Combining Support Vector Machines for Credit Scoring

Ralf Stecking and Klaus B. Schebesch

Department of Economics, University of Bremen, Germany
stecking@uni-bremen.de
kbsbase@gmx.de

Summary. Support vector machines (SVM) from statistical learning theory are powerful classification methods with a wide range of applications including credit scoring. The urgent need to further boost classification performance in many applications leads the machine learning community into developing SVM with multiple kernels and many other combined approaches. Owing to the huge size of the credit market, even small improvements in classification accuracy might considerably reduce effective misclassification costs experienced by banks. Under certain conditions, the combination of different models may reduce or at least stabilize the risk of misclassification. We report on combining several SVM with different kernel functions and variable credit client data sets. We present classification results produced by various combination strategies and we compare them to the results obtained earlier with more traditional single SVM credit scoring models.

1 Introduction

Classifier combination is a recurring topic since classification models are successfully applied across disciplines ranging from pattern recognition in astronomy to automatic credit client scoring in finance. Classifier combination is done in the hope of improving the out-of-sample classification performance of single *base* classifiers. As is well known by now ([1],[2]), the results of such combiners can be both better or worse than expensively crafted single models. In general, as the base models are less powerful (and much more easy to produce) their combiners tend to yield much better results. However, this advantage is decreasing with the quality of the base models (e.g. [1]). Our past credit scoring single-SVM classifiers concentrate on misclassification performance obtainable by different SVM kernels, different input variable subsets and financial operating characteristics ([3],[4],[5],[6]). In credit scoring, classifier combination using such base models may be very useful indeed, as small improvements in classification accuracy matter and as *fusing* models on different inputs may be required by practice. Hence, the paper presents in sections 2 and 3 model combinations with base models on all available inputs using single classifiers with six different kernels, and finally in section 4 SVM model combinations of base models on reduced inputs using the same kernel classifier.

Table 1. Region error w.r.t. figure 1 below. For critical/critical region `V` (where $|f_k(x)| < 1$ for both models) classification error is highest. For censored combination models these input region would not be predictable.

Region	`II`	`III`	`IV`	`V`	`VI`	`VII`	`VIII`	**Total**
Error in %	15.0	12.2	22.9	39.7	22.9	9.1	11.1	28.0
N	20	82	35	360	35	99	27	658

2 Censored Combination of SVM

The basic data set for our past credit scoring models is an equally distributed sample of 658 clients for a building and loan credit with a total number of 40 input variables. In order to forecast the defaulting behavior of new credit clients, a variety of SVM models were constructed, in part also for addressing special circumstances like asymmetric costs of misclassification and unbalanced class sizes in the credit client population [3]. Using a **fixed** set of 40 **input variables**, six different SVM with **varying kernel functions** are combined for classifying good and bad credit clients. Detailed information about kernels, hyperparameters and tuning can be found in [6].

Our combined approach of different SVM will be described as follows: Let $f_k(x) = \langle \Phi_k(x), w_k \rangle + b_k$ be the output of the kth SVM model for unknown pattern x, with b_k a constant, Φ_k the (usually unknown) feature map which lifts points from the input space $\mathbf{X}$ into feature space $\mathcal{F}$, hence $\Phi : \mathbf{X} \to \mathcal{F}$. The weight vector w_k is defined by $w_k = \sum_i \alpha_i y_i \Phi_k(x_i)$ with α_i the (C bounded) dual variables ($0 \leq \alpha_i \leq C$), and y_i be the binary output of input pattern x. Note also, that $\langle \Phi_k(x), \Phi_k(x_i) \rangle = K(x, x_i)$, where K is a kernel function, for example $K(x,x_i) = \exp\left(-s\|x - x_i\|^2\right)$, i.e. the well known RBF kernel with user specified kernel parameter s. In previous work [4] it was shown that SVM **output regions** can be defined in the following way: (1) if $|f_k(x)| \geq 1$, then x is called a *typical* pattern with low classification error, (2) if $|f_k(x)| < 1$, then x is a *critical* pattern with high classification error. Combining SVM models for classification we calculate sign $\left(\sum_k f_k(x)\right)$ introducing the restriction, that $f_k(x) = 0$, if $|f_k(x)| < 1$, which means: SVM model k has zero contribution for its critical patterns. For illustrative purpose we combine two SVM models (RBF and second degree polynomial) and mark nine regions (see figure 1): typical/typical regions are `I, III, VII, IX`, critical/critical region is `V` and typical/critical regions are `II, IV, VI, VIII`. Censored classification uses only typical/typical regions (with a classification error of 10.5 %) and typical/critical regions (where critical predictions are set to zero) with a classification error of 18.8 %. For the critical/critical region `V` no classification is given, as the expected error within this region would be 39.7 % (see table 1). The crucial point for this combination strategy is the fairly high number of unpredictable patterns (360 out of 658) in this case. However, by enhancing the diversity and by increasing the number of SVM models used in combinations, the number of predictable patterns will also increase, as will be shown in the following section.