right) < \varepsilon$$

After the data is partitioned into clusters using the NIFC method, NIFCF method is applied. In this method, one classifier for each cluster is constructed by LR using membership values obtained from NIFC and their transformations as new independent variables. The algorithm of the NIFCF method with LR, which can be used for the training data sets, can be seen below.

Steps of training algorithm for NIFCF:

1. Set initial parameter, α, which is the level used for eliminating the points farther away from the cluster centers.

2. Calculate cluster centers for input-output variables, $\boldsymbol{v}(\boldsymbol{X}Y)_{\boldsymbol{i}}$ and interim fuzzy functions for each cluster using NIFC algorithm.

$$\boldsymbol{v}(\boldsymbol{X}Y)_{\boldsymbol{i}} = \left\{v(x_1)_i, \dots, v\left(x_p\right)_i, v(y)_i\right\}$$

where

$v(x_j)_i$:cluster center of the j^{th} independent variable for the i^{th} cluster,

$v(y)_i$:cluster center of dependent variable for the i^{th} cluster,

3. For each cluster $i = 1, \dots, n$

3.1. For each observation number $k = 1, \dots, N$

Using cluster centers for input space, $\boldsymbol{v(X)_i} = \{v(x_1)_i, \dots, v(x_p)_i\}$

3.1.1. Calculate membership values for input space, u_{ik}.

$$u_{ik} = \left(\sum_{j=1}^{n} \left[\frac{\|\boldsymbol{X_k} - \boldsymbol{v(X)_i}\|^2 + SE_{ik}}{\|\boldsymbol{X_k} - \boldsymbol{v(X)_j}\|^2 + SE_{jk}} \right]^{\frac{1}{m-1}} \right)^{-1} \tag{5.6}$$

where

SE_{ik}: squared error term between the actual output and predicted output value of the k^{th} observation using interim fuzzy function calculated for the i^{th} cluster at step 2.

3.1.2. Calculate alpha-cut membership values, μ_{ik}.

$$\mu_{ik} = \{u_{ik} \geq \alpha\} \tag{5.7}$$

3.1.3. Calculate normalized membership values, γ_{ik}.

$$\gamma_{ik} = \frac{\mu_{ik}}{\sum_{j=1}^{n} \mu_{jk}} \tag{5.8}$$

3.2. Determine the new augmented input matrix for each cluster *i*, $\mathbf{\Phi_i}$, using observations selected according to α-cut level. $\mathbf{\Phi_i}$ matrix is composed of input variable matrix, $\mathbf{X_i^\alpha}$, vector of normalized membership values for the cluster *i*, $\boldsymbol{\gamma_i}$, and the matrix composed of their selected transformations, $\boldsymbol{\gamma_i}'$, such as $\boldsymbol{\gamma_i}^2, \boldsymbol{\gamma_i}^3, \boldsymbol{\gamma_i}^m$, exp($\boldsymbol{\gamma_i}$), log(($\mathbf{1}$-$\boldsymbol{\gamma_i}$)/ $\boldsymbol{\gamma_i}$).

$$\mathbf{\Phi_i}(\mathbf{X}, \boldsymbol{\gamma_i}) = [\mathbf{X_i^\alpha} \quad \boldsymbol{\gamma_i} \quad \boldsymbol{\gamma_i}']$$

where

$$\mathbf{X_i^\alpha} = \{\boldsymbol{x_k} \in \mathbf{X} \,|u_{ik}(\boldsymbol{x_k}) \geq \alpha, k = 1, \dots, N\}$$

3.3. Using LR as a classifier, calculate a fuzzy classifier function using new augmented matrix $\mathbf{\Phi_i}(\mathbf{X}, \boldsymbol{\gamma_i})$.

$$P_i\left(y_i^\alpha = 1 \middle| \mathbf{\Phi_i}(\mathbf{X}, \boldsymbol{\gamma_i})\right) = 1/\left(1 + e^{-\beta_0 + \boldsymbol{\beta^T} \mathbf{\Phi_i}(\mathbf{X}, \boldsymbol{\gamma_i})}\right) \tag{5.9}$$

where

$$y_i^\alpha = \{y_k \in \mathbf{y} \,|u_{ik}(y_k) \geq \alpha, k = 1, \dots, N\},$$

$\boldsymbol{\beta}$: vector of estimated regression coefficients,

β_0: estimated regression coefficient for the intercept.

3.3.1. For each observation $k = 1, \dots, N$

3.3.1.1. Using the fuzzy classifier function constructed at step 3.3, calculate posterior probabilities, $\hat{p}_{ik}\left(y_k^\alpha = 1 \middle| \mathbf{\Phi_i}(\mathbf{X}, \boldsymbol{\gamma_i})\right)$.

4. For each observation $k = 1, \dots, N$

4.1. Calculate a single probability output $\hat{p}_k$, weighting the posterior probabilities, $\hat{p}_{ik}$, with their corresponding membership values, γ_{ik}.

$$\hat{p}_k = \frac{\sum_{i=1}^{n} \gamma_{ik} \hat{p}_{ik} \left(y_k^{\alpha} = 1 \middle| \mathbf{\Phi_i}(\mathbf{X}, \boldsymbol{\gamma_i}) \right)}{\sum_{i=1}^{n} \gamma_{ik}} \tag{5.10}$$

After construction of classification models for each cluster using training data set, the following algorithm is used to classify new data.

Steps of testing algorithm for NIFCF:

1. Set initial parameter k, which is a positive integer used for k-nearest neighbor algorithm.

2. Standardize testing data.

3. For each observation $r = 1, \dots, N^{test}$

   3.1. Find k-nearest neighbors from training data set, the Euclidean distances to corresponding testing data r and actual output values of which are represented by vectors $\boldsymbol{d_r}$ and $\boldsymbol{y_r}$:

   $$\boldsymbol{d_r} = [d_{r1} \dots d_{rj} \dots d_{rk}]$$

   $$\boldsymbol{y_r} = [y_{r1} \dots y_{rj} \dots y_{rk}]$$

   where

   d_{rj} : distance of the r^{th} observation to its j^{th} nearest neighbor,

   y_{rj} : actual output value of the j^{th} nearest neighbor of the r^{th} testing data observation.

   3.2. For each cluster $i = 1, \dots, n$

   3.2.1. For each nearest neighbor $j = 1, \dots, k$

3.2.1.1. Calculate squared error term, SE_{rij} .

$$SE_{rij} = \left[y_{rj} - f(\boldsymbol{\tau_{rij}})\right]^2 \tag{5.11}$$

where

$f(\boldsymbol{\tau_{rij}})$: predicted value of the j^{th} nearest neighbor of the r^{th} testing data vector for a given input vector, τ_{rij} , which is calculated at 4.1.2.1. step of the NIFC algorithm,

$\boldsymbol{\tau_{rij}}$: row vector, the elements of which are membership values of the j^{th} nearest neighbor of the r^{th} testing data vector for the i^{th} cluster.

3.2.1.2. Calculate weights, η_{rj} of the nearest neighbors by

$$\eta_{rj} = 1 - (d_{rj} / \sum_{s=1}^{k} d_{rs}) \tag{5.12}$$

3.2.2. Calculate weighted squared error term, SE_{ir}^{test} for testing data.

$$SE_{ir}^{test} = \sum_{q=1}^{k} \eta_{iq} SE_{riq} \tag{5.13}$$

3.2.3. Calculate improved membership values, u_{ir}^{test} .

$$u_{ir}^{test} = \left(\sum_{j=1}^{n} \left[\frac{\|\boldsymbol{x_r}^{test} - v(\boldsymbol{X})_i\|^2 + SE_{ir}}{\|\boldsymbol{x_r}^{test} - v(\boldsymbol{X})_j\|^2 + SE_{jr}} \right]^{\frac{1}{(m-1)}} \right)^{-1} \tag{5.14}$$

where

$\boldsymbol{x_r}^{test}$: testing data input vector for the r^{th} observation,

$v(\boldsymbol{X})_i$: the i^{th} cluster centers for input variables calculated using the NIFC algorithm.

3.2.4. Calculate alpha-cut membership values, μ_{ir}^{test}.

$$\mu_{ir}^{test} = \{u_{ir}^{test} \geq \alpha\} \tag{5.15}$$

3.2.5. Calculate normalized membership values, ${\gamma_{ir}}^{test}$.

$$\gamma_{ir}^{test} = \frac{\mu_{ir}^{test}}{\sum_{q=1}^{n} \mu_{qr}^{test}} \tag{5.16}$$

3.2.6. Determine the new augmented input vector, $\mathbf{\Phi_{ir}^{test}}$ which is composed of testing data input vector for the r^{th} observation, ${\boldsymbol{x_r}}^{test}$, normalized membership value of the r^{th} observation for the i^{th} cluster, γ_{ir}^{test}, and the vector composed of their transformations, $[\boldsymbol{\gamma_{ir}^{test}}]'$ used at the 3.2. step of training data algorithm of the NIFCF.

$$\mathbf{\Phi_{ir}^{test}}({\boldsymbol{x_r}}^{test}, \gamma_{ir}^{test}) = \begin{bmatrix}\mathbf{x_r}^{test} & \gamma_{ir}^{test} & [\boldsymbol{\gamma_{ir}^{test}}]'\end{bmatrix}$$

3.2.7. Using the fuzzy classifier function constructed at step 3.3 of the NIFCF training data algorithm, calculate posterior probabilities, $\hat{p}_{ir}^{test}(y_r^{test} = 1/\ \boldsymbol{\Phi_{ir}^{test}})$.

3.3. Calculate a single probability output $\hat{p}_r^{test}$, weighting the posterior probabilities, $\hat{p}_{ir}^{test}$, with their corresponding membership values, γ_{ir}^{test}.

$$\hat{p}_r^{test} = \frac{\sum_{r=1}^{n} \gamma_{ir}^{test}\, \hat{p}_{ir}^{test}\ (y_r^{test} = 1/\ \boldsymbol{\Phi_{ir}^{test}})}{\sum_{i=1}^{n} \gamma_{ir}^{test}} \tag{5.17}$$

All algorithms are coded in *MATLAB 7.3.0.* For the clustering part of the algorithm, functions of *Fuzzy Clustering and Data Analysis Toolbox* are used. The best models constructed by MARS are automatically read from *R 2.8.1* by using *MATLAB R-Link Toolbox.*

5.3. Applications and Performance Analysis

NIFCF method is applied on three data sets: customer satisfaction, casting and ionosphere, which are explained in detail in Chapter 3, and its performance is compared with FCF and LR methods.

5.3.1. Applications

Before the application of FCF and NIFCF methods, the data have to be fuzzy partitioned using the FCM and NIFC algorithms, respectively. However, initial parameters have to be determined before starting clustering. The optimal number of clusters and degree of fuzziness is determined in a way that is stated in the *Handbook of Fuzzy Clustering and Data Analysis Toolbox*. In the handbook, it is suggested that the optimal values of parameters should be determined using and comparing the results of several validity indices since no validation index is reliable only by itself. Validity indices in the toolbox: partition coefficient, classification entropy, partition index, separation index, Xie Beni's index, Dunn's index and alternative Dunn's index, which are explained in the Section 2.4.3, are used for the determination of optimum parameter values. Optimum number of clusters is determined using the outputs of *optnumber* function of the toolbox. The *optnumber* function provides visual graphs for each validity index, which shows the change of the values of the validity index for different levels of cluster numbers. By using these graphs, the optimum number of clusters is determined by selecting the most appropriate value that optimizes the validity index values. For the determination of optimum degree of fuzziness, a new function is coded in MATLAB called *optfuzziness*, which is very similar to *optnumber*. Both functions *optnumber* and *optfuzziness* use the membership values obtained from the FCM algorithm, so the values of the parameters determined using these functions are optimum for the FCM clustering method. By using the NIFC algorithm instead of the FCM algorithm in the functions, optimum parameter values for the NIFC clustering method are obtained. By using *optnumber* and *optfuzziness* functions, the optimum pairs for number of clusters and degree of fuzziness values are determined by using training data sets for both methods, FCF and NIFCF. The optimal values of number of clusters and degree of fuzziness for each data set are given in Appendix B.

Other parameters, termination constant, ε, maximum number of iterations, n_{max} and alpha-cut level, α are determined as 10^{-12}, 100 and 0.1, respectively. For the testing algorithm of the NIFCF method 5 nearest neighbors are decided to be used to find the estimated membership values of testing data.

Finally, transformations of membership values to be used for the FCF and NIFCF methods are determined. Logit transformation, $\log((1-\boldsymbol{\gamma_i})/\,\boldsymbol{\gamma_i})$, which is one of the suggested transformations that can increase the performance of the model more than others since the distribution of the membership values are generally Gaussian (Çelikyılmaz, 2008), and power transformation, ${\gamma_i}^3$ to represent higher order terms in case of significance of higher order terms, are selected to be used for both of the methods. Thus, the input matrix used for constructing fuzzy functions for each cluster is composed of original input variables, membership values of the related cluster and their log-odds and power transformations, $\boldsymbol{\Phi}_\mathbf{i}(\mathbf{X},\boldsymbol{\gamma_i}) = [\mathbf{X}_\mathbf{i}^{\boldsymbol{\alpha}} \quad \boldsymbol{\gamma_i} \quad \boldsymbol{\gamma_i^3} \quad \log((\mathbf{1}-\boldsymbol{\gamma_i})/\,\boldsymbol{\gamma_i})]$.

In addition to the methods based on fuzzy functions, LR is also applied for three data sets in order to compare the performance of NIFCF with a conventional statistical classification method based on probability theory. The residual checks are performed before the application of LR.

Finally, using the posterior probabilities obtained, observations are classified according to 0.5 threshold level for all of the methods applied.

5.3.2. Performance Analysis

Performances of the methods, FCF, NIFCF and LR applied on three data sets are tested using several classification performance measures. The average values of these classification measures calculated for each set of each replication can be seen in Table 5.1.

Table 5.1: Average Application Results for LR, FCF and NIFCF

	Customer Satisfaction			**Casting**			**Ionosphere**		
Measure (Avg.)	LR	FCF	NIFCF	LR	FCF	NIFCF	LR	FCF	NIFCF
MCR	0.222	0.205	0.197	0.123	0.087	0.069	0.116	0.099	0.097
PCC	0.778	0.795	0.803	0.877	0.913	0.931	0.884	0.901	0.903
KAPPA	0.768	0.786	0.794	0.874	0.911	0.930	0.878	0.896	0.898
Precision	0.663	0.690	0.729	0.963	0.969	0.967	0.882	0.904	0.907
Recall	0.611	0.681	0.736	0.888	0.926	0.952	0.947	0.947	0.947
Specificity	0.852	0.846	0.833	0.822	0.844	0.822	0.772	0.820	0.825
F05	0.639	0.670	0.694	0.947	0.959	0.963	0.894	0.912	0.914
F1	0.618	0.658	0.685	0.923	0.946	0.958	0.913	0.925	0.926
F2	0.610	0.664	0.707	0.901	0.934	0.954	0.933	0.938	0.938
Log-Odds Ratio	2.158	2.572	2.684	3.367	3.574	3.775	4.254	4.458	4.504
Stability	0.059	0.077	0.023	0.066	0.046	0.013	0.023	0.035	0.014
AUC	0.721	0.747	0.772	0.786	0.815	0.819	0.851	0.873	0.876

Table 5.1 indicates that the NIFCF method gives more satisfactory results than others for almost all classification measures for each data set. The only measure that the NIFCF does not outperform other methods is specificity for two data sets, which measures the proportion of the customers somewhat satisfied with the driver seat, which are correctly identified. In addition, FCF gives more satisfactory results compared with LR for all performance measures except stability. It seems that LR gives more stable results.

These 3 methods are statistically compared by using two-way ANOVA in order to see whether there is a statistically significant difference between these methods according to classification measures mentioned above. Two-way ANOVA test is performed using *Minitab 15*. The classification measures are entered as response variable and the methods and the data sets that represent blocking variable are entered as factors. Firstly, the assumptions of ANOVA for each measure are checked with residual plots. For the performance measures that the assumptions of ANOVA are not satisfied, logarithmic transformations of them are used. The two-way ANOVA results for each classification measure can be seen in Table 5.2.

Table 5.2 shows that there are significant differences among the methods according to classification measures MCR, PCC, KAPPA, $F_{0.5}$ and AUC at the significance level of α=0.05.

Table 5.2: Two-Way ANOVA Results of LR, FCF and NIFCF methods

Measure (Avg.)	Source	DF	SS	MS	F	P
MCR, PCC	Method	2	0.0155	0.0077	9.7300	0.0290
	Dataset	2	0.2177	0.1089	137.0500	0.0000
	Interaction	4	0.0032	0.0008	0.3500	0.8400
	Error	72	0.1611	0.0022		
	Total	80	0.3975			
KAPPA	Method	2	0.0165	0.0082	9.8300	0.0290
	Dataset	2	0.2410	0.1205	143.8500	0.0000
	Interaction	4	0.0034	0.0008	0.3400	0.8480
	Error	72	0.1760	0.0024		
	Total	80	0.4368			
Precision	Method	2	3.3810	1.6900	3.6700	0.1240
	Dataset	2	90.8460	45.4230	98.7300	0.0000
	Interaction	4	1.8400	0.4600	0.5300	0.7120
	Error	72	62.1110	0.8630		
	Total	80	158.1780			
Recall	Method	2	7.1590	3.5790	2.7800	0.1750
	Dataset	2	46.1970	23.0980	17.9600	0.0010
	Interaction	4	5.1440	1.2860	0.7100	0.5870
	Error	72	130.2720	1.8090		
	Total	80	188.7720			
Specificity	Method	2	0.3906	0.1953	1.3900	0.3480
	Dataset	2	2.4402	1.2201	8.7000	0.0350
	Interaction	4	0.5610	0.1402	0.2400	0.9170
	Error	72	42.6514	0.5924		
	Total	80	46.0431			
$F_{0.5}$	Method	2	1.6440	0.8220	11.7500	0.0210
	Dataset	2	88.9620	44.4810	636.0300	0.0000
	Interaction	4	0.2800	0.0700	0.2600	0.9020
	Error	72	19.2620	0.2680		
	Total	80	110.1470			
F_1	Method	2	2.3710	1.1850	5.0600	0.0800
	Dataset	2	78.7690	39.3840	168.1000	0.0000
	Interaction	4	0.9370	0.2340	0.7300	0.5750
	Error	72	23.1180	0.3210		
	Total	80	105.1940			

Table 5.2 (cont'd): Two-Way ANOVA Results of LR, FCF and NIFCF methods

F_2	Method	2	4.0847	2.0424	3.0100	0.1590
	Dataset	2	71.9752	35.9876	53.0900	0.0010
	Interaction	4	2.7113	0.6778	0.9700	0.4300
	Error	72	50.4193	0.7003		
	Total	80	129.1906			
Log-Odds Ratio	Method	2	23.2280	11.6140	5.5500	0.0700
	Dataset	2	33.4510	16.7260	7.9900	0.04
	Interaction	4	8.3740	2.0930	0.8400	0.503
	Error	72	178.8280	2.4840		
	Total	80	243.8820			
Stability	Method	2	0.0212	0.0106	5.9700	0.0630
	Dataset	2	0.0115	0.0058	3.2400	0.1450
	Interaction	4	0.0071	0.0018	1.1800	0.3270
	Error	72	0.1082	0.0015		
	Total	80	0.1480			
AUC	Method	2	0.8691	0.4346	33.8800	0.0030
	Dataset	2	7.3853	3.6926	287.9300	0.0000
	Interaction	4	0.0513	0.0128	0.0700	0.9910
	Error	72	12.9762	0.1802		
	Total	80	21.2819			

The multiple comparison analysis is applied in order to see which method performs significantly better than others for the classification measures in which methods significantly differ. The methods are compared in pairs using Tukey's test. The test is conducted using "ANOVA with General Linear Model (GLM)" tool of *Minitab 15* with the family error rate 0.05. The p-values of Tukey's multiple comparison test for the classification measures MCR, PCC, KAPPA and AUC, in which methods significantly differ, are given in Table 5.3. Notations, ">" and "<" indicate that the method in the column list shows higher and lower performance than the method in the row list, respectively.

Table 5.3: p-values of Tukey's Multiple Comparison Test for MCR, PCC, KAPPA, $F_{0.5}$ and AUC

Measures	Methods	FCF	NIFCF
	LR	0.0797	0.0276 (>)
MCR, PCC	FCF		0.4956
	LR	0.0782	0.0272 (>)
KAPPA	FCF		0.4954
$F_{0.5}$	LR	0.0715	0.0193 (>)
	FCF		0.3461
	LR	0.0117(>)	0.0029 (>)
AUC	FCF		0.128

Table 5.3 shows that the NIFCF method has significantly better performance than LR according to MCR, PCC, KAPPA and AUC classification measures and the FCF gives significantly better results for AUC measure than LR. However, there is not any significance difference detected between fuzzy classification methods, FCF and NIFCF.

5.4. Discussion

Performance analysis study on LR, FCF and NIFCF indicates that fuzzy classification methods give better performance results than the conventional statistical classification method, LR, for data sets including fuzzy type of uncertainty. In addition, NIFCF gives more satisfactory results than FCF when average values of classification measures are considered even if there is not any significant difference detected.

Moreover, the use of a nonparametric method, MARS in the clustering phase of the algorithm, enables clustering of the data for each level of parameters. So, the selection of optimum model parameters by examining validity measures depending on the clustering results at different levels of the parameters is achieved and proper application of the method is guaranteed. In addition, the selection of variable transformations is automated.

Apart from the advantages mentioned above, there is also a disadvantage of using NIFCF method according to the other methods, FCF and LR. Native communication between two softwares, *MATLAB 7.3.0* and *R 2.8.1*, which is necessary for getting MARS results automatically from *R 2.8.1*, causes program to slow down. So, the run of the *MATLAB 7.3.0* program for the NIFCF method takes considerably longer time than the other methods. For example, while the run time of FCF method is approximately 15 seconds for one training data set of Ionosphere data, it takes almost 6 minutes to run NIFCF method for the same data set.

CHAPTER 6

CONCLUSIONS AND FURTHER STUDIES

Classification techniques have a wide range of application area in industrial engineering such as determination of customer satisfaction, dividing customers into groups according to their specific features, assignment of personnel into appropriate occupation groups (Zopounidis and Doumpos, 2002). Statistical classification methods are mainly used for these classification problems. However, statistical classification methods, which handle crisp data, may not answer the needs of this application area since in many applications in industrial engineering, we should deal with non crisp data resulted from the use of human judgments or linguistic terms. For example, qualitative information exists as well as quantitative information in regarding relationship between customer attributes and engineering characteristics in quality function deployment (Kim, Moskowitz and Köksalan, 1996) or expert judgments have an important role in the decision making processes in industrial engineering. Such data brings fuzzy type of uncertainty, which should be handled by the use of an appropriate fuzzy classification method. These fuzzy classification methods are very limited in number in the literature but needed for many applications from many fields especially for industrial engineering area, where qualitative information has an important role as well as quantitative information. This study brings two contributions to the field of fuzzy classification by developing Tanaka based fuzzy classification models and NIFCF.

Firstly, different classification models based on Tanaka's FLR approach are developed and compared using the customer satisfaction survey data. We have shown how Tanaka's FLR approach can be used for classification problems after converting the discrete dependent variable to a continuous variable by considering different types of uncertainties in data and interpreting and reflecting these uncertainties in different ways. Three alternative approaches are developed and tested using several classification measures. These alternative approaches differ only in the way the discrete dependent variable is converted to the equivalent continuous variable. After the conversion takes place, Tanaka's FLR approach is applied in all alternative approaches without any modification. As a result of the comparison of these approaches, it is observed that the alternative approaches, which consider both randomness and fuzziness included in the data inherently, outperform the other alternative approach, which considers only fuzzy type of uncertainty.

As a second contribution, the IFCF method is further improved by using a nonparametric method, MARS, in the clustering phase of it, instead of multiple linear regression. By the use of MARS, better fitting models can be developed in the clustering phase and the selection of appropriate variable transformations is automated. The proposed NIFCF method is compared with the FCF and LR methods using three data sets: customer satisfaction, casting and ionosphere. The NIFCF method gives more satisfactory results compared with the other methods. The only disadvantage of the NIFCF method is identified as slower running of the program as a result of native communication between two programs, which are *MATLAB 7.3.0* and *R 2.8.1.*

As a future study, different classification models can be developed based on different FLR methods such as Peters' (1994) and Ozelkan and Duckstein's (2000) using similar conversions of discrete dependent variables into continuous fuzzy variables or different fuzzy linear regression methods can be developed and used for classification purposes. As an example, the objective function of the LP model of

Tanaka's FLR approach can be changed to minimize the maximum fuzziness of the predicted fuzzy intervals instead of total fuzziness.

In addition, performance of the NIFCF method can be tested using more data sets from several fields. Moreover, the methods can be compared according to their sensitivity to the parameters such as the number of clusters or the degree of fuzziness. Furthermore, the computational time of the NIFCF algorithm can be reduced by increasing the efficiency of the code and/or use of different nonparametric/robust methods instead of MARS.

REFERENCES

Abe, S. and Thawonmas, R. (1997). A fuzzy classifier with ellipsoidal regions. *IEEE Trans. Fuzzy Systems, 5(*3), 358-368.

Ayhan, D. (2009). "Multi-Class Classification Methods Utilizing Mahalanobis Taguchi System and A Re-Sampling Approach for Imbalanced Data Sets" M.S Thesis, Middle East Technical University, Ankara, Turkey.

Bakır, B. (2007). "Defect Cause Modeling with Decision Tree And Regression Analysis: A Case Study in Casting Industry" M.S Thesis, Middle East Technical University, Ankara, Turkey.

Balasko, B., Abonyi, J. and Feil, B. Fuzzy Clustering and Data Analysis Toolbox For Use with MATLAB. http://www.fmt.vein.hu/softcomp/fclusttoolbox.

Bolotin, A. (2005). Fuzzification of linear regression models with indicator variables in medical decision making. *Proceeding of International Conference on Computational Intelligence for Modeling, Control and Automation*, 572-576.

Celmins, A. (1987). Least squares model fitting to fuzzy vector data. *Fuzzy Sets and Systems, 22*(3), 245-269.

Chang, Y.H. and Ayyub, B. M. (2001). Fuzzy regression methods - a comparative assessment. *Fuzzy Sets and Systems, 119* (2), 187-203.

Çelikyılmaz, A. (2008). "Modeling Uncertainty with Evolutionary Improved Fuzzy Functions" Ph. D. Thesis, University of Toronto, Canada.

Çelikyılmaz, A. and Türkşen, I.B. (2008). Enhanced fuzzy system models with improved fuzzy clustering algorithm. *IEEE Transactions on Fuzzy Systems, 16*(3), 779-794

Çelikyılmaz, A., Türkşen, I.B., Aktaş, R., Doğanay, M.M. and Ceylan, N.B. (2007). A new classifier design with fuzzy functions. *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, 4482, 136-143.

Diamond, P. (1988). Fuzzy least squares. *Information Sciences 46*(3), 141-157.

D'Urso, P. (2003). Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data. *Computational Statistics and Data Analysis, 42*(1-2), 47-72.

D'Urso, P. and Gastaldi, T. (2000). A least-squares approach to fuzzy linear regression analysis. *Computational Statistics and Data Analysis, 34*(4), 427-440.

D'Urso, P. and Santoro, A. (2006). Goodness of fit and variable selection in the fuzzy multiple linear regression. *Fuzzy Sets and Systems, 157*(19), 2627-2647.

Friedman J. (1991). Multivariate adaptive regression splines. *Annals of Statistics*, 19, 1-141.

Hathaway, R.J. and Bezdek, J.C. (1993). Switching regression models and fuzzy clustering. *IEEE Transactions on Fuzzy Systems*, *1*(3), 195-203.

Hojati, M., Bector, C.R. and Smimou, K. (2005). A simple method for computation of fuzzy linear regression. *European Journal of Operational Research, 166* (1), 172-184.

Hosmer, D.W. and Lemeshow, S. (2000). *Applied Logistic Regression.* John Willey & Sons, Inc. New York, NY.

Huang, M.L., Chen, H.Y. and Huang, J.J. (2007). Glaucoma detection using adaptive neuro- fuzzy inference system. *Expert Systems with Applications 32*(2), 458–468.

Hung, W. and Yang, M. (2006). An omission approach for detecting outliers in fuzzy regression models. *Fuzzy Sets and Systems, 157*(23), 3109-3122.

Ishibuchi, H. (1992). Fuzzy regression analysis. *Fuzzy Theory and Systems, 4*, 137-148.

Kao, C. and Chyu, C.-L. (2002). A fuzzy linear regression model with better explanatory power. *Fuzzy Sets and Systems, 126*(3), 401-409.

Kim, B. and Bishu, R.R. (1998). Evaluation of fuzzy linear regression models by comparing membership functions. *Fuzzy Sets and Systems, 100* (1-3), 343-352.

Kim, K.J., Moskowitz, H and Köksalan, M. (1996). Fuzzy versus statistical linear regression. *European Journal of Operational Research, 92,* 417-434.

Klir, G.J. and Folger, T.A. (1988). *Fuzzy Sets, Uncertainty and Information*. Printice Hall, Inc. New Jersey.

Köksal, G., Batmaz, İ. And Testik, M.C. (2008). Data mining processes and a review of their applications for product and process quality improvement in manufacturing industry. Tech. Report No: 08-03, Middle East Technical University, Ankara, Turkey.

Nasrabadi, M.M. and Nasrabadi, E. (2004). A mathematical-programming approach to fuzzy linear regression analysis. *Applied Mathematics and Computation*, 155, 873-881.

Özelkan, E.C. and Duckstein, L. (2000). Multi-objective fuzzy regression: A general framework. *Computers and Operations Research*, 27(7-8), 635–652.

Özer, G., Kılıç, T., Kartal, E., Batmaz, İ., Bayrak, Ö.T., Köksal, G. and Türkşen, İ.B. (2009). A nonparametric improved fuzzy classifier function approach for classification based on customer satisfaction survey data. *Book of Abstracts of The 23rd European Conference on Operational Research*. Bonn, Germany.

Özer, G., Köksal, G., Batmaz, İ., Bayrak, Ö.T., Kartal, E. and Kılıç, K. (2009). Classification models based on Tanaka's fuzzy linear regression approach: The case of customer satisfaction modeling. *1st International Fuzzy Systems Symposium* (Accepted to be presented).

Peters, G. (1994). Fuzzy linear regression with fuzzy intervals. *Fuzzy Sets and Systems*, *63*(1), 45-55.

Ross, T.J. (2004). *Fuzzy Logic with Engineering Applications.* John Willey & Sons, Inc. New York, NY.

Sakawa, M. and Yano, H. (1992). Multiobjective fuzzy linear regression analysis for fuzzy input output data. *Fuzzy Sets and Systems, 47*(2), 173 -181.

Setnes, M. and Babushka, R. (1999). Fuzzy relational classifier trained by fuzzy clustering. *IEEE Transactions on Systems, Man and Cybernetics, 29* (5) 619-625.

Sugeno, M. and Yasukawa, T. (1993). A Fuzzy Logic Based Approach to Qualitative Modeling. *IEEE Transactions on Fuzzy Systems, 1*(1), 7-31.

Takagi, T. and Sugeno, M. (1985). Fuzzy Identification of Systems and Its Applications to Modeling and Control. *IEEE Transactions on Systems, Man and Cybernetics*, *15*(1), 116-132.

Tanaka, H., Hayashi, I. and Watada, J. (1989). Possibilistic linear regression analysis for fuzzy data. *European Journal of Operational Research, 40*(3), 389-396.

Tanaka, H., Uejima, S. and Asai, K. (1982). Linear regression analysis with fuzzy model. *IEEE Transactions on Systems Man Cybernetics, 12*(6), 903-907.

Taylan, P., Weber, G.-W., and Yerlikaya, F. (2008). Continuous optimization applied in MARS for modern applications in finance, science and technology. *Proceeding of*

20th Mini-EURO Conference Continuous Optimization and Knowledge-Based Technologies.

Türkşen, I.B. (2008). Fuzzy functions with LSE. *Applied Soft Computing 8*(3), 1178-1188.

Türkşen, I.B. and Çelikyılmaz, A. (2006). Comparison of fuzzy functions with fuzzy rule base approaches. *International Journal of Fuzzy Systems, 8*(3), 137-149.

UCI Machine Learning Data Repository, http://archive.ics.uci.edu/ml/datasets.html

Wang, H.F. and Tsaur, R.C. (2000). Insight of a fuzzy regression model. *Fuzzy Sets and Systems, 112*(3), 355-369.

Wang, H.F. and Tsaur, R.C. (2000). Resolution of fuzzy regression model. *European Journal of Operational Research, 126*(3), 637-650.

Weiss, S.M. and Zhang T. (2003). Performance Analysis and Evaluation. *The Handbook of Data Mining*. Edited By Nong Ye.

Yerlikaya, F. (2008). "A New Contribution to Nonlinear Robust Regression and Classification with MARS and Its Applications to Data Mining for Quality Control in Manufacturing" M.S Thesis, Middle East Technical University, Ankara, Turkey.

Zadeh, L.A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems, 1*(1), 3-28.

Zadeh, L.A. (2000). Outline of a Computational Theory of Perceptions Based on Computing with Words. *Soft Computing and Intelligent Systems*, Academic Press, London, 3–22.

Zopounidis, C. (2002). MCDA methodologies for classification and sorting. *European Journal of Operational Research, 138*(2), 227-228.

Zopounidis, C. (2002). Multicriteria classification and sorting methods: A literature review. *European Journal of Operational Research, 138*(2), 229-246.

APPENDIX A

VARIABLE SELECTION

A.1. Variable Selection Output for Casting Data Set

Logistic Regression

```
      Total number of cases:      92 (Unweighted)
      Number of selected cases:   92
      Number of unselected cases: 0

      Number of selected cases:                 92
      Number rejected because of missing data:  0
      Number of cases included in the analysis: 92



Dependent Variable Encoding:

Original       Internal
Value          Value
     ,00       0
    1,00       1




Dependent Variable..   Y

Beginning Block Number  0.  Initial Log Likelihood Function

-2 Log Likelihood   81,821557

* Constant is included in the model.


Beginning Block Number  1.  Method: Forward Stepwise (COND)


       Improv.              Model              Correct
Step   Chi-Sq.  df   sig    Chi-Sq.  df   sig Class %      Variable
   1    15,631   1  ,000     15,631   1  ,000   80,43    IN: X21
   2     6,477   1  ,011     22,108   2  ,000   82,61    IN: X8
```

```
   3     8,621   1  ,003    30,729   3  ,000   84,78   IN: X7
   4     7,203   1  ,007    37,932   4  ,000   88,04   IN: X14
   5     6,980   1  ,008    44,912   5  ,000   90,22   IN: X29
   6     6,623   1  ,010    51,535   6  ,000   93,48   IN: X35
   7     6,305   1  ,012    57,840   7  ,000   93,48   IN: X11
   8    23,982   1  ,000    81,822   8  ,000  100,00   IN: X27

No more variables can be deleted or added.


End Block Number 1   PIN =     ,1500  Limits reached.

Final Equation for Block 1

Estimation terminated at iteration number 26 because
a perfect fit is detected.  This solution is not unique.

 -2 Log Likelihood         ,000
 Goodness of Fit           ,000
 Cox & Snell - R^2         ,589
 Nagelkerke - R^2         1,000

                    Chi-Square    df Significance

 Model                  81,822     8        ,0000
 Block                  81,822     8        ,0000
 Step                   23,982     1        ,0000

>Warning # 18582
>Covariance matrix cannot be computed.  Remaining statistics will
be
>omitted.
```

A.2. Variable Selection Output for Ionosphere Data Set

Logistic Regression

```
      Total number of cases:      351 (Unweighted)
      Number of selected cases:   351
      Number of unselected cases: 0

      Number of selected cases:                 351
      Number rejected because of missing data:  0
      Number of cases included in the analysis: 351

The variable X2 is constant for all selected cases.
Since a constant was requested in the model,
it will be removed from the analysis.



Dependent Variable Encoding:

Original       Internal
Value          Value
     ,00       0
    1,00       1

Dependent Variable..   Y

Beginning Block Number  0.  Initial Log Likelihood Function

-2 Log Likelihood   458,28371

* Constant is included in the model.


Beginning Block Number  1.  Method: Forward Stepwise (COND)


       Improv.              Model              Correct
Step   Chi-Sq.  df   sig    Chi-Sq.  df   sig Class %     Variable
   1   105,715   1  ,000    105,715   1  ,000   79,49   IN: X3
   2    74,512   1  ,000    180,227   2  ,000   84,05   IN: X1
   3    34,880   1  ,000    215,107   3  ,000   86,61   IN: X5
   4    26,718   1  ,000    241,824   4  ,000   88,03   IN: X8
   5    12,569   1  ,000    254,393   5  ,000   88,89   IN: X34
   6     7,241   1  ,007    261,634   6  ,000   89,74   IN: X7
   7     7,248   1  ,007    268,882   7  ,000   89,17   IN: X6
   8     6,434   1  ,011    275,316   8  ,000   88,32   IN: X22
   9     7,694   1  ,006    283,010   9  ,000   90,03   IN: X27
  10    13,312   1  ,000    296,322  10  ,000   90,31   IN: X30
```

```
  11    11,871   1  ,001   308,193  11  ,000   90,88   IN: X25
  12     7,643   1  ,006   315,836  12  ,000   90,60   IN: X31
  13     2,877   1  ,090   318,714  13  ,000   91,17   IN: X16
  14     3,364   1  ,067   322,078  14  ,000   92,31   IN: X18

No more variables can be deleted or added.


End Block Number 1   PIN =     ,1500  Limits reached.

Final Equation for Block 1

Estimation terminated at iteration number 10 because
Log Likelihood decreased by less than ,01 percent.

 -2 Log Likelihood      136,206
 Goodness of Fit        240,675
 Cox & Snell - R^2         ,601
 Nagelkerke - R^2          ,824

                     Chi-Square    df Significance

 Model                  322,078    14        ,0000
 Block                  322,078    14        ,0000
 Step                     3,364     1        ,0666

Classification Table for Y
The Cut Value is ,50
                    Predicted
                  ,00    1,00     Percent Correct
                    0  ⇔    1
Observed         ⇕⇓⇓⇓⇓⇓⇓⇓⇕⇓⇓⇓⇓⇓⇓⇓⇕
   ,00      0    ⇔  217  ⇔    8  ⇔   96,44%
                 ⇕⇓⇓⇓⇓⇓⇓⇓⇕⇓⇓⇓⇓⇓⇓⇓⇕
   1,00     1    ⇔   19  ⇔  107  ⇔   84,92%
                 ⇕⇓⇓⇓⇓⇓⇓⇓⇕⇓⇓⇓⇓⇓⇓⇓⇕
                           Overall  92,31%

---------------------- Variables in the Equation ------------------
-----

Variable            B      S.E.     Wald    df      Sig       R
Exp(B)

X1          -19,8136   24,5570    ,6510     1    ,4198   ,0000
,0000
X3           -2,7225     ,7517  13,1172     1    ,0003  -,1558
,0657
X5           -2,3033     ,8941   6,6367     1    ,0100  -,1006
,0999
X6           -3,2288     ,8540  14,2956     1    ,0002  -,1638
,0396
X7           -2,3778     ,8170   8,4694     1    ,0036  -,1188
,0928
X8           -2,7251     ,7105  14,7122     1    ,0001  -,1665
,0655
```

X16	1,7593	,9591	3,3645	1	,0666	,0546	5,8085
X18	-1,1656	,6647	3,0749	1	,0795	-,0484	,3117
X22	2,4357	,7382	10,8879	1	,0010	,1393	11,4237
X25	-1,8347	,6374	8,2846	1	,0040	-,1171	,1597
X27	3,9619	,9008	19,3435	1	,0000	,1945	52,5577
X30	-3,9982	1,0390	14,8076	1	,0001	-,1672	,0183
X31	-2,1467	,8735	6,0401	1	,0140	-,0939	,1169
X34	3,4899	,9033	14,9274	1	,0001	,1680	32,7811
Constant	23,0048	24,5930	,8750	1	,3496		

APPENDIX B

OPTIMUM PARAMETER VALUES FOR FCM AND NIFC ALGORITHMS

Table B.1: Optimum parameter values for three data sets

	Customer Satisfaction				Casting				Ionosphere			
	FCM		NIFC		FCM		NIFC		FCM		NIFC	
	m	n	m	n	m	n	m	n	m	n	m	n
R1Tr1	1.65	3	1.65	3	2	3	1.85	3	1.75	6	1.8	6
R1Tr2	1.65	2	1.6	3	1.95	2	1.9	2	1.75	4	1.75	5
R1Tr3	1.6	3	1.65	3	1.95	2	1.9	2	1.7	6	1.65	5
R2Tr1	1.65	2	1.65	3	2.15	3	1.75	3	1.8	6	1.8	6
R2Tr2	1.6	3	1.7	3	2.05	3	2.05	3	1.7	5	1.6	5
R2Tr3	1.65	3	1.7	3	1.8	3	1.8	3	1.7	6	1.7	5
R3Tr1	1.6	4	1.65	3	2	2	1.7	3	1.65	7	1.6	5
R3Tr2	1.65	3	1.65	3	2	2	1.95	3	1.65	5	1.7	6
R3Tr3	1.65	2	1.65	3	1.8	3	2.05	3	1.8	6	1.8	5

m: degree of fuzziness

n: number of clusters