ELSEVIER

Expert Systems with Applications 28 (2005) 603–614

Expert Systems with Applications

www.elsevier.com/locate/eswa

Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters

Jae H. Min[a], Young-Chan Lee[b,*]

[a]College of Business Administration, Sogang University, CPO Box 1142, Seoul 121-742, South Korea
[b]College of Commerce and Economics, Dongguk University, Gyeongju, Gyeongbuk, 780-714, South Korea

Abstract

Bankruptcy prediction has drawn a lot of research interests in previous literature, and recent studies have shown that machine learning techniques achieved better performance than traditional statistical ones. This paper applies support vector machines (SVMs) to the bankruptcy prediction problem in an attempt to suggest a new model with better explanatory power and stability. To serve this purpose, we use a grid-search technique using 5-fold cross-validation to find out the optimal parameter values of kernel function of SVM. In addition, to evaluate the prediction accuracy of SVM, we compare its performance with those of multiple discriminant analysis (MDA), logistic regression analysis (Logit), and three-layer fully connected back-propagation neural networks (BPNs). The experiment results show that SVM outperforms the other methods.

Keywords: Bankruptcy prediction; Support vector machine; Grid-search; Kernel function; Back-propagation neural networks

1. Introduction

Over the past decade, several financial crises observed in some emerging markets enjoying financial liberalization showed that debt financing could result in large and sudden capital outflows, thereby causing a domestic ‘credit crunch.’ This experience made banking authorities such as Bank for International Settlements (BIS) learn a number of lessons, among which they all encourage commercial banks to develop internal models to better quantify financial risks (Basel Committee on Banking Supervision, 1999). Decision making problems in the area of credit evaluation have been considered very important but difficult ones for financial institutions due to a high level of risk from wrong decisions. In order to effectively manage the credit risk exposures of financial institutions, there is a strong need for sophisticated decision support systems backed by analytical tools to measure, monitor, manage, and control financial and operational risks as well as inefficiencies (Emel, Oral, Reisman, & Yolalan, 2003; Park & Han, 2002).

A potential client’s credit risk is often evaluated by financial institutions’ internal credit scoring models, which aim to determine whether an applicant has the capability to repay by evaluating the risk of his loan application. Such models offer them a means for evaluating the risk of their credit portfolio, in a timely manner, by centralizing global-exposures data and by analyzing marginal as well as absolute contributions to risk components. Armed with these models, quantitative risk management systems can thus provide the financial institutions early warning signals for potential business failures of the clients (Chen & Huang, 2003; Lee, Chiu, Lu, & Chen, 2002; Lopez & Saidenberg, 2000; West, 2000).

So far, linear probability and multivariate conditional probability models, the recursive partitioning algorithm, artificial intelligence, multi-criteria decision making, mathematical programming have been proposed to support the credit decision (Bryant, 1997; Butta, 1994; Cielen & Vanhoof, 1999; Coakley & Brown, 2000; Davis, Edelman, & Gammerman, 1992; Diakoulaki, Mavrotas, & Papayan nakis, 1992; Dimitras, Zanakis, & Zopounidis, 1996;

* Corresponding author. Tel.: +82 54 770 2317; fax: +82 54 770 2532.
E-mail addresses: jaemin@ccs.sogang.ac.kr (J.H. Min), chanlee@dongguk.ac.kr (Y.-C. Lee).

0957-4174/$ - see front matter © 2005 Published by Elsevier Ltd.
doi:10.1016/j.eswa.2004.12.008

Emel et al., 2003; Falbo, 1991; Frydman, Altman, & Kao, 1985; Jo & Han, 1996; Lee, Sung, & Chang, 1999; Martin, 1997; Reichert, Cho, & Wagner, 1983; Roy, 1991; Tam & Kiang, 1992; Troutt et al., 1996; Zopounidis & Doumpos, 1998). In particular, artificial neural networks (ANNs) are most frequently used in previous literature since the power of prediction is known to be better than the others; however, it has been commonly reported that ANN models require a large amount of training data to estimate the distribution of input pattern, and they have difficulties of generalizing the results because of their overfitting nature. In addition, it fully depends on researchers' experience or knowledge to preprocess data in order to select control parameters including relevant input variables, hidden layer size, learning rate, and momentum (Gemen et al., 1992; Lawrence, Giles, & Tsoi, 1997; Moody, 1992; Sarle, 1995; Smith, 1993; Weigend, 1994).

The purpose of this paper is to apply support vector machines (SVMs), a relatively new machine learning technique, to bankruptcy prediction problem and to provide a new model improving its prediction accuracy. Developed by Vapnik (1998), SVM is gaining popularity due to many attractive features and excellent generalization performance on a wide range of problems. Also, SVM embodies the structural risk minimization principle (SRM), which has been shown to be superior to traditional empirical risk minimization principle (ERM) employed by conventional neural networks. SRM minimizes an upper bound of generalization error as opposed to ERM that minimizes the error on training data. Therefore, the solution of SVM may be global optimum while other neural network models tend to fall into a local optimal solution, and overfitting is unlikely to occur with SVM (Cristianini & Shawe-Taylor, 2000; Gunn, 1998; Hearst, Dumais, Osman, Platt & Scholkopf, 1998; Kim, 2003). In addition, bearing in mind that the optimal parameter search plays a crucial role to build a bankruptcy prediction model with high prediction accuracy and stability, we employ a grid-search technique using 5-fold cross-validation to find out the optimal parameter values of kernel function of SVM. To evaluate the prediction accuracy of SVM, we also compare its performance with those of multiple discriminant analysis (MDA), logistic regression analysis (Logit), and three-layer fully connected back-propagation neural networks (BPNs).

2. Theoretical background

2.1. Statistical methods

In the late 1960s, discriminant analysis (DA) was introduced to create a composite empirical indicator of financial ratios. Using financial ratios, Beaver (1966) developed an indicator that best differentiated between failed and non-failed firms using univariate analysis techniques. The univariate approach was later improved and extended to multivariate analysis by Altman (1968). During the years that followed, many researchers attempted to increase the success of multiple discriminant analysis (MDA) in predicting business failures (Altman, Marco, & Varetto, 1994; Eisenbeis, 1978; Falbo, 1991; Peel, Peel, & Pope, 1986). Linear probability and multivariate conditional probability models (Logit and Probit) were introduced to the business failure prediction literature in late 1970s. The contribution of these methods was in estimating the odds of a firm's failure with probability (Martin, 1997; Ohlson, 1980a,b). In the 1980s, the recursive partitioning algorithm (RPA) based on a binary classification tree rationale was applied to this business failure prediction problem by Frydman et al. (1985); Srinivasan and Kim (1998).

2.2. Artificial intelligence methods

From the late 1980s, artificial intelligence (AI) techniques, particularly rule-based expert systems, case-based reasoning systems, and machine learning techniques such as artificial neural networks (ANNs) have been successfully applied to bankruptcy prediction (Desai, Conway, & Overstreet, 1997; Elmer & Borowski, 1988; Jensen, 1992; Malhotra & Malhotra, 2002; Markham & Ragsdale, 1995; Patuwo, Hu, & Hung, 1993; Srinivasan & Ruparel, 1990; West, 2000; Zhang, 2000; Zhang, Hu, Patuwo, & Indro, 1999). Recently, comparative studies between ANN and other techniques have been conducted. The results of these studies indicate that ANN shows better prediction accuracy (Barniv, Agarwal, & Leach, 1997; Bell, Ribar, & Verchio, 1990; Coates & Fant, 1993; Curram & Mingers, 1994; Desai, Crook, & Overstreet, 1996; Fanning & Cogger, 1994; Fletcher & Goss, 1993; Jo & Han, 1996; Lee, Jo, & Han, 1997; Odom & Sharda, 1990; Tam & Kiang, 1992; Wilson & Sharda, 1994); however, ANN has a difficulty in explaining the prediction results due to the lack of explanatory power, and suffers from difficulties with generalization because of overfitting. In addition, it needs too much time and efforts to construct a best architecture (Lawrence et al., 1997; Sarle, 1995).

2.3. Support vector machines

Support vector machines (SVMs) use a linear model to implement nonlinear class boundaries through some non-linear mapping input vectors into a high-dimensional feature space. The linear model constructed in the new space can represent a nonlinear decision boundary in the original space. In the new space, an optimal separating hyperplane (OSH) is constructed. Thus, SVM is known as the algorithm that finds a special kind of linear model, the *maximum margin hyperplane*. The maximum margin hyperplane gives the maximum separation between decision classes. The training examples that are closest to the maximum margin hyperplane are called *support vectors*. All other training examples are irrelevant for defining the binary class boundaries

(Cristianini & Shawe-Taylor, 2000; Gunn, 1998; Hearst et al., 1998; Vapnik, 1998).

SVM is simple enough to be analyzed mathematically since it can be shown to correspond to a linear method in a high dimensional feature space nonlinearly related to input space. In this sense, SVM may serve as a promising alternative combining the strengths of conventional statistical methods that are more theory-driven and easy to analyze, and more data-driven, distribution-free and robust machine learning methods. Recently, the SVM approach has been introduced to several financial applications such as credit rating, time series prediction, and insurance claim fraud detection (Fan & Palaniswami, 2000; Gestel et al., 2001; Huang, Chen, Hsu, Chen, & Wu, 2004; Kim, 2003; Tay & Cao, 2001; Viaene, Derrig, Baesens, & Dedene, 2002). These studies reported that SVM was comparable to and even outperformed other classifiers including ANN, CBR, MDA, and Logit in terms of generalization performance. Motivated by these previous researches, we apply SVM to the domain of bankruptcy prediction, and compare its prediction performance with those of MDA, Logit, and BPNs.

A simple description of the SVM algorithm is provided as follows. Given a training set $D=\{x_i, y_i\}_{i=1}^{N}$ with input vectors $x_i=(x_i^{(1)},\ldots,x_i^{(n)})^T \in \mathbb{R}^n$ and target labels $y_i \in \{-1,+1\}$, the support vector machine (SVM) classifier, according to Vapnik's original formulation, satisfies the following conditions

$$\begin{cases} \mathbf{w}^T\phi(x_i)+b \geq +1, & \text{if } y_i=+1 \\ \mathbf{w}^T\phi(x_i)+b \leq -1, & \text{if } y_i=-1 \end{cases} \tag{1}$$

which is equivalent to

$$y_i[\mathbf{w}^T\phi(x_i)+b] \geq 1, \quad i=1,\ldots,N \tag{2}$$

where $\mathbf{w}$ represents the weight vector and b the bias. Nonlinear function $\phi(\bullet):\mathbb{R}^n \to \mathbb{R}^{n_k}$ maps input or measurement space to a high-dimensional, and possibly infinite-dimensional, feature space. Eq. (2) then comes down to the construction of two parallel bounding hyperplanes at opposite sides of a separating hyperplane $\mathbf{w}^T\phi(x)+b=0$ in the feature space with the margin width between both hyperplanes equal to $2/(||\mathbf{w}||^2)$. In primal weight space, the classifier then takes the decision function form (3)

$$\operatorname{sgn}(\mathbf{w}^T\phi(x)+b) \tag{3}$$

Most of classification problems are, however, linearly non-separable. Therefore, it is general to find the weight vector using slack variable (ξ_i) to permit misclassification. One defines the primal optimization problem as

$$\underset{w,b,\xi}{\text{Min}} \frac{1}{2}\mathbf{w}^T\mathbf{w}+C\sum_{i=1}^{N}\xi_i \tag{4}$$

subject to

$$\begin{cases} y_i(\mathbf{w}^T\phi(x_i)+b) \geq 1-\xi_i, & i=1,\ldots,N \\ \xi_i \geq 0, & i=1,\ldots,N \end{cases} \tag{5}$$

where ξ_i's are slack variables needed to allow misclassifications in the set of inequalities, and $C \in \mathbb{R}^+$ is a tuning hyperparameter, weighting the importance of classification errors vis-à-vis the margin width. The solution of the primal problem is obtained after constructing the Lagrangian. From the conditions of optimality, one obtains a quadratic programming (QP) problem with Lagrange multipliers α_i's. A multiplier α_i exists for each training data instance. Data instances corresponding to non-zero α_i's are called *support vectors*.

On the other hand, the above primal problem can be converted into the following dual problem with objective function (6) and constraints (7). Since the decision variables are support vector of Lagrange multipliers, it is easier to interpret the results of this dual problem than those of the primal one.

$$\underset{\alpha}{\text{Max}} \frac{1}{2}\boldsymbol{\alpha}^T Q\boldsymbol{\alpha}-\mathbf{e}^T\boldsymbol{\alpha} \tag{6}$$

subject to

$$\begin{cases} 0 \leq \alpha_i \leq C, & i=1,\ldots,N \\ \mathbf{y}^T\boldsymbol{\alpha}=0 \end{cases} \tag{7}$$

In the dual problem above, $\mathbf{e}$ is the vector of all ones, Q is a $N\times N$ positive semi-definite matrix, $Q_{ij}=y_iy_jK(x_i,x_j)$, and $K(x_i,x_j)\equiv\phi(x_i)^T\phi(x_j)$ is the kernel. Here, training vectors x_i's are mapped into a higher (maybe infinite) dimensional space by function ϕ. As is typical for SVMs, we never calculate $\mathbf{w}$ or $\phi(x)$. This is made possible due to Mercer's condition, which relates mapping function $\phi(x)$ to kernel function $K(\cdot,\cdot)$ as follows.

$$K(x_i,x_j)=\phi(x_i)^T\phi(x_j) \tag{8}$$

For kernel function $K(\cdot,\cdot)$, one typically has several design choices such as the linear kernel of $K(x_i,x_j)=x_i^Tx_j$, the polynomial kernel of degree d of $K(x_i,x_j)=(\gamma x_i^Tx_j+r)^d$, $\gamma>0$, the radial basis function (RBF) kernel of $K(x_i,x_j)=\exp\{-\gamma||x_i-x_j||^2\}$, $\gamma>0$, and the sigmoid kernel of $K(x_i,x_j)=\tanh\{\gamma x_i^Tx_j+r\}$, where d, $r\in\mathbb{N}$ and $\gamma\in\mathbb{R}^+$ are constants. Then one constructs the final SVM classifier as

$$\operatorname{sgn}\left(\sum_{i}^{N}\alpha_iy_iK(x,x_i)+b\right) \tag{9}$$

The details of the optimization are discussed in (Chang & Lin, 2001; Cristianini & Shawe-Taylor, 2000; Gunn, 1998; Vapnik, 1998).

3. Research design

3.1. Data collection and preprocessing

The database used in this study was obtained from the Korea's largest credit guarantee organization that has the public policy of promoting growth of small and medium-sized enterprises in the nation. All of the non-bankruptcy cases are from medium-sized heavy industry firms in 2002 and the corresponding bankruptcy cases are also from the same industry. In general, however, the number of bankruptcy cases is smaller than that of non-bankruptcy cases; hence, we collected additional bankruptcy cases from the years of 2000, 2001, and 2002 in order to make the experimental data set consist of even number of bankruptcy and non-bankruptcy cases. The final sample of 1888 firms includes 944 bankruptcy and 944 non-bankruptcy cases placed in random order.

The original data are scaled into the range of $(-1, 1)$. The goal of linear scaling is to independently normalize each feature component to the specified range. It ensures the larger value input attributes do not overwhelm smaller value inputs; hence helps to reduce prediction errors (Hsu, Chang, & Lin, 2004).

In choosing financial ratios, we apply stepwise logistic regression analysis. Most previous studies using statistical methods such as discriminant analysis and logistic regression have selected independent variables through the stepwise regression analysis. In this study, several financial ratios are initially selected by the principal component analysis and *t*-test for the graphical analysis. Using the stepwise logistic regression, we also reduce the number of financial variables to a manageable set of 11.

In cases of SVM, MDA, and Logit, each data set is split into two subsets: a training set of 80% (1510) and a holdout set of 20% (378) of the total data (1888), respectively. The holdout data is used to test the results, which is not utilized to develop the model. In case of BPN, each data set is split into three subsets: a training set of 60% (1132), a validation set of 20% (378), and a holdout set of 20% (378) of the total data (1888) respectively, where the validation data is used to check the results.

3.2. Analysis steps

This study is conducted according to the analysis steps in Fig. 1.

In Step 1, we reduce the number of multi-dimensional financial ratios to two factors through the principal component analysis (PCA)[1] and calculate factor scores of all companies using factor loadings equal to or greater than 0.5. In Step 2, we compare the training performance among SVM models graphically using the factor scores.

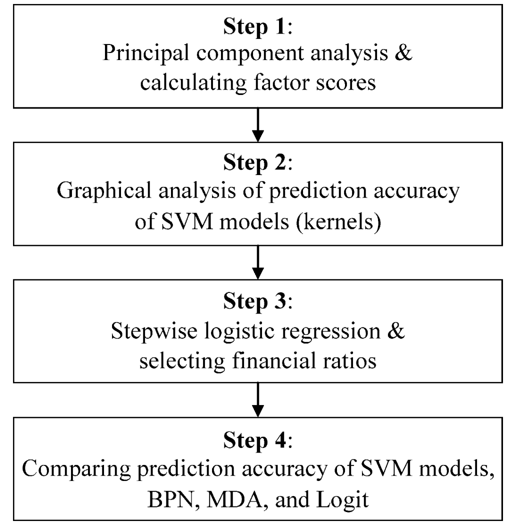


Fig. 1. Analysis steps.

In Step 3, we select financial ratios to be used in the bankruptcy prediction model via stepwise logistic regression. Lastly in Step 4, we compare the prediction accuracy of SVM, BPN, MDA, and Logit using the financial ratios selected in Step 3.

3.3. Support vector machines

In this study, the radial basis function (RBF) is used as the basic kernel function of SVM. There are two parameters associated with RBF kernels: C and γ. The upper bound C and the kernel parameter γ play a crucial role in the performance of SVMs (Hsu et al., 2004; Tay & Cao, 2001). Therefore, improper selection of these two parameters can cause overfitting or underfitting problems. Nevertheless, there is little general guidance to determine the parameter values of SVM. Recently, Hsu et al. (2004) suggested a practical guideline to SVM using grid-search and cross-validation, and this study will utilize it.

The goal is to identify optimal choice of C and γ so that the classifier can accurately predict unknown data (i.e. holdout data). Note that it may not be useful to achieve high training accuracy (i.e. classifiers accurately predict training data whose class labels are indeed known). Therefore, a common way is to separate training data into two parts, of which one is considered unknown in training the classifier. Then the prediction accuracy on this set can more precisely reflect the performance on classifying unknown data. An improved version of this procedure is cross-validation.

In v-fold cross-validation, we first divide the training set into v subsets of equal size. Sequentially one subset is tested using the classifier trained on the remaining $(v-1)$ subsets. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data that are correctly classified.

The cross-validation procedure can prevent the overfitting problem. In this study, we use a grid-search on C and γ using

[1] For the principal component analysis, we use statistically significant financial ratios by *t*-test.

5-fold cross-validation. Basically, all the pairs of (C, γ) are tried and the one with the best cross-validation accuracy is selected. We realize that trying exponentially growing sequences of C and γ is a practical method to identify optimal parameters (for example, $C=2^{-5}, 2^{-3},\ldots, 2^{15}$, $\gamma=2^{-15}, 2^{-13},\ldots, 2^{3}$).

The reasons why we use the grid-search are as follows. One is that psychologically we may not feel safe to use approximation methods or heuristics that avoid exhaustive parameter searches. The other reason is that the computational time to find the optimal parameter values by the grid-search is not much more than those by advanced methods since there are only two parameters. Furthermore, the grid-search can be easily parallelized because each pair (C, γ) is independent (Hsu et al., 2004). In addition, we show the prediction accuracy of other kernel functions such as linear, polynomial, and sigmoid functions to validate the RBF kernel adoption. We use *LIBSVM* software system (Chang & Lin, 2004) to perform SVM experiments.

3.4. Back-propagation neural networks

In this study, MDA, Logit and a three-layer fully connected back-propagation neural network (BPN) are used as benchmarks. In BPN, this study varies the number of nodes in the hidden layer and stopping criteria for training. In particular, 8, 12, 16, 24, 32 hidden nodes are used for each stopping criterion because BPN does not have a general rule for determining the optimal number of hidden nodes (Kim, 2003). For the stopping criteria of BPN, this study allows 100, 200, 300 learning epochs per one training example since there is little general knowledge for selecting the number of epochs. The learning rate is set to 0.1 and the momentum term is to 0.7. The hidden nodes use the hyperbolic tangent transfer function and the output node uses the same transfer function. We use *NeuroSolutions 4.32* to perform the BPN experiments.

3.5. Multiple discriminant analysis

Multiple discriminant analysis (MDA) tries to derive a linear combination of two or more independent variables that best discriminates among a priori defined groups, which in our case are bankruptcy and non-bankruptcy companies. This is achieved by the statistical decision rule of maximizing the between-group variance relative to the within-group variance. This relationship is expressed as the ratio of the between-group to the within-group variance. The MDA derives the linear combinations from an equation that takes the following form

$$Z = w_1x_1 + w_2x_2 + \ldots + w_nx_n \tag{10}$$

where Z is a discriminant score, $w_i(i=1, 2,\ldots, n)$ are discriminant weights, and $x_i(i=1, 2,\ldots, n)$ are independent variables, the financial ratios. Thus, each firm receives a single composite discriminant score which is then compared to a cut-off value, and with this information, we can determine to which group the firm belongs.

MDA does very well provided that the variables in every group follow a multivariate normal distribution and the covariance matrix for every group is equal. However, empirical experiments have shown that especially failed firms violate the normality condition.[2] In addition, the equal group variance condition is often violated. Moreover, multi-collinearity among independent variables may cause a serious problem, especially when the stepwise procedures are employed (Hair, Anderson, Tatham & Black, 1998).

3.6. Logistic regression analysis

Logistic regression (Logit) analysis has also been used to investigate the relationship between binary or ordinal response probability and explanatory variables. The method fits linear logistic regression model for binary or ordinal response data by the method of maximum likelihood. Among the first users of Logit analysis in the context of financial distress was Ohlson (1980a,b). Like MDA, this technique weights the independent variables and assigns a Z score in a form of failure probability to each company in a sample. The advantage of this method is that it does not assume multivariate normality and equal covariance matrices as MDA does. Logit analysis incorporates nonlinear effects, and uses the logistical cumulative function in predicting a bankruptcy, i.e.

$$\text{Probability of default} = \frac{1}{1+e^{-Z}} = \frac{1}{1+e^{-(w_0+w_1x_1+\ldots+w_nx_n)}} \tag{11}$$

Logit analysis uses the stepwise method to select final variables (financial ratios). The procedure starts by estimating parameters for variables forced into the model, i.e. intercept and the first possible explanatory variables. Next, the procedure computes the adjusted chi-squared statistic for all the variables not in the model and examines the largest of these statistics. If it is significant at the specified level, 0.01 in our study, the variable is entered into the model. Each selection step is followed by one or more elimination step, i.e. the variables already selected into the model do not necessarily stay. The stepwise selection process terminates if no further variable can be added to the model, or if the variable just entered into the model is the only variable removed in the subsequent elimination.

[2] Nevertheless, empirical studies have shown that the problems concerning normality assumptions do not weaken its classification capability, but its prediction ability.

Table 1
The result of principal component analysis

Variables	Factor1	Factor2
Growth rate of tangible assets	−0.058	0.091
Ordinary income to total assets[a]	0.541	0.698
Net income to total assets[a]	0.529	0.713
Ordinary income to stockholders' equity	0.374	0.446
Net income to stockholders' equity[a]	0.444	0.521
Ordinary income to sales[a]	0.342	0.761
Net income to sales[a]	0.327	0.771
Variable costs to sales[a]	0.544	−0.313
EBITDA to sales[a]	−0.246	0.714
Depreciation ratio	0.400	0.177
Interest expenses to total borrowings and bonds payable	0.008	0.048
Interest expenses to total expenses[a]	−0.761	−0.130
Interest expenses to sales[a]	−0.770	−0.220
Net interest expenses to sales[a]	−0.727	−0.199
Interest coverage ratio[a]	0.408	0.604
Break-even point ratio	0.430	−0.210
Stockholders' equity to total assets	0.053	0.304
Cash flow to previous year's short term loan	−0.103	0.431
Cash flow to short term loan	−0.128	0.448
Cash flow to total loan[a]	−0.103	0.541
Cash flow to total debt[a]	−0.174	0.568
Cash flow to interest expenses[a]	−0.136	0.519
Fixed ratio	−0.346	−0.066
Fixed assets to stockholders' equity and long-term liabilities	−0.214	−0.051
Total borrowings and bonds payable to sales[a]	−0.749	−0.279
Total assets turnover[a]	0.889	0.066
Stockholders' equity turnover[a]	0.503	−0.112
Capital stock turnover[a]	0.523	−0.027
Operating assets turnover[a]	0.866	0.067
Fixed assets turnover[a]	0.758	−0.157
Tangible assets turnover[a]	0.634	−0.178
Inventories turnover	0.141	0.266
Payables turnover	0.017	0.157
Gross value added to total assets and productivity of capital	0.485	0.451
Gross value added to property, plant, and equipment[a]	0.557	−0.071
Gross value added to sales	−0.396	0.485
Solvency ratio[a]	−0.593	0.062
Ordinary income to ordinary expenses[a]	0.027	0.566

[a] The financial ratios whose factor loading is equal to or greater than 0.5.

4. Empirical analysis

4.1. Principal component analysis

We conducted the principal component analysis (varimax rotation) to 38 financial ratios with significant *t*-value. Among them, 23 financial ratios turned out to have their factor loadings equal to or greater than 0.5. The list of 38 financial ratios and their respective factor loadings are summarized in Table 1.

The scatter plots of two factor scores calculated by using the factor loadings in Table 1 is represented in Fig. 2. In Fig. 2, two different colors denote two classes of the training and the holdout examples. Blue and purple bullets represent these two respective classes.

This study graphically compared the training performance between SVM models using the RBF and polynomial kernel function to solve the problem whose split boundary is complex like Fig. 2.

4.2. Graphical analysis of prediction accuracy of SVM models

In this study, we set the upper bound C and kernel parameter γ to 2^{15} and 2^{-1}, respectively. These are identified from the grid-search using 5-fold cross-validation (see Fig. 3). A detailed description of this process is discussed in Section 4.4.

As shown in Fig. 3, the optimal pair of (C, γ) is found at the green zone which have $C=2^{15}$ and $\gamma=2^{-1}$, respectively. Using these parameters, we run the SVM models in case of the RBF and polynomial kernels. Figs. 4 and 5 graphically compare the prediction performance of SVM model using the RBF with one using the polynomial kernel function.

First, the result of implementing RBF kernel SVM for the training data and holdout data is shown in Fig. 4, where target variable is bankruptcy or non-bankruptcy (0 or 1), and input variables are two factor scores.

SVM using the RBF kernel trains the data corresponding to each Gaussian in order to determine the support vectors.

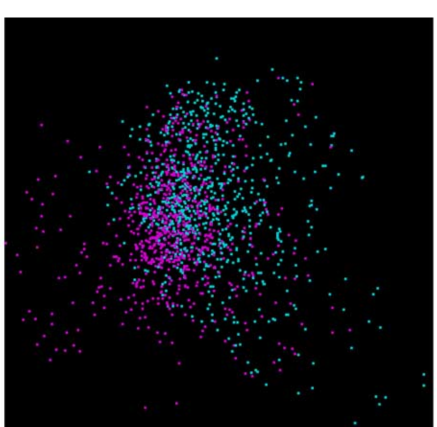
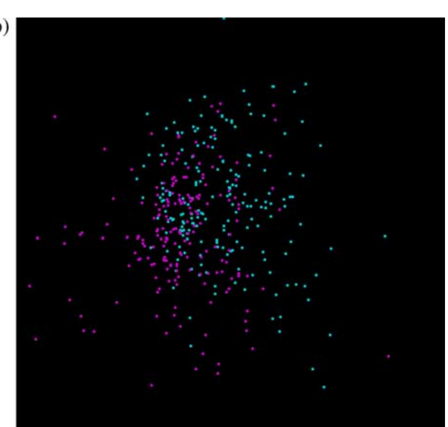
(a)

(b)


Fig. 2. The scatter plot of factor scores. (a) The scatter plot of training data. (b) The scatter plot of holdout data.

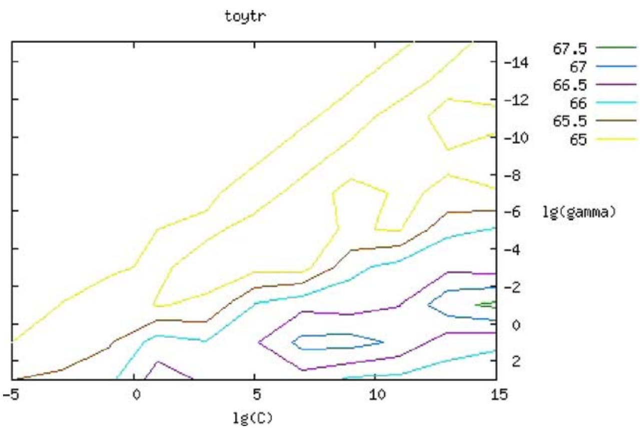


Fig. 3. Grid-search on $C=2^{-5}, 2^{-3}, \ldots, 2^{15}$ and $\gamma=2^{-15}, 2^{-13}, \ldots, 2$.

These support vectors allow the network to rapidly converge on the data boundaries and consequently classify the inputs. As shown in Fig. 4, SVM classifies two groups appropriately, where the prediction accuracy is 67.2185% (training data) and 66.1376% (holdout data), respectively.

Second, the result of implementing the polynomial kernel SVM according to training data and test data is represented in Fig. 5. In the polynomial kernel, there exists additional kernel parameter, degree d. We vary degree d from 1 to 3.

As shown in Fig. 5, the underfitting or overfitting problem is occurred in case of the polynomial kernel function. In particular, the prediction performance of the training data is rather lower than that of the holdout data when degree d is either 1 or 2, for example, $65.5629\% < 70.1058\%$ ($d=1$) and $66.8212\% < 68.5185\%$ ($d=2$), respectively. This is a typical underfitting phenomenon in machine learning. On the other hand, in case of $d=3$, although the prediction performance of the training data is higher than that of the RBF kernel ($67.8808\% > 67.2185\%$), the prediction performance of the holdout data is rather lower than that of the RBF kernel ($65.6085\% < 66.1376\%$). This is a typical overfitting phenomenon in machine learning. Therefore, we need to make extra efforts to find the best value of degree d in the polynomial kernel SVM model.

4.3. Selecting financial ratios

In Section 4.2, we graphically showed the training performance of two SVM models using the principal component of 38 financial ratios and found out the performance of SVM was attractive. Based on these results, we conducted a comparative analysis of the prediction accuracy among SVM, ANN, MDA, and Logit. For the analysis, we selected the final financial ratios through the stepwise logistic regression analysis. The financial ratios finally used in the analysis are summarized in Table 2.

The performance of bankruptcy prediction fully depends on the input variables. In previous literature (Jo & Han, 1996; Lee, Han, & Kwon, 1996; Lee et al., 1997; Park & Han, 2002), there are many cases that the economic interpretation of them is very difficult but they are significant to classify business units into bankruptcy or non-bankruptcy group. We select the final input variables (financial ratios) among the ratios that are mainly used for the corporate analysis at the Bank of Korea. Therefore, the financial ratios selected in this study could be considered economically interpretable as well as useful to figure out the financial credibility.

4.4. Comparing prediction accuracy of SVM models, ANN, MDA, and Logit

4.4.1. SVM models

In SVM, each data set is split into two subsets: a training set of 80% (1510) and a holdout set of 20% (378) of the total data (1888), respectively.

We must first decide which kernels to select for implementing SVM; and then the penalty parameter C and kernel parameters are chosen. One of the advantages of the linear kernel SVM is that there are no parameters to tune except for constant C. But the upper bound C on coefficient α_i affects the prediction performance for the cases where the training data is not separable by a linear SVM (Drucker, Wu, & Vapnik, 1999). For the nonlinear SVM, there is an additional parameter, the kernel parameter, to tune. There are

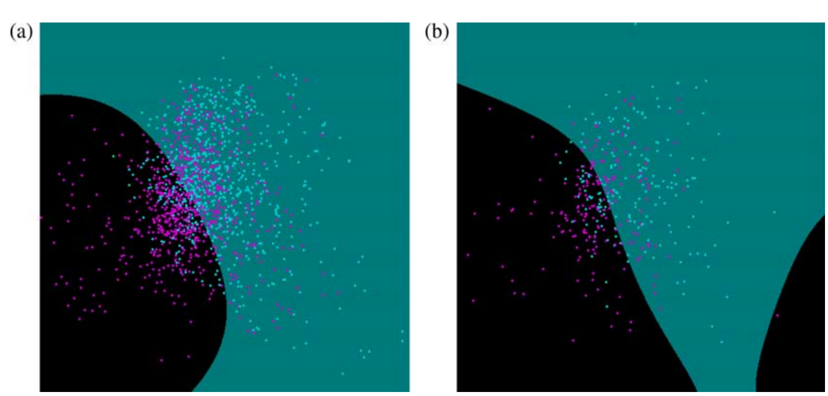


Fig. 4. Graphical display of the prediction performance of RBF kernel SVM. (a) Training data. (b) Holdout data.

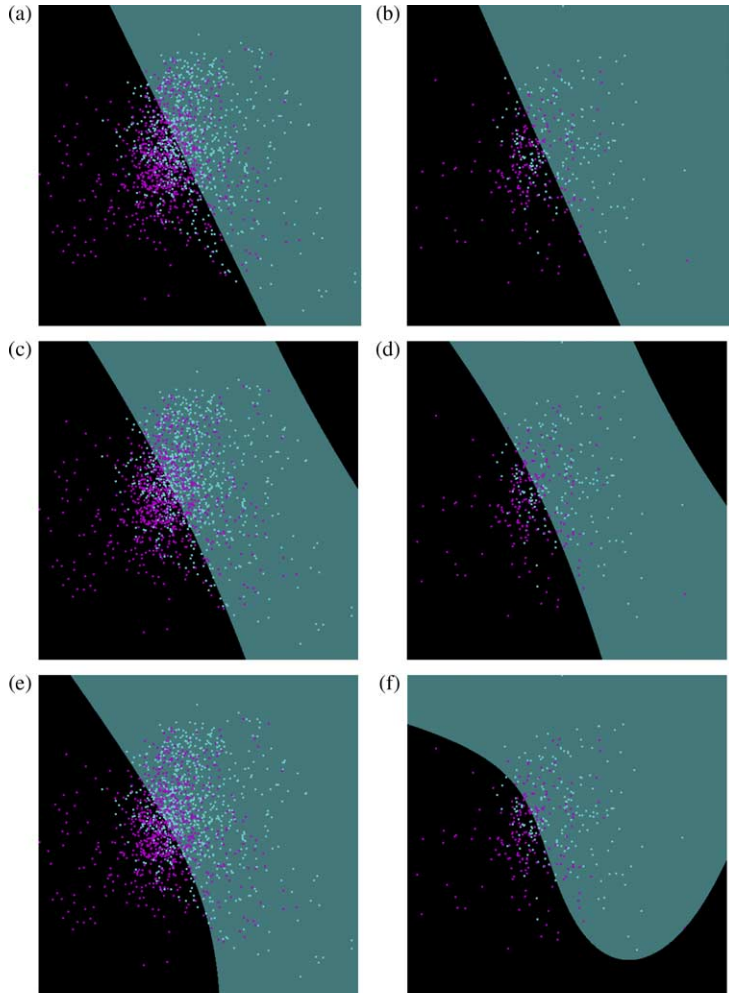


Fig. 5. Graphical display of the prediction performance of polynomial kernel SVM. (a) Training data ($d=1$). (b) Holdout data ($d=1$). (c) Training data ($d=2$). (d) Holdout data ($d=2$). (e) Training data ($d=3$). (f) Holdout data ($d=3$).

three kernel functions for nonlinear SVM including the radial basis function (RBF), the polynomial, and the sigmoid.

The RBF kernel nonlinearly maps the samples into a higher dimensional space unlike the linear kernel, so it can handle the case when the relation between class labels and attributes is nonlinear. The sigmoid kernel behaves like the RBF for certain parameters; however, it is not valid under some parameters (Vapnik, 1998). The polynomial function takes a longer time in the training stage of SVM, and it is reported to provide worse results than the RBF function in the previous studies (Huang et al., 2004; Kim, 2003; Tay & Cao, 2001). In addition, the polynomial kernel has more hyperparameters than the RBF kernel and may go to infinity or zero while the degree is large. Thus, this study uses the RBF kernel SVM as the default model.

There are two parameters associated with the RBF kernels: C and γ. It is not known beforehand which values of C and γ are the best for one problem; consequently, some kind of model selection (parameter search) approach must be employed (Hsu et al., 2004). This study conducts a grid-search to find the best values of C and γ using 5-fold

Table 2
The selected financial ratios

Variables	Formula
Ordinary income to stockholders' equity	Ordinary income ÷ total assets
Variable costs to sales	Variable costs ÷ sales
Interest expenses to total borrowings and bonds payable	Interest expenses ÷ total borrowings and bonds payable
Interest expenses to sales	Interest expenses ÷ sales
Break-even point ratio	Sales at break-even point ÷ sales Sales at break-even point: ÷ fixed costs-non-operating revenue) ÷ (1 − (variable costs ÷ sales))
Stockholders' equity to total assets	Stockholders' equity ÷ total assets
Cash flow to short term loan	Cash flow from operating ÷ short term loan
Fixed ratio	Fixed assets ÷ stockholders' equity
Capital stock turnover	Sales ÷ capital stock
Operating assets turnover	Sales ÷ operating assets
Fixed assets turnover	Sales ÷ fixed assets

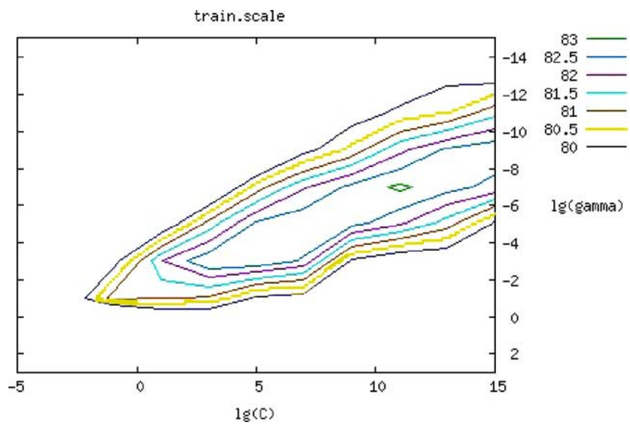


Fig. 6. Grid-search on $C=2^{-5}, 2^{-3}, \ldots, 2^{15}$ and $\gamma=2^{-15}, 2^{-13}, \ldots, 0$.

cross-validation. Pairs of (C, γ) are tried and the one with the best cross-validation accuracy is picked. After conducting the grid-search for training data, we found that the optimal (C, γ) was $(2^{11}, 2^{-7})$ with the cross-validation rate of 83.1126% (see Fig. 6). Table 3 summarizes the results of the grid-search using 5-fold cross-validation.

After the optimal (C, γ) was found, the whole training data was trained again to generate the final classifier. The overall prediction accuracy of the holdout data turned out to be 83.0688% (bankruptcy: 82.0106%, non-bankruptcy: 84.1270%), where the prediction accuracy of the training data was 88.0132%. This study conducts additional experiments with other kernel functions such as the linear, polynomial, and sigmoid. The performance of four different kernel functions is summarized in Table 4.

As shown in Table 4, the polynomial kernel shows a bit higher performance than the RBF kernel when degree d is set to 2 for the holdout data. However, the polynomial kernel, as described earlier, needs more hyperparameters than the RBF kernel and may go to infinity or zero while degree is large. In particular, the underfitting or overfitting problem can occur with specific C and γ. Therefore, we adopt the RBF kernel as the final kernel function of SVM model.

4.4.2. BPN models

In BPN, each data set is split into three subsets: a training set of 60% (1132), a holdout set of 20% (378), and a validation set of 20% (378) of the total data (1888), respectively. The results of three-layer BPN according to parameter adjustment are summarized in Table 5. We can see that the prediction accuracy of training data tends to be higher as the learning epoch increases. The best prediction accuracy for the holdout data was found when the epoch was 300 and the number of hidden nodes was 24. The prediction accuracy of the holdout data turned out to be 82.5397% (bankruptcy: 79.3651%, non-bankruptcy: 85.7143%), and that of the training data was 85.2474%.

As shown in Table 5, the best prediction accuracy of BPN (88.1625%) for the training data is almost same as that of SVM (88.0132%). BPN, however, reveals an overfitting problem when the learning epoch is 300 compared with SVM. As the learning epoch increases, the prediction accuracy of the training data tends to improve while the prediction accuracy of the holdout data rather decreases.

From the results of the empirical experiment, we can conclude that SVM shows better performance than BPN in bankruptcy prediction while avoiding overfitting problem and exhaustive parameter search.

4.4.3. MDA and Logit

In MDA and Logit, each data set is split into two subsets: a training set of 80% and a holdout set of 20% of the total data (1888), respectively. Table 6 summarizes the prediction performance of MDA and Logit. As shown in Table 6, Logit slightly outperforms MDA for the holdout data.

4.4.4. Prediction performance comparisons

Table 7 compares the best prediction performance of SVM, BPN, MDA, and Logit in training and holdout data, and shows that SVM outperforms BPN, MDA and Logit by 0.5, 4.8, and 3.9%, respectively for the holdout data.

In addition, we conducted McNemar test to examine whether SVM significantly outperformed the other three

Table 3
The result of grid-search using 5-fold cross-validation

C	γ									
	2^{3}	2^{1}	2^{-1}	2^{-3}	2^{-5}	2^{-7}	2^{-9}	2^{-11}	2^{-13}	2^{-15}
2^{-5}	47.2185	47.2185	74.4371	71.9205	57.2848	47.2185	47.2185	47.2185	47.2185	47.2185
2^{-3}	47.2185	62.9801	79.1391	77.2848	74.4371	67.8808	47.2185	47.2185	47.2185	47.2185
2^{-1}	48.1457	74.3709	81.2583	79.6689	77.7483	75.3642	68.1457	47.2185	47.2185	47.2185
2^{1}	62.5828	77.5497	81.0596	81.9868	79.2715	78.1457	75.6954	68.0795	47.2185	47.2185
2^{3}	62.5828	77.6159	80.9272	82.9139	81.1921	78.8079	78.0795	75.5629	67.8808	47.2185
2^{5}	62.5828	77.5497	79.8675	82.9139	82.5828	80.7285	78.4106	77.8808	75.5629	67.7483
2^{7}	62.5828	77.5497	79.7351	82.4503	82.8477	81.9205	79.7351	78.0795	77.6159	75.5629
2^{9}	62.5828	77.5497	79.7351	79.9338	82.7152	82.6490	80.6623	79.6026	78.0795	77.5497
2^{11}	62.5828	77.5497	79.7351	79.3377	82.1192	83.1126	81.8543	80.1325	79.3377	78.0132
2^{13}	62.5828	77.5497	79.7351	79.3377	81.2583	82.7152	82.5828	80.4636	79.8013	79.4702
2^{15}	62.5828	77.5497	79.7351	79.2053	79.9338	82.3179	82.8477	81.3245	79.6026	79.8013

Table 4
The performance of SVM models

Model (kernel)	C	γ	d	Prediction accuracy (%)	
				Training data	Holdout data
Linear	2^{11}	N/A	N/A	80.2649	77.2487
RBF	2^{11}	2^{-7}	N/A	88.0132	83.0688
Polynomial[a]	2^{11}	2^{-7}	1	80.3311	77.2487
			2	86.6225	83.8624
			3	88.4768	82.0106
Sigmoid[a]	2^{11}	2^{-7}	N/A	72.6490	71.1640

[a] Parameter r is set to 1.

Table 5
The performance of BPN

Learning epoch	Hidden nodes	Prediction accuracy (%)	
		Training data	Holdout data
100	8	83.1272	80.6878
	12	82.9505	80.4233
	16	81.5371	80.1587
	24	81.0954	79.8942
	32	83.5689	81.2169
200	8	84.8057	80.9524
	12	86.7491	81.2169
	16	82.6855	81.2169
	24	85.1590	81.2169
	32	85.2474	81.7460
300	8	84.8940	82.0106
	12	86.1307	81.4815
	16	87.0141	80.6878
	24	85.2474	82.5397
	32	88.1625	81.4815

models. As a nonparametric test for two related samples, it is particularly useful for before-after measurement of the same subjects (Kim, 2003).

Table 8 shows the results of the McNemar test to statistically compare the prediction performance for the holdout data among four models. As shown in Table 8, SVM outperforms MDA and Logit at 5% and 10% statistical significance level, respectively. However, SVM does not significantly outperform BPN. In addition, Table 8 also shows that the prediction performance among BPN, MDA, and Logit do not significantly differ each other.

Table 6
The performance of MDA and Logit

Model	Prediction accuracy (%)	
	Training data	Holdout data
MDA	78.8079	79.1391
Logit	79.8676	78.3069

Table 7
The best prediction accuracy of SVM, BPN, MDA, and Logit (hit ratio: %)

	SVM	BNN	MDA	Logit
Training data	88.0132	85.2474	78.8079	79.8676
Holdout data	83.0688	82.5397	79.1391	78.3069

5. Conclusions

In this paper, we applied SVM to bankruptcy prediction problem, and showed its attractive prediction power compared to the existing methods. Mapping input vectors into a high-dimensional feature space, SVM transforms complex problems (with complex decision surfaces) into simpler problems that can use linear discriminant functions, and it has been successfully introduced in several financial applications recently. Achieving similar to or better performance than BPNs in practical applications, SVM can conduct classification learning with relatively small amount of data. Also, embodying the structural risk minimization principle (SRM), SVM may prevent the overfitting problem and makes its solution global optimum since the feasible region is convex set.

Particularly in this study, we utilize a grid-search technique using 5-fold cross-validation in order to choose optimal values of the upper bound C and the kernel parameter γ that are most important in SVM model selection. Selecting the optimal parameter values through the grid-search, we could build a bankruptcy prediction model with high stability and prediction power. To validate the prediction performance of this model, we statistically compared its prediction accuracy with those of standard three-layer fully connected BPNs, MDA, and Logit, respectively. The results of empirical analysis showed that SVM outperformed the other methods. With these results, we claim that SVM can serve as a promising alternative for the bankruptcy prediction.

While this study used RBF kernel as a basic kernel function of SVM model, it should be noted that the appropriate kernel function can be problem-specific; hence it remains an interesting topic for further study to derive judicious procedures to select proper kernel functions and the corresponding parameter values according to the types of classification problems.

Table 8
McNemar values (p-values) for the pairwise comparison of performance

	BPN	MDA	Logit
SVM	1.750[c] (0.186)[d]	4.470 (0.035)**	2.857 (0.091)*
BPN		0.356 (0.551)	0.103 (0.749)
MDA			0.595 (0.440)

*$P<0.1$, **$P<0.05$.
[b] P-value.
[a] McNemar statistic.

References

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*(4), 589–609.

Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis comparisons using linear discriminant analysis and neural networks. *Journal of Banking and Finance*, *18*(3), 505–529.

Barniv, R., Agarwal, A., & Leach, R. (1997). Predicting the outcome following bankruptcy filing: A three-state classification using neural networks. *International Journal of Intelligent Systems in Accounting, Finance and Management*, *6*(3), 177–194.

Basel Committee on Banking Supervision (1999). Credit Risk Modeling: Current Practices and Applications. Basel Committee Publications.

Beaver, W. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, *5*, 71–111.

Bell, T., Ribar, G., Verchio, J. (1990). Neural nets vs. logistic regression: A comparison of each model's ability to predict commercial bank failures. In Proceedings of the 1990 Deloitte and Touche/University of Kansas Symposium on Auditing Problems (pp. 29–58).

Bryant, S. M. (1997). A case-based reasoning approach to bankruptcy prediction modeling. *International Journal of Intelligent Systems in Accounting, Finance and Management*, *6*(3), 195–214.

Butta, P. (1994). Mining for financial knowledge with CBR. *AI Expert*, *9*(2), 34–41.

Chang, C.-C., Lin, C.-J. (2004). LIBSVM: A library for support vector machines. Technical Report, Department of Computer Science and Information Engineering, National Taiwan University. Available at http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf.

Chen, M.-C., & Huang, S.-H. (2003). Credit scoring and rejected instances reassigning through evolutionary computation techniques. *Expert Systems with Applications*, *24*, 433–441.

Cielen, A., & Vanhoof, K. (1999). *Bankruptcy prediction using a data envelopment analysis Manuscript*. Diebenpeek: Limburg University.

Coakley, J. R., & Brown, C. E. (2000). Artificial neural networks in accounting and finance: Modeling issues. *International Journal of Intelligent Systems in Accounting, Finance and Management*, *9*(2), 119–144.

Coates, P., & Fant, L. (1993). Recognizing financial distress patterns using a neural network tool. *Financial Management*, *22*(3), 142–155.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines*. Cambridge, England: Cambridge University Press.

Curram, S. P., & Mingers, J. (1994). Neural networks, decision tree induction and discriminant analysis: An empirical comparison. *Journal of Operational Research Society*, *45*(4), 440–450.

Davis, R. H., Edelman, D. B., & Gammerman, A. J. (1992). Machine learning algorithms for credit-card applications. *IMA Journal of Mathematics Applied in Business and Industry*, *4*, 43–51.

Desai, V. S., Conway, J. N., & Overstreet, G. A. (1997). Credit scoring models in the credit union environment using neural networks and genetic algorithms. *IMA Journal of Mathematics Applied in Business and Industry*, *8*, 324–346.

Desai, V. S., Crook, J. N., & Overstreet, G. A. (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research*, *95*, 24–37.

Diakoulaki, D., Mavrotas, G., & Papayannakis, L. (1992). A multicriteria approach for evaluating the performance of industrial firms. *Omega*, *20*(4), 467–474.

Dimitras, A. I., Zanakis, S. H., & Zopounidis, C. (1996). A survey of business failure with an emphasis on prediction methods and industrial applications. *European Journal of Operational Research*, *90*, 487–513.

Drucker, H., Wu, D., & Vapnik, V. N. (1999). Support vector machines for spam categorization. *IEEE Transactions Neural Networks*, *10*(5), 1048–1054.

Eisenbeis, R. A. (1978). Problems in applying discriminant analysis in credit scoring models. *Journal of Banking and Finance*, *2*, 205–219.

Elmer, P. J., & Borowski, D. M. (1988). An expert system approach to financial analysis: The case of S&L bankruptcy. *Financial Management Autumn*, *17*(3), 66–76.

Emel, A. B., Oral, M., Reisman, A., & Yolalan, R. (2003). A credit scoring approach for the Commercial Banking Sector. *Socio-Economic Planning Sciences*, *37*, 103–123.

Falbo, P. (1991). Credit-scoring by enlarged discriminant models. *Omega*, *19*(4), 275–289.

Fan, A., Palaniswami, M. (2000). Selecting bankruptcy predictors using a support vector machine approach. In Proceedings of the international joint conference on neural networks.

Fanning, K., & Cogger, K. (1994). A comparative analysis of artificial neural networks using financial distress prediction. *International Journal of Intelligent Systems in Accounting, Finance and Management*, *3*(3), 241–252.

Fletcher, D., & Goss, E. (1993). Forecasting with neural networks and application using bankruptcy data. *Information and Management*, *24*, 159–167.

Frydman, H. E., Altman, E. I., & Kao, D. (1985). Introducing recursive partitioning for financial classification: The case of financial distress. *The Journal of Finance*, *40*(1), 269–291.

Gestel, T. V., Suykens, J. A. K., Baestaens, D.-E., Lambrechts, A., Lanckriet, G., Vandaele, B., et al. (2001). Financial time series prediction using least squares support vector machines within the evidence framework. *IEEE Transactions Neural Networks*, *12*(4), 809–821.

Gunn, S.R. (1998). Support vector machines for classification and regression. Technical Report, University of Southampton.

Hair, J. F., Anderson, R. E., Tatham, R. E., & Black, W. C. (1998). *Multivariate data analysis with readings*. Prentice Hall.

Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent System*, *13*(4), 18–28.

Hsu, C.-W., Chang, C.-C., Lin, C.-J. (2004). A practical guide to support vector classification. Technical Report, Department of Computer Science and Information Engineering, National Taiwan University. Available at http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., & Wu, S. (2004). Credit rating analysis with support vector machine and neural networks: A market comparative study. *Decision Support Systems*, *37*, 543–558.

Jensen, H. L. (1992). Using neural networks for credit scoring. *Managerial Finance*, *18*, 15–26.

Jo, H., & Han, I. (1996). Integration of case-based forecasting, neural network, and discriminant analysis for bankruptcy prediction. *Expert Systems with Applications*, *11*, 415–422.

Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, *55*, 307–319.

Lawrence, S., Giles, C.L., Tsoi, A.-C. (1997). Lessons in neural network training: Overfitting may be harder than expected. In Proceedings of the fourteenth national conference on artificial intelligence, AAAl-97, (pp. 540–545). Mento Park, CA: AAAl Press.

Lee, G., Sung, T. K., & Chang, N. (1999). Dynamics of modeling in data mining: Interpretive approach to bankruptcy prediction. *Journal of Management Information Systems*, *16*, 63–85.

Lee, H., Jo, H., & Han, I. (1997). Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis. *Expert Systems With Applications*, *13*, 97–108.

Lee, K., Han, I., & Kwon, Y. (1996). Hybrid neural networks for bankruptcy predictions. *Decision Support Systems*, *18*, 63–72.

Lee, T.-S., Chiu, C.-C., Lu, C.-J., & Chen, I.-F. (2002). Credit scoring using hybrid neural discriminant technique. *Expert Systems with Applications*, *23*, 245–254.

Lopez, J. A., & Saidenberg, M. R. (2000). Evaluating credit risk models. *Journal of Banking and Finance*, *24*(1–2), 151–165.

Malhotra, R., & Malhotra, D. K. (2002). Differentiating between good credits and bad credits using neuro-fuzzy systems. *European Journal of Operational Research*, *136*(2), 190–211.

Markham, I. S., & Ragsdale, C. T. (1995). Combining neural networks and statistical predictions to solve the classification problem in discriminant analysis. *Decision Sciences*, *26*(2), 229–242.

Martin, D. (1997). Early warning of bank failure: A logit regression approach. *Journal of Banking and Finance*, *1*, 249–276.

Moody, J. E. (1992). The effective number of parameters: An analysis of generalization and regularization in nonlinear learning systems. *NIPS*, *4*, 847–854.

Odom, M., Sharda, R. (1990). A neural network model for bankruptcy prediction. In Proceedings of the international joint conference on neural networks (II-163-II-168).

Ohlson, J. A. (1980a). Financial ratios and probabilistic prediction of bankruptcy. *Journal of Accounting Research*, *18*(1), 109–131.

Ohlson, J. A. (1980b). Financial ratios and probabilistic prediction of bankruptcy prediction. *Journal of Accounting Research*, *18*(1), 109–131.

Park, C.-S., & Han, I. (2002). A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction. *Expert Systems with Applications*, *23*(1), 255–264.

Patuwo, E., Hu, M. H., & Hung, M. S. (1993). Two-group classification using neural networks. *Decisions Science*, *24*(4), 825–845.

Peel, M. J., Peel, D. A., & Pope, P. F. (1986). Predicting corporate failure-some results for the UK corporate sector. *Omega*, *14*(1), 5–12.

Reichert, A. K., Cho, C. C., & Wagner, G. M. (1983). An examination of the conceptual issues involved in developing credit-scoring models. *Journal of Business and Economic Statistics*, *1*, 101–114.

Roy, B. (1991). The outranking approach and the foundations of ELECTRE methods. *Theory and Decision*, *31*, 49–73.

Sarle, W.S. (1995). Stopped training and other remedies for overfitting. In Proceedings of the twenty-seventh symposium on the interface of computing science and statistics (pp. 352–360).

Smith, M. (1993). *Neural networks for statistical modeling*. New York: Van Nostrand Reinhold.

Srinivasan, V., & Ruparel, B. (1990). CGX: An expert support system for credit granting. *European Journal of Operational Research*, *45*, 293–308.

Srinivasan, V., & Kim, Y. H. (1988). Designing expert financial systems: A case study of corporate credit management. *Financial Management*, *5*, 32–43.

Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science*, *38*(7), 926–947.

Tay, F. E. H., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, *29*, 309–317.

Vapnik, V. (1998). *Statistical learning theory*. New York: Springer.

Viaene, S., Derrig, R. A., Baesens, B., & Dedene, G. (2002). A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection. *The Journal of Risk and Insurance*, *69*(3), 373–421.

Weigend, A. (1994). On overfitting and the effective number of hidden units. In Proceedings of the 1993 connectionist models summer school (pp. 335–342).

West, D. (2000). Neural network credit scoring models. *Computers and Operations Research*, *27*, 1131–1152.

Wilson, R., & Sharda, R. (1994). Bankruptcy prediction using neural networks. *Decision Support Systems*, *11*, 545–557.

Zhang, G. P. (2000). Neural networks for classification: A survey. *IEEE Transactions Systems, Man, and Cybernetics-Part C: Applications and Reviews*, *30*(4), 451–462.

Zhang, G. P., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, *116*, 16–32.