# Financial Distress Model Prediction using SVM+

Bernardete Ribeiro, *Member, IEEE* and Catarina Silva, *Member, IEEE*
and Armando Vieira, and A. Gaspar-Cunha, and João C. das Neves

***Abstract*— Financial distress prediction is of great importance to all stakeholders in order to enable better decision-making in evaluating firms. In recent years, the rate of bankruptcy has risen and it is becoming harder to estimate as companies become more complex and the asymmetric information between banks and firms increases. Although a great variety of techniques have been applied along the years, no comprehensive method incorporating an holistic perspective had hitherto been considered. Recently, SVM+ a technique proposed by Vapnik [17] provides a formal way to incorporate privileged information onto the learning models improving generalization. By exploiting additional information to improve traditional inductive learning we propose a prediction model where data is naturally separated into several groups according to the size of the firm. Experimental results in the setting of a heterogeneous data set of French companies demonstrated that the proposed model showed superior performance in terms of prediction accuracy in bankruptcy prediction and misclassification cost.**

## I. Introduction

The hit rate of firms insolvency has increased exponentially during last year due to the global economic crisis. As a consequence, there is an ever-increasing need for fast automated recognition systems for bankruptcy prediction. The extensive recent literature shows that at the heart of the business failure problem is the asymmetric information between banks and firms. As a consequence, the development of analytical tools to determine which financial information is more relevant to predict financial distress has gained popularity along with the design of early warning systems that predict bankruptcy [10].

Enterprise bankruptcy forecasting is very important to all stakeholders (banks, insurance firms, creditors, and investors) to manage credit risk associated with counterparts. Although it is a widely studied topic, it is becoming harder to estimate as companies become more complex and develop more sophisticated schemes to hide their real situation. On the other hand as the inability to discharge all debts as they come due (insolvency) increases, the need for substantially more accurate predicting models and, at the same time, for faster decision-maker systems becomes crucial.

Bernardete Ribeiro is with Department of Informatics Engineering and CISUC, Center of Informatics and Systems of University of Coimbra, Portugal (bribeiro@dei.uc.pt).

Catarina Silva is with School of Technology and Management, Polytechnic Institute of Leiria (IPL) and CISUC, Portugal (catarina@dei.uc.pt).

Armando Vieira is with Physics Department, Polytechnic Institute of Porto (ISEP), Portugal (asv@isep.pt).

A. Gaspar-Cunha is with Institute of Polymers and Composites (IPC/I3N), University of Minho, Guimarães, Portugal (agc@dep.minho.pt).

João C. das Neves is with School of Economics and Management (ISEG), Technical University of Lisbon (jcneves@iseg.utl.pt).

The health of a firm in a highly competitive business environment is dependent upon its capability of achieving profitability and financial solvency. This means that a firm becomes unhealthy, or deteriorates to the point where it is in danger of suffering business failure, when it loses its competence to maintain profitability and financial solvency [19]. Business failure is not only common with new start-ups but also with listed companies, and it can easily happen to firms of any and all sizes.

In Portugal bankruptcies increased by 49% and start-ups fell 15% in 2009. In fact, more than 1250 companies were declared insolvent in 2009, representing an increase of 49 percent with respect to the previous year, while only 30,412 new businesses were initiated, which means a decrease of 15 percent. According to the Annual Survey of Insolvency and Constitutions Company Coface, during the year 2009, the court declared 1251 bankruptcy of enterprises, more 410 than in the previous year, corresponding to the above increase.

For this study we used a large database of French companies. This database is very detailed containing information on a wide set of financial ratios spawning over a period of several years. It contains up to three thousands distressed companies and about sixty thousand healthy ones. The financial Coface Data set (French credit risk provider) is strongly heterogeneous with regards to the type of companies and their financial statuses. A great deal of research has been pursued disregarding this aspect. In this paper we focus on improving financial distress decision-making by structuring information into heterogeneous groups of companies and by using advanced SVM+ techniques.

The rest of the paper is organized as follows. Section II describes relevant background knowledge on bankruptcy prediction and related work to easily understand the analysis conducted in our experiments to be presented and discussed further in Section IV. In Section III we introduce SVM+ algorithm. In Section IV we include the description of the historic solvent (and default) firm data collected and labeled appropriately for bankruptcy prediction model in a case study of the French Market, describe performance metrics and present (and discuss) the results. Finally, in Section V we present the conclusions and point out further lines of work.

## II. Background on Bankruptcy Prediction

The problem is stated as follows: given a set of parameters (mainly of financial nature) that describe the situation of a company over a given period, predict the probability that the company may become bankrupted during the following year. Neural Networks (NNs) are particularly suited for predicting the bankrupt probability, thus they are a strategic choice

978-1-4244-8126-2/10/$26.00 ©2010 IEEE

among other methods. Likewise, their properties make them often used in financial applications because of their excellent performances of treating non-linear data with self-learning capability [7]. A review of the topic of bankruptcy prediction with emphasis on NN models is given in [3]. More recently, in [11] there is a broad coverage of a wide range of other intelligent techniques such as fuzzy set theory (FS), decision trees (DT), rough sets, case-based reasoning (CBR), support vector machines (SVM), data envelopment analysis and soft computing. Although these models have been widely used in the last decades, still the pioneer statistical techniques are worth mentioning in the modeling of corporate bankruptcy prediction such as univariate and multivariate discriminant analysis [1], [2]. Classification algorithms like linear discriminant analysis (LDA) and logistic regression (LR) are also popular linear approaches. All these techniques aim at finding an optimal linear combination of explanatory input variables, such as, e.g., solvency and liquidity ratios, in order to analyze, model and predict corporate default risk. Unfortunately the financial ratios violate the assumptions of (i) linear separability, (ii) multivariate normality and (iii) independence of the predictive variables. Therefore, the models overlook the complex nature, boundaries and interrelationships of the financial ratios.

Most of the prediction models use financial ratios as predictor variables, by employing the selection of only a few financial ratios according to a choice based criteria. Model selection of corporate distress prediction is advisable for reducing problem complexity saving computational costs. In [9] a linear pre-processing stage using principal component analysis (PCA) for dimensionality reduction purposes is tested. However, nonlinear projection methods (e.g. ISOMAP) have been successfully used [14] making them more suitable for this problem. With the same goal, non-negative matrix factorization (NMF) is used in [12] for extracting the most discriminative features.

While the forecast of bankruptcy is of paramount importance to all stakeholders, to estimate the probability of a corporate failure can prevent the adverse effects that such event can provoke. In [13] probabilistic framework for bankruptcy detection based on a Relevance Vector Machine (RVM) is described. It is shown therein that the classifier can yield a decision function that is much sparser (than the SVM) leading to significant reduction in the computational complexity while the prediction accuracy is competitive. In [10] a Gaussian Process is used to estimate bankruptcy probabilities.

In [18] a comprehensive review of hybrid and ensemble-based soft computing techniques applied to bankruptcy prediction is presented. Moreover, a variety of soft computing techniques applied to bankruptcy prediction have been referred. Despite the numerous papers dealing with the problem, it is often difficult to compare the techniques due to possible differences in assumptions, data sets, time periods and failure definitions.

In this paper we look at a new learning paradigm [17], [16] SVM+ and propose a financial distress prediction model using priveleged information regarding heterogeneous financial ratios grouped by the type of firms according to the number of employees and annual turnover or global balance. In this regard the approach takes an holistic view of the overall process enhancing the learning inductive process by improving generalization.

## III. Learning Models with SVM+

In [17], [16] Vapnik discusses in detail the extension of the new formulation of the SVM algorithm presented formerly in [15] and demonstrates its superiority toward other machine learning techniques. The new paradigm SVM+ while upholding the main principles of SVM extends its concept, by incorporating the essence of 'untold' information often not handled in a learning problem. In the scope of many practical problems, it is showed that in terms of the capability of transmitting priveleged or hidden information, the role of a supervisor (or even of an oracle) leverages the machine in classification (or regression) tasks. It is a new step in machine learning paradigms which had never been put before. In [8] a learning paradigm for multi-task learning (MTL) is able to solve a problem with heterogeneous data and lateral information. The authors compare in several papers [8], [4] SVM+ and MTL approaches demonstrating their similarities and differences.

### A. Support Vector Machines (SVM)

Support Vector Machines (SVMs) are maximum margin classifiers with low capacity and good generalization. The SVM trains a classifier by finding an optimal separating hyperplane which maximizes the margin between two classes of data in the kernel induced feature space.

Suppose we are given $l$ instances of training data. Each instance consists of a $(\mathbf{x}_i, y_i)$ pair where $\mathbf{x}_i \in \mathbb{R}^N$ is a vector containing $N$ attributes of the instance $i$, and $y_i \in \{+1, -1\}$ is the correspondent class label. The method uses input-output training pairs from the data set $\mathcal{D} = \{(\mathbf{x}_i, y_i) \in \mathcal{X} \subseteq \mathbb{R}^N \times \mathcal{Y} : 1 \leq i \leq l\}$ such that the SVM classifies correctly unobserved data $(\mathbf{x}, y)$.

Each $\mathbf{x}$ in $\mathcal{X}$ is then mapped to a $\phi(\mathbf{x})$ in the kernel-induced feature space, which is related to the kernel function $K$ by $\phi(x)^T\phi(\mathbf{x}') = K(\mathbf{x}, \mathbf{x}')$ for any $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$. SVM tries to find the optimal separating hyperplane $\mathbf{w}^T\phi(\mathbf{x}) + b$ that has large margin and small training error.

The quadratic programming optimization problem originally proposed in [5] is:

$$\min_{\mathbf{w}, b \in \mathbb{R}, \xi} \Phi(\mathbf{w}, b, \xi) = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{l}\sum_{i=1}^{l}\xi_i \qquad (1)$$

subject to constraints

$$\begin{aligned} y_i(\mathbf{w}^T\phi(\mathbf{x}) + b) &\geq 1 - \xi_i \quad & i = 1, \cdots, l \\ \xi_i &\geq 0 & i = 1, \cdots, l \end{aligned} \qquad (2)$$

Here $\xi = [\xi_1, \cdots, \mathbf{x}_l]^T$ is the vector of slack variables upholding the errors, and $C$ is a user-defined regularization parameter that trades-off the margin error.

The problem is solved by maximizing the equivalent form (3) with Lagrange multipliers:

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (3)$$

with respect to $\alpha_i$, under the constraints where $0 \leq \alpha_i \leq \frac{C}{l}$, $i = 1...,l$ and $\sum_{i=1}^{l} \alpha_i y_i = 0$. The solution (4) is a linear combination of the input data points $\mathbf{x}_i$ for which $\alpha_i$ is different zero (the so-called the support vectors (SVs)) and is given by:

$$f(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b \quad (4)$$

with $\alpha_i, b \quad \mathbb{R}$. The SVM finds the class of a given test point $\mathbf{x}_j$ by computing $f(\mathbf{x}_j)$ and by checking which side of the hyperplane it falls on.

### *B. SVM+*

Recently, a generalization of a support vector machine (SVM) technique, called support vector machine plus (SVM+) , was proposed by Vapnik and co-workers in [15].

Suppose that training data are the union of $t > 1$ groups. Let us denote the indices of samples from group $r$ by $T_r = \; i_{nl}, \cdots, i_{nr} \;\; , r = 1, \cdots t$ . Then the total training data set is a union of $t$ groups:

$\mathcal{D}_r = \left\{ (\mathbf{x}_{r_i}, y_{r_i}) \quad \mathcal{X} \subseteq \mathbb{R}^N \times \mathcal{Y} : n_1 \leq i \leq n_r \right\}$

To account for the group information, Vapnik [15] proposed to define the slack variables within each group by so-called 'correcting function'

$$\xi_i = \xi_r(\mathbf{x}_i) = \phi_r(\mathbf{x}_i, \mathbf{w}_r), i \quad T_r, r = 1, \cdots, \mathbf{t}.$$

To define the correcting function $\xi_i = \xi_r(\mathbf{x}_i) = \phi_r(\mathbf{x}_i, \mathbf{w}_r)$ for group $T_r$ Vapnik [15] proposed to map the input training vectors $\mathbf{x}_i, i \quad T_r$ onto two different Hilbert spaces (i) the decision function space and (ii) the space of the correcting functions for a given group $r$. In SVM+ the slack variables are restricted by the correcting functions, and the correcting functions represent additional information about the data. Vapnik uses this concept to control the learning machine by establishing a privileged information setting which leads to an holistic view of the whole process.

The SVM+ approach is designed to take advantage of the structure in the training data (for example, noise present in data, or invariants in the data) . By leveraging this structure, the SVM+ technique can attain better generalization by lowering the overall system's VC-dimension.

$$\min_{\mathbf{w}, \mathbf{w}_r, b \in \mathbb{R}, \xi^r} \Phi(\mathbf{w}, b, \xi) = \frac{1}{2} \quad \mathbf{w} \quad ^2 + \frac{\gamma}{2} \quad \mathbf{w}_r \quad ^2 + \frac{C}{l} \sum_{i=1}^{l} \sum_{i=1}^{l} \xi_i^r \quad (5)$$

subject to:

$$y_i \left( (\mathbf{w} \cdot \mathbf{z}_i) + b \right) \geq 1 - \xi_i^r, i \quad T_r, r = 1 \cdots t \quad (6)$$

The capacity of a set of decision functions is reflected by $\mathbf{w}$ and the capacity of a set of correcting functions for group $r$ is $\mathbf{w}_r$ . SVM+ directly controls the capacity of the decision functions and the correcting function. $\gamma$ adjusts the relative weight of these two capacities. $C$ controls the trade-off between complexity and proportion of nonseparable samples. Figure 1 illustrates the SVM+ learning procedure.

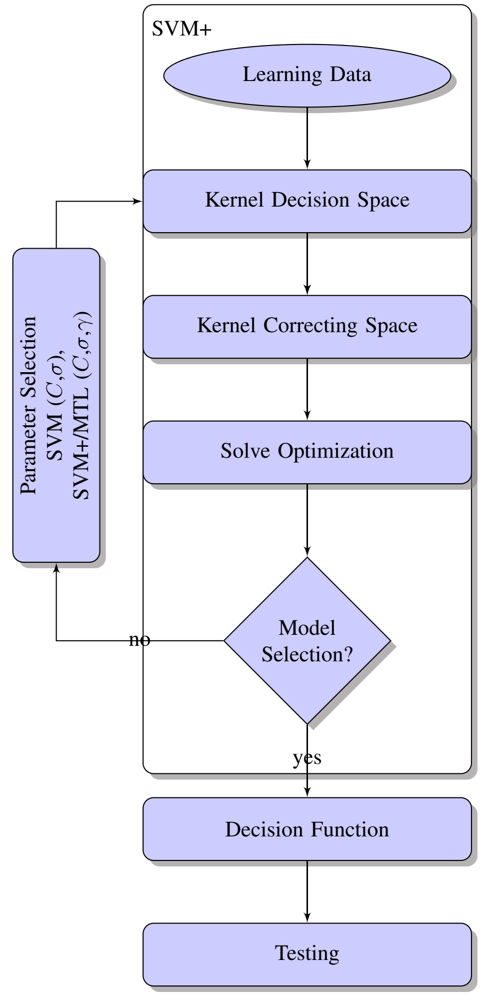


Fig. 1. SVM+ Learning Diagram

In this problem, the slack variables are represented as $(\phi(\mathbf{x}_i), \mathbf{w}_r) + d$ and must be non-negative

$$f(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b + \frac{1}{\gamma} + \sum_{i=1}^{l} \alpha_i K(\mathbf{x}, \mathbf{x}_{i^r}) + d_r \quad (7)$$

with $r = 1 \cdots T$.

From a practical point of view the SVM classifier uses as free parameters the parameter C (in the case the linear SVM is used), and 2 parameters $C$, $\sigma$ (RBF kernel is used). The SVM+ classifier, where linear kernel is used for the decision space, and RBF kernel is used for correcting space, requires 3 parameters: C (as in standard linear SVM), $\gamma$ and $\sigma$ (RBF witdh). In the case RBF kernel is used for the decision space, the SVM+ classifier needs 4 parameters ($C, \gamma, \sigma_1$ and $\sigma_2$).

TABLE I
DIANE DATA BASE FINANCIAL RATIOS

| | Variable Description | | | |
|---|---|---|---|---|
| DIANA DATA BASE | $x_1$ - | Number of Employees Previous year | $x_{16}$ - | Cashflow / Turnover |
| | $x_2$ - | Capital Employed / Fixed Assets | $x_{17}$ - | Working Capital / Turnover days |
| | $x_3$ - | Financial Debt / Capital Employed | $x_{18}$ - | Net Current Assets/Turnover days |
| | $x_4$ - | Depreciation of Tangible Assets | $x_{19}$ - | Working Capital Needs / Turnover |
| | $x_5$ - | Working Capital / Current Assets | $x_{20}$ - | Export |
| | $x_6$ - | Current ratio | $x_{21}$ - | Added Value per Employee in k Euros |
| | $x_7$ - | Liquidity Ratio | $x_{22}$ - | Total Assets Turnover |
| | $x_8$ - | Stock Turnover days | $x_{23}$ - | Operating Profit Margin |
| | $x_9$ - | Collection Period days | $x_{24}$ - | Net Profit Margin |
| | $x_{10}$ - | Credit Period days | $x_{25}$ - | Added Value Margin |
| | $x_{11}$ - | Turnover per Employee k Euros | $x_{26}$ - | Part of Employees |
| | $x_{12}$ - | Interest / Turnover | $x_{27}$ - | Return on Capital Employed |
| | $x_{13}$ - | Debt Period days | $x_{28}$ - | Return on Total Assets |
| | $x_{14}$ - | Financial Debt / Equity | $x_{29}$ - | EBIT Margin |
| | $x_{15}$ - | Financial Debt / Cashflow | $x_{30}$ - | EBITDA Margin |

## IV. EXPERIMENTAL SETUP AND DISCUSSION

### A. Data Description

We used Diane database which contains financial statements of French companies. The initial sample consisted of financial ratios of about 60 000 industrial French companies (for the years of 2002 to 2006) with at least 10 employees. From these companies, about 3000 were declared bankrupted in 2007 or presented a restructuring plan to the court for approval by the creditors.

The 30 financial ratios produced by COFACE are described in Table I. These financial predictors allow to describe firms in terms of the financial strength, liquidity, solvability, productivity of labor and capital, margins, net profitability and return on investment. Although, in the context of linear statistical models, some of these variables have small discriminatory capabilities for default prediction, non-linear approaches may extract relevant information contained in these ratios to improve the classification accuracy without compromising generalization. The ultimate goal is class (healthy, distress) prediction.

The data set is quite mixed in terms of enterprises sectors (construction firms, real estate firms, manufacturing, IT firms, etc.), size of the company (Nr Employees $< 10$, $>= 10$ and $< 100$, $>= 100$ and $<= 6000$), and includes structured and heterogeneous information regarding the companies financial statuses. For a more accurate prediction group information should be included in the model in order to attain better predictability.

The following strategy is pursued from original financial database to appropriately set up a prediction model.

- A set of 600 default companies are selected with at most 10 missing data;
- A set of 600 non-default companies is sampled randomly to obtain a balanced data set;
- The missing values are replaced by the value of the closest available year;
- The ratios are preprocessed by logarithmized operation to decrease the scatter of data distribution;

$$y = \begin{cases} log(x+1) & \text{if } x > 0 \\ -log(1-x) & \text{otherwise} \end{cases} \quad (8)$$

- The features are then normalized for the purpose of equal influence on classification. We use the linear normalization which transforms the maximum value to 1 and the minimum value to 0.

$$y = \frac{x - min(x)}{max(x) - min(x)} \quad (9)$$

- The companies are grouped by their category of large, medium and small regarding their size according to the number of employees and annual turnover or global balance.

### B. Evaluation metrics

Performance metrics were evaluated based on the classification contingency matrix defined in Table III. Here $tp, fp, tn, fn$ represent the usual notation for the confusion matrix in terms of true (or false) and positive (or negative) results from the classifier. Also important are the Recall ($\frac{tp}{tp+fn}$) and Precision ($\frac{tp}{tp+fp}$) measures which are good indicators of the classifier performance.

In this classification problem, two types of misclassification carry different weights. This is due to the fact that a potentially distressed ('bad') company is classified as financially healthy ('good') then the amount of loss incurred by a stakeholder is entirely different from the other type of misclassification. Therefore, the two possible types of errors have to be accounted for. A 'Type I error' (or false positive rate ($fpr$), i.e. $\frac{fp}{fp+tn}$) indicates the misclassification of a healthy firm as distressed. Conversely, a "Type II

TABLE II

CLASSIFICATION RESULTS ON DIANE DATA.

| Class | SVM | | | SVM+ | | | MTL | | |
|---|---|---|---|---|---|---|---|---|---|
| Metrics | F-score | Type I | Type II | F-score | Type I | Type II | F-score | Type I | Type II |
| | 87.55 | 3.82 | 19.26 | **91.03** | 12.97 | **5.92** | 88.00 | 3.82 | 18.5 |
| Metrics | Recall | Precision | Accuracy | Recall | Precision | Accuracy | Recall | Precision | Accuracy |
| | 95.61 | 80.74 | 88.35 | 88.19 | 94.07 | **90.60** | 88.19 | 94.07 | 88.72 |

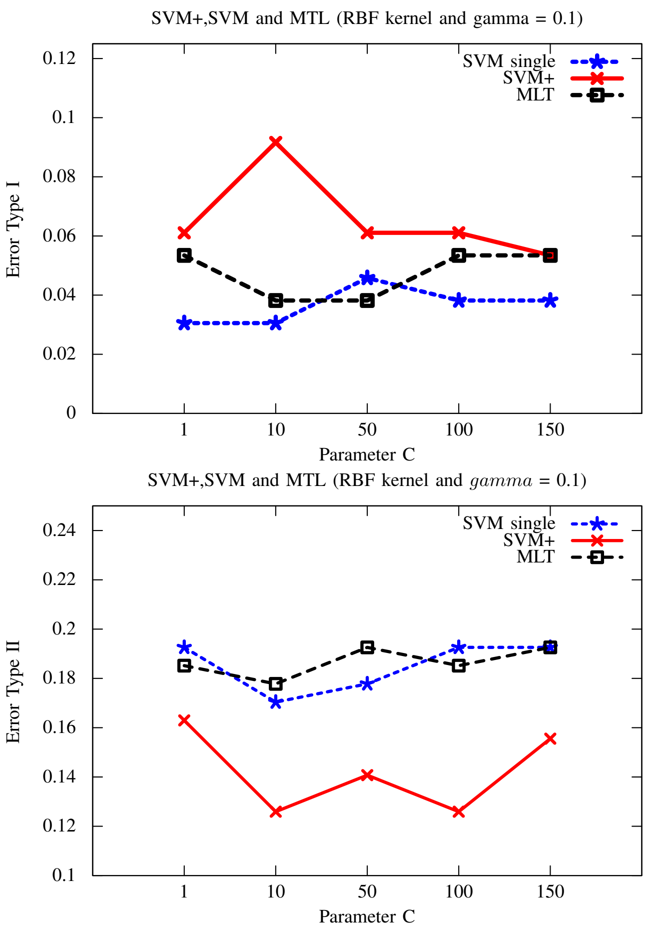


Fig. 2. Errors Type I and Type II

error' (or false negative rate ($fnr$),i.e. $\frac{fn}{fn+tp}$) is one in which a distressed firm is misclassified by the predictor as viable. This error is very important since the predictor should not make a mistake preventing the decision maker to take a wrong decision. Another performance metric is ROC (Receiver Operating Characteristic) curve [6] which is obtained by plotting $tpr$ versus $fpr$. The curve depicts the tradeoffs between $tp$ and $fp$.

TABLE III

CONTINGENCY MATRIX.

| real class | predicted class Bankrupt | Healthy | total |
|---|---|---|---|
| Bankrupt | $tp$ | $fn$ | pos |
| Healthy | $fp$ | $tn$ | neg |
| total | $pos_p$ | $neg_p$ | T |

An "overall hit" refers to the total correct classifications for the set ($\frac{tp+tn}{tp+fp+fn+tn}$) regardless of type. We also illustrate the results with F1-score which quantifies the tradeoff between Recall (R) and Precision (P) and is fairly indicative of the performance of the overall algorithm ($F1 = 2\frac{P*R}{P+R}$). All the results represent mean values obtained in test financial data.

### C. Results

The entire data set is divided randomly into five folds for cross-validation, in which $4$ folds are used for model training, and the remaining is used for testing the generalization capability of the built model. In each trial the SVM, SVM+ and MTL are applied to the learning dataset. For validation, each sample of the test data set is input to the resultant model and the predicted class is assigned as the predicted label. After the experiment is repeated $5$ times, the confusion matrix is calculated by comparing the real class and predicted class for the entire data. Then the evaluation criteria are obtained from the confusion matrix.

In all the experiments we decided to use RBF kernel since it was shown to be the best in previous empirical results running in the same data set [14], [12]. As for the correcting kernel space, although we run some experiments with the linear kernel, we decided to use also the RBF kernel due to better results.

Analysing Table II we observe that F1-score increased by $4\%$ for the SVM+ w.r.t. the baseline SVM, while in MTL approach the F1-score performance improved by $3\%$ w.r.t. the SVM baseline. We observe in the middle column of Table II that the significant measures F1-score ($91.03\%$), predictability accuracy ($90.60\%$) and Error Type II ($5.92\%$) (in bold) are better than the same measures in the baseline (SVM) and MTL approach. Since the misclassification cost on 'bankrupt' class is higher the classifier achieving less error type II is preferred in practice. In Figure 2 a comparison of both type of errors is given, for the SVM+ and SVM single approach as well as MTL showing the former is better.

A study fixing both parameters of the kernel decision space $\sigma_1$ and of the kernel correcting space $\sigma_2$, and parameter $\gamma$, with varying $C$, shows that $F1$-score for the three methods as indicated in the Figure 3 (upper part) is higher for SVM+, while MTL performs better than single SVM (baseline). In a similar way, the decision (and correcting) space kernel free parameters are kept constant as well as the trade-off

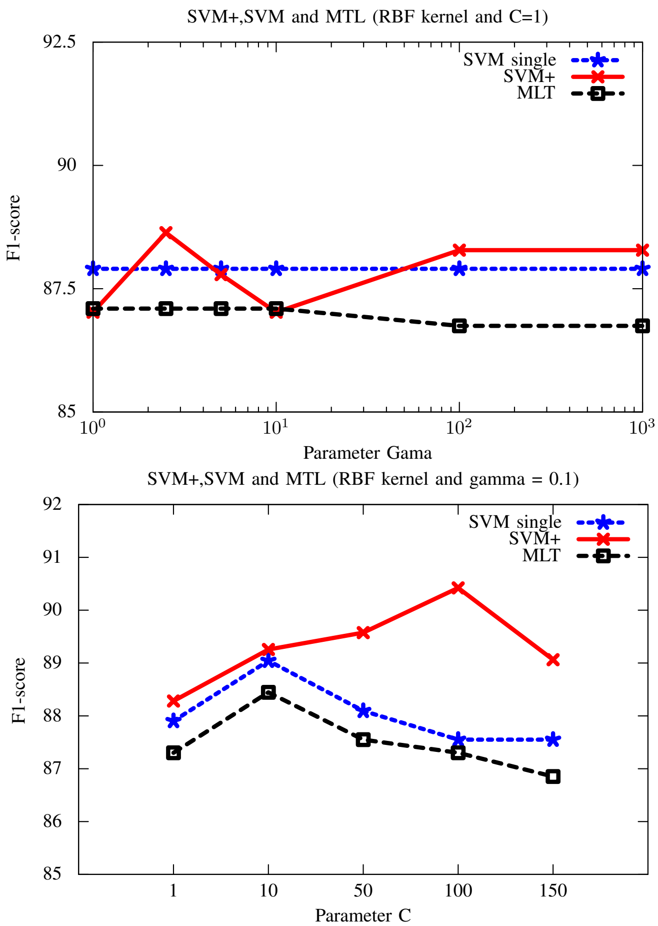


Fig. 3. F1-score versus Parameter $\gamma$ and $C$

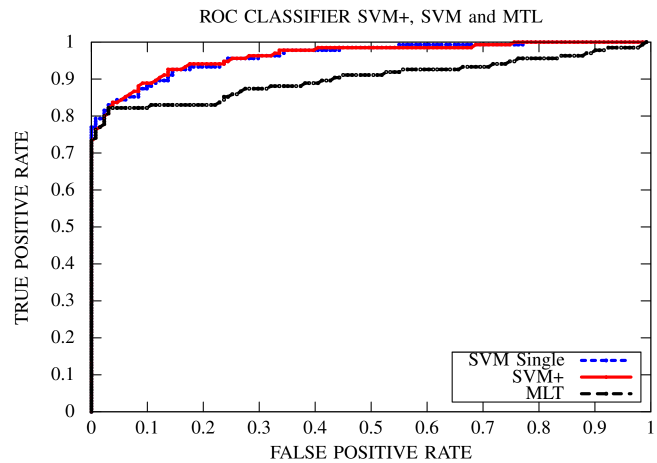


Fig. 4. ROC Curves

parameter $C$, while varying $\gamma$ in the interval $(1 - 1000)$[1]. Figure 3 (lower part) plots the values of $F1$-score under previous conditions, showing also better performance of SVM+ as compared to the other methods. An overall view of the binary classifier performance is observed in Figure 4 which depicts the ROC curves demonstrating the high performance of SVM+.

## V. Conclusion and Future Work

In response to the recent growth of the credit industry and to the world economic crisis early planning for declaring bankruptcy is of great importance to various stakeholders. In this study we use 30 financial ratios as inputs to the failure corporate prediction model using structured and heterogeneous information grouped by the companies financial statuses. The companies are grouped by their category of large, medium, small regarding their size according to the number of employees and annual turnover or global balance. By taking an holistic perspective it was possible to incorporate companies privileged information into the model. As a consequence, different optimized parameters both in the kernel decision space and kernel correcting space are selected, resulting in better overall-predictability performance. The SVM+ model yields improvement of F1-score performance measure while decreasing type II error which quantifies the cost of missing a company in a bad status. Future work will extend this study.

[1]Logscale is used for better visualization. Notice also in this Figure that F1-score is constant for the baseline SVM since its formulation does not include $\gamma$.

## Acknowledgment

This work was partially supported by “Fundação da Ciência e Tecnologia’ under grant no. PTDC/GES/70168/2006.

## References

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

[2] E. I. Altman. *Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting and Avoiding Distress and Pro ting from Bankruptcy*. John Wiley & Sons, 2nd edition, 1993.

[3] A.F. Atiya. Bankruptcy prediction for credit risk using neural networks: A survey and new results. *Neural Networks, IEEE Transactions on*, 12(4):929 – 935, Jul 2001.

[4] Feng Cai and Vladimir Cherkassky. SVM+ regression and multi-task learning. In *Proceedings of International Joint Conference on Neural Networks*, Atlanta, Georgia, 2009.

[5] Corinna Cortes and Vladimir Vapnik. Support-vector networks. In *Machine Learning*, pages 273–297, 1995.

[6] T. E. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, (27):861–874, 2006.

[7] Huang Fu-yuan. A genetic fuzzy neural network for bankruptcy prediction in chinese corporations. In *International Conference on Risk Management & Engineering Management (ICRMEM 08)*, pages 542 – 546, Nov. 2008.

[8] L. Liang and V. Cherkassky. Connection between SVM+ and multitask learning. In *Proc of the IJCNN*, Hong-Kong, 2008.

[9] G.A. Rekba Pai, R. Annapoorani, and G.A. Vijayalakshmi Pai. Performance analysis of a statistical and an evolutionary neural network based classifier for the prediction of industrial bankruptcy. In *IEEE Conference on Cybernetics and Intelligent Systems*, volume 2, pages 1033 – 1038, 2004.

[10] Tonatiuh Pena, Serafin Martinez, and Bolanle Abudu. Bankruptcy prediction: A comparison of some statistical and machine learning techniques. Technical Report TR-Nº 2009-18, 2009.

[11] P. Ravi Kumar and V. Ravi. Bankruptcy prediction in banks and firms via statistical and intelligent techniques - a review. *European Journal of Operational Research*, 180(1):1 – 28, July 2007.

[12] B Ribeiro, C Silva, A Vieira, and J Carvalho das Neves. Extracting discriminative features using non-negative matrix factorization in financial distress data. In M. Kolehmainen et al. (Eds.), editor, *Int Conf on Adaptive and Natural Computing Algorithms*, volume 4432, pages 537–547, Berlin Heidelberg, April 2009. Lecture Notes in Computer Science (LNCS), Springer-Verlag.

[13] B. Ribeiro, A. Vieira, and J.C. das Neves. Sparse bayesian models: Bankruptcy-predictors of choice? In *Neural Networks, 2006. IJCNN 06. International Joint Conference on*, pages 3377–3381, July 2006.

[14] B Ribeiro, A Vieira, J Duarte, C Silva, J C das Neves, Q Liu, and A H. Sung. Learning manifolds for bankruptcy analysis. In M. Köppen et al. (Eds.), editor, *Int. Conf. on Neural Information Processing*, volume 5506, pages 722–729, Berlin Heidelberg, 2009. Lecture Notes in Computer Science (LNCS), Springer-Verlag.

[15] V. Vapnik. *Estimation of Dependences Based on Empirical Data: Empirical Inference Science*. Springer-Verlag New York, Inc., 1982–2006.

[16] V. Vapnik, A. Vashist, and N. Pavlovitch. Learning using hidden information (learning with teacher). In *Proceedings of International Joint Conference on Neural Networks*, pages 3188–3195, 2009.

[17] Vladimir Vapnik and Akshay Vashist. A new learning paradigm: Learning using privileged information. *Neural Networks*, 22(5-6):544–557, July 2009.

[18] A. Verikas, Z. Kalsyte, M. Bacauskiene, and A. Gelzinis. Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: A survey. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, 2009.

[19] Wei Wen Wu. Beyond business failure prediction. *Expert Systems with Applications*, (37):2371–2376, 2010.