Available online at www.sciencedirect.com

ScienceDirect

Expert Systems with Applications 36 (2009) 3302–3308

Expert Systems with Applications

www.elsevier.com/locate/eswa

Support vector machines for credit scoring and discovery of significant features

Tony Bellotti [*], Jonathan Crook

Credit Research Centre, Management School and Economics, University of Edinburgh, William Robertson Building, 50 George Square, Edinburgh EH8 9JY, UK

Abstract

The assessment of risk of default on credit is important for financial institutions. Logistic regression and discriminant analysis are techniques traditionally used in credit scoring for determining likelihood to default based on consumer application and credit reference agency data. We test support vector machines against these traditional methods on a large credit card database. We find that they are competitive and can be used as the basis of a feature selection method to discover those features that are most significant in determining risk of default.

Keywords: SVM; Credit scoring; Feature selection

1. Introduction

Credit scoring is the set of decision models and techniques that aid lenders in granting consumer credit by assessing the risk of lending to different consumers. It is an important area of research that enables financial institutions to develop lending strategies to optimize profit. Additionally, bad debt is a growing social problem that could be tackled partly by better informed lending enabled by more accurate credit scoring models. A range of different data mining and statistical techniques have been used since the 1930s when numerical score cards were first introduced by mail-order companies (Thomas, Edelman, & Crook, 2002, Section 1.3). It is now common for financial institutions to use statistical methods such as logistic regression (LR) and linear discriminant analysis (LDA) to build credit scoring models. Potential borrowers are classified according to their probability to default on a loan, based on application and credit reference agency data collected about them. Such models are used by setting a threshold on the probability to default and rejecting loan applications that fall below this level.

Support vector machines (SVMs) have been applied successfully in many classification problems such as text categorization, image recognition and gene expression analysis (e.g. see Cristianini & Shawe-Taylor, 2000, Chapter 8). Experiments using SVM for credit scoring are relatively new, however. Several papers have recently been published assessing the performance of SVM for credit scoring. Baesens et al. (2003) apply SVMs, along with other classifiers to several data sets. They report that SVMs perform well in comparison with other algorithms, but do not always give the best performance. Schebesch and Stecking (2005) apply SVM to a database of applicants for building and loan credit. They conclude that SVMs perform slightly better than LR, but not significantly so. They also use SVMs, with its capacity to output support vectors, to discover typical and critical regions of the problem space. Both papers report using linear SVM and a Gaussian radial basis function (RBF) kernel. In both cases, the size of the credit database is much smaller than would typically be used in a real application. Van Gestel et al. (2006) use least squares

[*] Corresponding author.
E-mail address: tony.bellotti@ed.ac.uk (T. Bellotti).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.01.005

SVMs with a Bayesian kernel to derive classifiers for corporate bankruptcy. They find no significant difference between SVM, LR and LDA in terms of proportion of test cases correctly classified and no difference between LR and SVM in terms of area under the ROC curve. Li, Shiue, and Huang (2006) find SVMs outperform multi-layer perceptrons for consumer credit data, but their results are also based on a small sample size. Huang, Chen, Hsu, Chen, and Wu (2004) compare SVMs with a back-propagation neural network to predict corporate credit ratings but find inconsequential differences in performance. Lee (2007) find a similar result for corporate loans. Dikken (2005) finds SVMs to be inferior to LR when modelling corporate credit risk. Huang, Chen, and Wang (2007) find SVMs classify credit applications no more accurately than artificial neural nets (ANN), decision trees or genetic algorithms (GA), and compared the relative importance of using features selected by GA and SVM along with ANN and genetic programming. However, they use data sets far smaller and with fewer features than would be used by a financial institution and do not compare the features selected by SVM alone, nor do they compare with methods used in practice such as LR.

In this paper, our general framework is to compare the performance of SVM against several other well-known algorithms: LR, LDA and *k*-nearest neighbours (*k*NN). We extend the work on assessing SVM for credit scoring in several ways.

1. SVM is tested against a much larger database of credit card customers than has been considered in the literature so far. We restrict our attention to those accounts opened in the same three month period. Hand (2006) points out that for many classification problems, the data suffers from population drift, in that the class distributions shift over time. This is particularly true of credit data with customer behaviour changing over time due to economic circumstances or changes in product development and marketing. For this reason a clearer model can be developed if it is based on data taken from a narrow time period within which there is likely to be less variability in these circumstances.
2. SVM is tested with a polynomial kernel to determine if a non-linear polynomial decision space yields better performance than linear SVM or using the Gaussian RBF kernel.
3. SVM performance is assessed in light of the number of support vectors required to model the data.
4. Financial institutions are primarily interested in determining which consumers are most likely to default on loans. However, they are also interested in knowing which characteristics of a consumer are most likely to affect their likelihood to default. For example, is a home-owner less likely to default than a tenant? This information allows credit modellers to stress test their predictions. Traditionally a test of significance of features is used to discover these characteristics. When a LR problem is solved using maximum likelihood estimation, a Wald statistic can be computed for each feature which is then used to determine significance. As an alternative, we follow Guyon, Weston, Barnhill, and Vapnik (2002) who select significant features in the data using the square of weights on features output by SVM. We apply this approach to select the top ranking features that are significant for credit scoring. This selection is compared with that given using LR.

2. Data

A data set of approximately 25,000 customers with credit cards opened in the same three month period of 2004 was provided by a major financial institution. Four different credit card products are represented in the data. A customer is defined to have *defaulted* if he or she has become three months or more behind in their payments within the first 12 months after the account is opened. Other definitions of default can be used, but this one is common in credit scoring (Thomas et al., 2002, Section 8.3). Defaulting customers are referred to as *bad* cases and all others as *good* cases. The data includes 34 features taken from each customer's original application along with features extracted from a credit reference agency at the time of application. The data is standardized before use, so each feature has the same mean (0) and variance (1).

Typically, credit data is not easily separable by any decision surface. This is natural since the data at time of application cannot capture the complexities in each individual customer's life that may lead to default. The application data can at best only provide an indication of default. Consequently, it is usual for the rates of misclassification on credit data to be between around 20% and 30% (e.g. see Baesens et al., 2003). This would be considered a poor result for many other classification problems but is typical of credit data. The poor separability of the credit data is illustrated in Fig. 1. The good cases tend to cluster towards the bottom-right and the bad towards the top-left, but this is only a very general trend and there is no clear separation.

3. Methods

The SVM is a relatively new learning algorithm that can be used for classification. We compare its performance against three older statistical classification methods: LR, LDA and *k*NN. All algorithms are described briefly below for a sequence of n training examples $(\mathbf{x}_1, y_1, \ldots, \mathbf{x}_n, y_n)$ with feature vectors $\mathbf{x}_i$ and class labels y_i. For credit scoring, the class label is either *bad* or *good*.

3.1. Support vector machine (SVM) classifier

SVM separates binary classified data by a hyperplane such that the margin width between the hyperplane and the examples is maximized. Statistical learning theory shows that maximizing the margin width reduces the