Article

Credit Scoring by Fuzzy Support Vector Machines with a Novel Membership Function

Jian Shi and Benlian Xu *

School of Electrical & Automatic Engineering, Changshu Institute of Technology, Changshu 215500, China;
shijian@cslg.edu.cn
* Correspondence: xu_benlian@cslg.edu.cn

Abstract: Due to the recent financial crisis and European debt crisis, credit risk evaluation has become an increasingly important issue for financial institutions. Reliable credit scoring models are crucial for commercial banks to evaluate the financial performance of clients and have been widely studied in the fields of statistics and machine learning. In this paper a novel fuzzy support vector machine (SVM) credit scoring model is proposed for credit risk analysis, in which fuzzy membership is adopted to indicate different contribution of each input point to the learning of SVM classification hyperplane. Considering the methodological consistency, support vector data description (SVDD) is introduced to construct the fuzzy membership function and to reduce the effect of outliers and noises. The SVDD-based fuzzy SVM model is tested against the traditional fuzzy SVM on two real-world datasets and the research results confirm the effectiveness of the presented method.

Keywords: fuzzy support vector machine; support vector data description; credit scoring

1. Introduction

During the recent financial crisis, many financial institutions endured great losses from customers' defaults on loans such as the subprime mortgage crisis in the USA. However, credit-granting institutions cannot simply refuse all applicants to avoid credit risk as the competition in the growing credit market has become fierce. Effective credit scoring has become one of the primary techniques for gaining competitive advantages in credit market which can help financial institutions to increase credit volume without excessively increasing their exposure to default.

Credit scoring models are developed to discriminate applicants as either accepted or rejected with respect to the customers' application form and credit report, which is built on the basis of past applicants' characteristics [1–3]. Since even a fraction of improvement in accuracy of credit scoring might translate into noteworthy future savings, numerous data mining and statistical techniques have been proposed to derive a satisfied credit scoring model over the past decades. Generally, these methods can be classified to statistics approaches (e.g., discriminant analysis and logistic regression), and machine learning approaches (e.g., artificial neural network and support vector machine). Though traditional statistical methods are relatively simple and explainable, their discriminating ability is still an argumentative problem due to the nonlinear relationship between default probability and credit patterns. Additionally, statistical methods must assume posterior probability models while machine learning approaches commonly operate without this limitation.

Regarding recent credit scoring techniques, artificial neural network [4,5] has been criticized for its poor performance when incorporating irrelevant attributes or small data sets, while support vector machine, motivated by statistical learning theory [6,7], is particularly well suited for coping with a large number of explanatory attributes or sparse data sets [8–11]. Baesens et al. studied the performance of various state-of-the-art classification algorithms on eight real-life credit scoring

data sets [12]. Thomas et al. tested 17 consumer credit modeling methods and report that SVM ranks the highest in term of classification accuracy [13]. Huang et al. constructed hybrid SVM-based credit scoring models with three strategies for feature selection and model parameter setting [14]. Martens et al. extracted rules from the trained SVM to obtain both accurate and comprehensible credit scoring models [15]. Niklis et al. developed linear and nonlinear support vector machines risk assessment and classification models, as well as additive SVM model that well suited the requirements of credit rating systems, which provided very competitive results [16]. Harris introduced the use of the clustered support vector machine for credit scorecard development and addresses some of the limitations associated with traditional nonlinear support vector machine based methods [17]. Cheng and Li proposed a rating model based on a support vector machine with monotonicity constraints derived from the prior knowledge of financial experts which helps correct the loss of monotonicity in data occurring during the collecting process [18].

However, in many real word credit scoring problems, some input training points are corrupted by noise, and some applicants may be assigned to the wrong side by accident. These points are all outliers, which do not completely belong to one class, but with different memberships in the two classes. In this case, standard SVM training algorithm will make the classification boundary to severely deviate from the optimal hyperplane as SVM is very sensitive to outliers. Lin and Wang introduced the concept of fuzzy SVM, which applies a fuzzy membership value to each input point according to its relative importance in the classification and reformulates the SVM such that allows different input point to make different contributions to the learning of decision surface [19]. Fuzzy SVM can be used as a remedy for unwanted over fitting caused by treating every data sample equally [20–22], however, the effects of the membership values when training fuzzy SVM is an interesting issue in the context of credit scoring modeling [23]. In this study, considering the methodological consistency, a novel support vector date description based fuzzy membership function is proposed to reduce the effects of outliers and improves the classification accuracy and generalization, and the effects of the membership values in fuzzy SVM is investigated and compared to those of SVM.

This paper is organized as follows. Section 2 recalls the background on SVM and fuzzy SVM, and reports two typical linear and nonlinear fuzzy membership functions. Section 3 highlights the main novelty of this work. The main aspects of the SVDD-based membership function are detailed. Section 4 collects the experimental results on two real world credit datasets. Finally, Section 5 draws the conclusion remarks.

2. SVM and Fuzzy SVM

2.1. Standard Support Vector Machines

In this section, the basic concept of SVM for classification problems is presented. Consider a two-class problem with a set of *l* sample points: (x_1, y_1), ..., (x_l, y_l). Each x_i has a class label $y_i \in \{-1, 1\}$ which denotes two classes separately. When the sample points are rigidly linear separable, SVM classifier tries to search for a hyperplane with the largest margin to separate them by solving the following quadratic program:

$$min \ \frac{1}{2}||w||^2$$
$$subject \ to \ y_i(w^T x_i + b) \geq 1 \ \ i = 1, 2, ..., l \tag{1}$$

where w is the weight vector and b is the bias term. For linear non-separable separable case, non-negative slack variables ξ_i are introduced to measure the amount of violation of the constraints in Equation (1). The QP problem turns to:

$$min \quad \frac{1}{2}||w||^2 + C\sum_{i=1}^{l}\xi_i$$
$$subject\ to \quad y_i(w^T x_i + b) \geq 1 - \xi_i \quad i = 1,2,...,l$$
$$\xi_i \geq 0 \qquad\qquad\qquad i = 1,2,...,l \tag{2}$$

where C is a regularized constant determining the tradeoff between margin maximization and classification violation. In many cases, it's difficult to obtain a suitable hyperplane in the original input space by such a linear classifier, thus a nonlinear mapping $\varphi(x)$ that satisfies Mercer's condition can be introduced to map the input variable x_i into a higher dimensional feature space. To solve the QP problem, the kernel function $K(x_i, x_j) = \varphi(x_i)\cdot\varphi(x_j)$ is introduced to compute the dot product of the data points in feature space, and the shape of the mapping φ is unnecessary to know. Then the optimal problem turns to the following dual form by constructing a Lagrangian:

$$min\ \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{l}\alpha_i$$
$$subject\ to\ \sum_{i=1}^{l}y_i\alpha_i = 0\ 0 \leq \alpha_i \leq C\ i = 1,2,...,l \tag{3}$$

where α are the non-negative Lagrange multipliers related to the constraints in Equation (2).

Some commonly used kernel functions are polynomial, sigmoid and radial basis function (RBF). RBF kernel is employed in this study because of its superiority, which is experimentally demonstrated by Gestel et al. [24].

2.2. Fuzzy Support Vector Machines

In credit scoring modeling, the applicants may not be exactly assigned to creditworthy or default ones. In other words, there is a fuzzy membership associated with each applicant, which can be regarded as the attitude of the corresponding applicants toward one class in the classification. Lin and Wang proposed the theory of fuzzy support vector machine based on standard SVM [19]. Suppose a set of labeled sample points with associated fuzzy membership: $(x_1,y_1,s_1), ..., (x_l,y_l,s_l)$. Each x_i has a class label $y_i \in \{-1,1\}$ and a fuzzy membership which satisfies $0 < s_i \leq 1$.

Since the fuzzy membership s_i is the attitude of the corresponding point toward one class and the parameter ξ_i is a measure of the constraints violation, the term $s_i\xi_i$ can be treated as a measure of constraints violation with different weights. Then the quadratic problem can be described as:

$$min \quad \frac{1}{2}||w||^2 + C\sum_{i=1}^{l}s_i\xi_i$$
$$subject\ to \quad y_i(w^T\varphi(x_i) + b) \geq 1 - \xi_i \quad i = 1,2,...,l$$
$$\xi_i \geq 0 \qquad\qquad\qquad\qquad i = 1,2,...,l \tag{4}$$

By constructing a Lagrangian, the quadratic programs can be solved in their dual space just as that for standard SVM:

$$min\ \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{l}\alpha_i$$
$$subject\ to\ \sum_{i=1}^{l}y_i\alpha_i = 0\ 0 \leq \alpha_i \leq s_i C\ i = 1,2,...,l \tag{5}$$

With different value of s_i, the tradeoff between the maximization of margin and the amount of constraints violation can be controlled. It is noted that a smaller s_i makes the corresponding point x_i less important in the training, thus choosing appropriate fuzzy memberships in a given problem is

of pivotal importance for FSVM. Lin proposed a model by setting a linear fuzzy membership as a function of the distance between each data point and its corresponding class center [19]. For the above training sample sequence: $(x_1, y_1, s_1), ..., (x_l, y_l, s_l)$, denote x_+ the mean of class with label +1 and x_- the mean of class -1. The radius of class +1 is as follows:

$$r_+ = max|x_+ - x_i| \ \ where \ \ y_i = 1 \tag{6}$$

and the radius of class -1 is:

$$r_- = max|x_- - x_i| \ \ where \ \ y_i = -1 \tag{7}$$

The fuzzy membership of each sample is

$$s_i = \begin{cases} 1 - |x_+ - x_i|/(r_+ + \delta) \ \ where \ \ y_i = 1 \\ 1 - |x_- - x_i|/(r_- + \delta) \ \ where \ \ y_i = -1 \end{cases} \tag{8}$$

with $\delta > 0$ to avoid the case $s_i = 0$.

The method performs well as the fuzzy membership is a function of the mean and radius of each class, and the effect of outlier can be reduced as it contributes little to the final decision plane. However, the algorithm is carried out in the original input space rather than the feature space, Tang further proposed a nonlinear membership function defined in the feature space with mapping function $\varphi(x)$ on the basis of the above method [25]. Define φ_+ and φ_- as the centers of the two classes by taking average on the points mapped to the feature space:

$$\varphi_+ = \frac{1}{n_+} \sum_{y_i=1} \varphi(x_i), \ \ \varphi_- = \frac{1}{n_-} \sum_{y_i=-1} \varphi(x_i) \tag{9}$$

where n_+ and n_- are the numbers of samples in two classes.

The radius is defined similarly to those of Lin:

$$r_+ = max|\varphi_+ - \varphi(x_i)| \ \ where \ \ y_i = 1 \\ r_- = max|\varphi_- - \varphi(x_i)| \ \ where \ \ y_i = -1 \tag{10}$$

Then the square of distance is calculated in the feature space:

$$d_{i+}^2 = ||\varphi_- - \varphi(x_i)||^2 = K(x_i, x_i) - \frac{2}{n_+} \sum_{y_j=1} K(x_j, x_i) + \frac{1}{n_+^2} \sum_i \sum_{y_j=1} K(x_i, x_j) \\ d_{i-}^2 = ||\varphi_+ - \varphi(x_i)||^2 = K(x_i, x_i) - \frac{2}{n_+} \sum_{y_j=1} K(x_j, x_i) + \frac{1}{n_+^2} \sum_i \sum_{y_j=1} K(x_i, x_j) \tag{11}$$

Finally the fuzzy membership of each sample is calculated as:

$$s_i = \begin{cases} 1 - \sqrt{d_{i+}^2/(r_+^2 + \delta)} \ \ where \ \ y_i = 1 \\ 1 - \sqrt{d_{i-}^2/(r_-^2 + \delta)} \ \ where \ \ y_i = -1 \end{cases} \tag{12}$$

with $\delta > 0$ to avoid the case $s_i = 0$.

The nonlinear version of FSVM outperforms the linear one as it could more accurately represent the contribution of each sample to the decision surface in the feature space. Both algorithm define the center and radius by taking average and maxim on sample points, however, they could be defined in a more explicable way.

3. Fuzzy SVM with SVDD Membership Function

Methodological consistency has been a major design principle and is expected to improve the comprehensibility of the modeling paradigm, which, in turn, may avail dispersion in practical applications [26]. Support Vector Data Description (SVDD) is inspired by the support vector machine classifier, which searches for a spherically shaped boundary around a dataset to detect novel data or outliers [27,28]. In this section, a SVDD member function is proposed for FSVM, which is defined in the feature space as well.

Similar to the SVM hyperplane approach with largest margin between two classes, SVDD estimates a hypersphere with the minimum volume to find an enclosed boundary containing almost all target objects, instead of using a hyperplane in SVM. Assume a hypersphere with center a and radius R, the cost function is defined as below:

$$min\ R^2 + C\sum_{i=1}^{l} \xi_i$$

$$subject\ to\ ||(\varphi(x_i) - a)||^2 \leq R^2 + \xi_i,\ \ \xi_i \geq 0\ i = 1, 2, ..., l \tag{13}$$

where ξ_i are slack variables, and the parameter C controls the trade-off between the volume and the violation.

The above optimal hypersphere problem can be solved by constructing the Lagrangian, which results in:

$$max\ \sum_{i=1}^{l} \alpha_i K(x_i, x_i) - \sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j K(x_i, x_j)$$

$$subject\ to\ \sum_{i=1} \alpha_i = 1,\ 0 \leq \alpha_i \leq C\ \ i = 1, 2, ..., l \tag{14}$$

The training samples with nonzero α_i are support vectors and are used to describe the hypersphere boundary.

Denote a_+ and a_- as the centers of class +1 and class −1 in the feature space respectively.

According to K-T condition, the center of the each class can be calculated as:

$$a_+ = \sum_{y_i=1} \alpha_i x_i,\ \ a_- = \sum_{y_i=-1} \alpha_i x_i \tag{15}$$

The radius of class +1 is then defined by

$$r_+ = max|a_+ - \varphi(x_i)|\ \ where\ y_i = 1 \tag{16}$$

and the radius of class −1 by

$$r_- = max|a_- - \varphi(x_i)|\ \ where\ y_i = -1 \tag{17}$$

The square of the distance between any sample x_i with class label $y_i = 1$ and the class center a_+ in the feature space can be calculated as:

$$d_{i+}^2 = K(x_i, x_i) - 2\sum_{y_j=1} \alpha_j^+ K(x_j, x_i) + \sum_{i}\sum_{y_j=1} \alpha_i^+ \alpha_j^+ K(x_i, x_j) \tag{18}$$

The square of the distance between the sample of class −1 and its corresponding class center a_- is derived similarly:

$$d_{i-}^2 = K(x_i, x_i) - 2\sum_{y_j=-1} \alpha_j^- K(x_j, x_i) + \sum_{i}\sum_{y_j=-1} \alpha_i^- \alpha_j^- K(x_i, x_j) \tag{19}$$

By definition , the radius r_+ is the distance from the center a_+ of the hypersphere of class +1 to any of its support vectors on the boundary, and r_- is the corresponding radius of class -1.

$$r_+ = d_{sv+}, \; r_- = d_{sv-} \tag{20}$$

Then the fuzzy membership s_i of each input sample x_i can be defined as follows:

$$s_i = \begin{cases} 1 - \sqrt{d_{i+}^2/(r_+^2 + \delta)} & where \; y_i = 1 \\ 1 - \sqrt{d_{i-}^2/(r_-^2 + \delta)} & where \; y_i = -1 \end{cases} \tag{21}$$

where $\delta > 0$ is a small constant used to avoid the case $s_i = 0$. In this paper, δ is set to 1×10^{-3} for all three models. Once the coefficients s_i are derived, fuzzy SVM for classification can be modeled according to Equation (4) in Section 2.

The fuzzy memberships are calculated in the feature space with a more explainable center and radius. As the class center and radius are defined in the minimum SVDD hypersphere including all the class targets, the fuzzy membership are more explicit and interpretable in the feature space compared to those average and maxim algorithm. In addition, the above formulas are similar to that of Jiang's, and are expected to reduce the effect of outliers as well.

4. Experimental Results and Discussion

In this section, two real-world credit datasets, the Australian and German credit data sets, are adopted to evaluate the performance of the proposed SVDD-FSVM method. Experimental results are compared with those of linear and nonlinear fuzzy SVMs. Australian credit data set, with 307 instances of accepted customers and 383 instances of rejected customers, contains 14 attributes, where six are continuous attributes and eight are categorical attributes. For confidentiality reasons, all attribute names and values have been changed to meaningless symbols. The German credit dataset is more unbalanced, consisting of 700 examples of creditworthy applicants and 300 examples of default applicants, with 24 numeric features describing 19 attributes. The attributes of each applicant are depicted in Table 1, with 4 attributes changed to dummy variables. Both data sets are made public from the UCI Repository of Machine Learning Databases, and are generally adopted as benchmarks to compare the performance of various classification models.

Table 1. Input variables of German data set.

Original Attributes	Input Variables	Variable Type	Attribute Description
A1	V1	qualitative	Status of existing checking account
A2	V2	numerical	Duration in month
A3	V3	qualitative	Credit history
A4	V4,V5	dummy	Purpose (V4: new car, V5: used car)
A5	V6	numerical	Credit amount
A6	V7	qualitative	Savings account/bonds
A7	V8	qualitative	Present employment since
A8	V9	qualitative	Personal status and sex
A9	V10,V11	dummy	Other debtors/guarantors (V10: none, V11: co-applicant)
A10	V12	numerical	Present residence since
A11	V13	qualitative	Property
A12	V14	numerical	Age in years
A13	V15	qualitative	Other installment plans
A14	V16,V17	dummy	Housing (V16: rent, V17: own)
A15	V18	numerical	Number of existing credits at this bank
A16	V19,V20,V21	dummy	Job (V19: unemployed/unskilled (non-resident), V20: unskilled (resident), V21: skilled employee/official)
A17	V22	numerical	Number of people being liable to provide maintenance for
A18	V23	qualitative	Telephone
A19	V24	qualitative	foreign worker

The input variables are normalized with respect to their maximum values and the fuzzy membership of each input instance is then derived by the SVDD operation. To visualize the SVDD hypersphere, the variables are projected to the two dimensional plane by principal component analysis, which is a popular multivariate statistic tool aiming to handle high dimensional, noisy, and correlated data by defining a reduced set of latent variables (referred to as principal components) [29]. The boundaries of class +1 and class –1 of Australian dataset for the first fold partition are pictured in Figure 1 and those of German dataset are pictured in Figure 2. Both hyperspheres are calculated with a confidence limit of 98%. Though the two principal components contain over 90% of the information (measured by variance) of the original data in both sets, the figures show that the rejected applicants and accepted applicants can hardly be distinguished by a simple boundary in two dimensional plane.
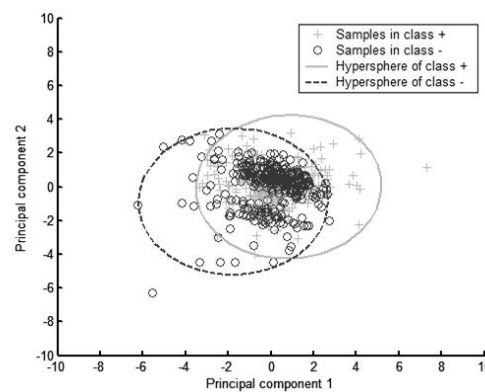

Figure 1. Principal components plot of SVDD for Australia data set.

For comparison purpose, the detailed performance of SVDD-FSVM is tested against linear FSVM and nonlinear FSVM. This study assessed the three credit scoring methods in terms of accuracy and other major assessment metrics such as sensitivity and specificity. Denote the number of default clients classified as accepted ones by DA and classified as rejected ones by DR. Let the number of creditworthy clients classified as accepted ones be CA and classified as rejected ones be CR. Then these evaluation criteria measuring the efficiency of the classification are defined as:

$$Sensitivity = \frac{CA}{CA+CR}$$

$$Specificity = \frac{DR}{DR+DA}$$

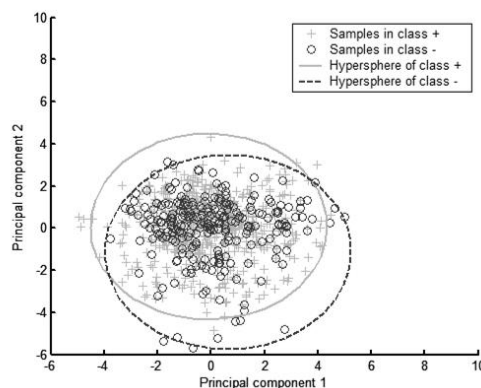$$Accuracy = \frac{CA+DR}{CA+CR+DA+DR}$$

(22)

Figure 2. Principal components plot of SVDD for German data set.

In this study, the credit dataset is randomly partitioned into identical training and test sets using 5-fold cross validation and a grid search is employed to find the optimal parameters [30]. The average comparison results on German and Australian dataset are shown in Tables 2 and 3 respectively.

Table 2. Performance on the Australian test data set.

Methods	Sensitivity (%)	Specificity (%)	Accuracy (%)
SVDD-FSVM	87.53	86.84	87.25
Nonlinear FSVM	89.87	85.13	87.10
Linear FSVM	86.95	86.48	86.67

Table 3. Performance on the German test data set.

Methods	Sensitivity (%)	Specificity (%)	Accuracy (%)
SVDD-FSVM	89.59	48.60	77.30
Nonlinear FSVM	92.15	41.75	77.00
Linear FSVM	95.18	23.42	73.60

The result listed in Table 2 indicates that the SVDD-FSVM outperformed the other two FSVMs on Australia data set, with overall accuracy of 87.25, comparing with those of 87.10 and 86.67 obtained from the corresponding nonlinear and linear FSVM models, respectively. Though the sensitivity, specificities and accuracy of German dataset in Table 3 are relatively lower than those of Australia data set as the German data structure is considered more unbalanced than the Australian data set, SVDD-FSVM still yields the best results among all the three approaches, with the specificity especially improved, which is considered more important than the sensitivity for the credit risk control of financial institutes.

5. Conclusions

This paper has presented methods for using fuzzy support vector machines to establish credit scoring models. Comparing with standard SVM, fuzzy SVM imposes a fuzzy membership to each input point such that different input points can make different contributions to the learning of decision surface, which enhances the SVM in reducing the effect of outliers and noises in data points with unmodeled characteristics. As choosing a proper fuzzy membership function is quite important to solving classification problem with FSVM, the SVDD version of fuzzy membership is proposed for methodological consistency consideration, which is a function of the distance between each input point and its corresponding SVDD hypersphere center. The SVDD-FSVM credit scoring model overall yields the best performance among all the three models when appropriately trained on two real-world credit data sets. The results indicate that the proposed method provides classification accuracy and reliability, and is supposed to have promising potential for practical use.

Author Contributions: Benlian Xu conceived and designed the experiments; Jian Shi performed the experiments, analyzed the data and wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Thomas, L.C.; Edelman, D.B.; Crook, J.N. *Credit Scoring and Its Applications*; Siam: Philadelphia, PA, USA, 2002.
2. Blöchlinger, A.; Leippold, M. Economic benefit of powerful credit scoring. *J. Bank. Financ.* **2006**, *30*, 851–873. [CrossRef]
3. Einav, L.; Jenkins, M.; Levin, J. The impact of credit scoring on consumer lending. *RAND J. Econ.* **2013**, *44*, 249–274. [CrossRef]
4. West, D. Neural network credit scoring models. *Comput. Oper. Res.* **2000**, *27*, 1131–1152. [CrossRef]
5. Chuang, C.L.; Huang, S.T. A hybrid neural network approach for credit scoring. *Expert Syst.* **2011**, *28*, 185–196. [CrossRef]
6. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995.
7. Vapnik, V.N. *Statistical Learning Theory*; John Wiley & Sons: New York, NY, USA, 1998.
8. Christmann, A.; Hable, R. Consistency of support vector machines using additive kernels for additive models. *Comput. Stat. Data Anal.* **2012**, *56*, 854–873. [CrossRef]
9. Jiang, H.; Yan, Z.; Liu, X. Melt index prediction using optimized least squares support vector machines based on hybrid particle swarm optimization algorithm. *Neurocomputing* **2013**, *119*, 469–477. [CrossRef]
10. Ruano, A.E.; Madureira, G.; Barros, O.; Khosravani, H.R.; Ruano, M.G.; Ferreira, P.M. Seismic detection using support vector machines. *Neurocomputing* **2014**, *135*, 273–283. [CrossRef]
11. Maldonado, S.; López, J. Imbalanced data classification using second-order cone programming support vector machines. *Pattern Recognit.* **2014**, *47*, 2070–2079. [CrossRef]
12. Baesens, B.; van Gestel, T.; Viaene, S.; Stepanova, M.; Suykens, J.; Vanthienen, J. Benchmarking state-of-the-art classification algorithms for credit scoring. *J. Oper. Res. Soc.* **2003**, *54*, 627–635. [CrossRef]
13. Thomas, L.C.; Oliver, R.W.; Hand, D.J. A survey of the issues in consumer credit modelling research. *J. Oper. Res. Soc.* **2005**, *56*, 1006–1015. [CrossRef]
14. Huang, C.-L.; Chen, M.-C.; Wang, C.-J. Credit scoring with a data mining approach based on support vector machines. *Expert Syst. Appl.* **2007**, *33*, 847–856. [CrossRef]
15. Martens, D.; Baesens, B.; van Gestel, T.; Vanthienen, J. Comprehensible credit scoring models using rule extraction from support vector machines. *Eur. J. Oper. Res.* **2007**, *183*, 1466–1476. [CrossRef]
16. Niklis, D.; Doumpos, M.; Zopounidis, C. Combining market and accounting-based models for credit scoring using a classification scheme based on support vector machines. *Appl. Math. Comput.* **2014**, *234*, 69–81. [CrossRef]
17. Harris, T. Credit scoring using the clustered support vector machine. *Expert Syst. Appl.* **2015**, *42*, 741–750. [CrossRef]
18. Chen, C.-C.; Li, S.-T. Credit rating with a monotonicity-constrained support vector machine model. *Expert Syst. Appl.* **2014**, *41*, 7235–7247. [CrossRef]
19. Lin, C.-F.; Wang, S.-D. Fuzzy support vector machines. *IEEE. Trans. Neural Netw.* **2002**, *13*, 464–471.
20. An, W.; Liang, M. Fuzzy support vector machine based on within-class scatter for classification problems with outliers or noises. *Neurocomputing* **2013**, *110*, 101–110. [CrossRef]
21. Chaudhuri, A. Modified fuzzy support vector machine for credit approval classification. *AI Commun.* **2014**, *27*, 189–211.
22. Wu, Z.; Zhang, H.; Liu, J. A fuzzy support vector machine algorithm for classification based on a novel PIM fuzzy clustering method. *Neurocomputing* **2014**, *125*, 119–124. [CrossRef]
23. Shieh, M.-D.; Yang, C.-C. Classification model for product form design using fuzzy support vector machines. *Comput. Ind. Eng.* **2008**, *55*, 150–164. [CrossRef]
24. Van Gestel, T.; Suykens, J.A.; Baesens, B.; Viaene, S.; Vanthienen, J.; Dedene, G.; De Moor, B.; Vandewalle, J. Benchmarking least squares support vector machine classifiers. *Mach. Learn.* **2004**, *54*, 5–32. [CrossRef]
25. Tang, W.M. Fuzzy SVM with a new fuzzy membership function to solve the two-class problems. *Neural Process. Lett.* **2011**, *34*, 209–219. [CrossRef]
26. Lessmann, S.; Voß, S. A reference model for customer-centric data mining with support vector machines. *Eur. J. Oper. Res.* **2009**, *199*, 520–530. [CrossRef]

27. Tax, D.M.; Duin, R.P. Support vector data description. *Mach. Learn.* **2004**, *54*, 45–66. [CrossRef]
28. Strack, R.; Kecman, V.; Strack, B.; Li, Q. Sphere Support Vector Machines for large classification tasks. *Neurocomputing* **2013**, *101*, 59–67. [CrossRef]
29. Jolliffe, I. *Principal Component Analysis*; John Wiley & Sons: New York, NY, USA, 2005.
30. Zhou, L.; Lai, K.-K.; Yu, L. Least squares support vector machines ensemble models for credit scoring. *Expert Syst. Appl.* **2010**, *37*, 127–133. [CrossRef]