

ROBUST SUPPORT VECTOR MACHINE

MÁY HỖ TRỢ VECTOR ROBUST

Le Hoang Son, Phan Nguyen Ky Phuc

Department of Industrial and System Engineering, International University (VNU-HCM)

ABSTRACT

Motivated by the fact the breast cancer has risen awareness recently, this paper considers well-known statistical learning technique, Support Vector Machines (SVM), to classify cancer. Moreover, to deal with the uncertainty data, a modified version of SVM, Robust Support Vector Machines (RSVM) is introduced with the application of robust optimization. Thus, the nature of classification model is also changed but still could be solved practically. To illustrate the computational performance, our method was tested on Wisconsin Breast Cancer dataset taken from UCI Machine Learning Repository. The result shows that our RSVM has significant improvement over regularized methods.

Keywords: Robust optimization; Binary Classification; Mixed-Integer Optimization.

TÓM TẮT

Được thúc đẩy bởi thực tế là ung thư vú đã gia tăng nhận thức gần đây, bài báo này xem xét kỹ thuật học thống kê nổi tiếng, Máy hỗ trợ véc tơ (SVM), để phân loại ung thư. Hơn nữa, để đối phó với dữ liệu không chắc chắn, một phiên bản sửa đổi của SVM, Máy vector hỗ trợ mạnh mẽ (RSVM) được giới thiệu với ứng dụng tối ưu hóa mạnh mẽ. Do đó, bản chất của mô hình phân loại cũng được thay đổi nhưng vẫn có thể giải quyết được trên thực tế. Để minh họa hiệu suất tính toán, phương pháp của chúng tôi đã được thử nghiệm trên bộ dữ liệu về Ung thư vú Wisconsin được lấy từ Kho lưu trữ Máy học UCI. Kết quả cho thấy RSVM của chúng tôi đã cải thiện đáng kể so với các phương pháp chính quy.

Từ khóa: Tối ưu Robust; Phân loại nhị phân; Tối ưu nguyên.



1. INTRODUCTION

Support vector machines (SVMs) are a group of related supervised learning methodologies used for classification and regression. Given a training set in which each data points are marked as belonging to two separate classes, an SVM training algorithm builds a model that aims to predicts whether a sample falls into one class or the others. Theoretically, SVMs constructs a hyperplane or a set of hyperplanes in a high or infinite dimensional space, which could be used for classification and regression, etc. Spontaneously, an optimal hyperplane should have the largest margin to the nearest training points of any specified class, since the larger distance usually implies the lower error of the classifier.

Given a set of training data points belonging to two classes, SVMs assume that the data point is given as a p - dimensional vector, and we would like to separate such points with a $p-1$ - dimensional hyperplane. This is so-called a “linear classifier”. As mentioned before, the best solution of hyperplane normally represents the largest separation, or margin. Hence, such hyperplane is chosen which has distance from it to the nearest point on each plane is maximized, so-called the maximum-margin hyperplane. The linear SVMs could be formulates as a quadratic program. The SVMs are important algorithms in data mining and machine learning [1]. However, it is usually that data contain noise, or with uncertainties. For instance, data points could be ranged within some region. Many studies were conducted in this subject, for example, the SVMs with interval uncertainty set in [2-6], the robust models of SVMs with ellipsoidal uncertainty in Bertsimas [7]. Additionally, data noises in some extensions of SVMs have been considered. With the SVM which constructs two

non-parallel hyperplanes and desires to solve two smaller quadratic programs, Twin support vector machines (TWSVMs) is a kind of SVM. Researchers already applied many types of uncertainty sets on TWSVMs. In [9], ellipsoidal uncertainty for TWSVM is solved by second order cone programming method. Mathematically, due to characteristics of uncertainty sets, there is an equivalence between regularized support vector machines and robust optimization for ellipsoidal uncertainty is studied in [11]. To handle the ellipsoidal uncertainty, robust optimization to regularized generalized eigenvalue classification is studied in [12]. Recently, chance-constrained robust optimization models were also used for classification with noisy data [14].

In this paper, we consider the robust optimization technique for SVMs in uncertainties of input data, for both features and labels. Recently, the uncertainty in labels of input data has been considered in Bertsimas. In mathematical programming, the robust optimization is an approach to tackle the uncertainties with unknown probabilities. Moreover, the robust optimization has been studied widely by [15, 16] and [17]. We also propose the equivalent formulation for SVMs with desired uncertainty sets.

Notation: let $\{\mathbf{x}_i, y_i\}_{i=1}^n$ be training set provided for the classification task, where $\mathbf{x}_i \in \mathbb{R}^p$ is the feature vector and $y_i \in \{-1, 1\}$ is the label for observation i .

$$\mathcal{D} = \{(\mathbf{x}_i, y_i) | \mathbf{x}_i \in \mathbb{R}^p, y_i \in \{-1, 1\}\}_{i=1}^n$$

2. METHODOLOGY

2.1. Soft-Margin Support Vector Machines for classification

Soft-Margin support vector machines

(SMSVM) is a variation of classical maximal margin classifier which relax the requirement of linearly separable data. Instead, it allows for points to be incorrectly classified. SMSVMs uses the hinge function as loss function and using parameter C to balance the minimization of total loss and maximization of margin, noted that C could be tuned through validation. Formulation of SMSVMs:

$$\min_{\omega, b} \frac{1}{2} \|\omega\|_2^2 + C \sum_{i=1}^n \max\{1 - y_i(\omega^T x_i - b), 0\} \quad (1)$$

And its equivalence is formulated as following problem:

$$\begin{aligned} \min_{\omega, b} \quad & \frac{1}{2} \|\omega\|_2^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i(\omega^T x_i - b) \geq 1 - \xi_i \quad i = 1, \dots, n \quad (2) \\ & \xi_i \geq 0 \quad i = 1, \dots, n \end{aligned}$$

2.2. Transductive Support Vector Machines with Soft-Margin

An extend variance of SVMs is Transductive SVMs (TSVMs) which use to treat partially labeled data in semi-supervised learning. In [15], [16], the formulation for TSVMs has been studied. Besides the training set D , we need to consider a test data set:

$$D^* = \{x_j^* | x_j^* \in \mathcal{R}^p\}_{j=1}^m$$

Where the point x_j^* is unclassified. The optimization model for TSVMs can be formulated as:

$$\begin{aligned} \min_{\omega, b, y_j^*} \quad & \frac{1}{2} \|\omega\|_2^2 + C \sum_{i=1}^n \xi_i \quad (3) \\ \text{s.t.} \quad & \end{aligned}$$

$$\begin{aligned} y_i(\omega^T x_i - b) &\geq 1 - \xi_i \quad i = 1, \dots, n \\ y_j^*(\omega^T x_j^* - b) &\geq 1 \quad j = 1, \dots, m \\ y_j^* &\in \{-1, 1\} \end{aligned}$$

2.3. Uncertainty and Robust optimization

2.3.1. Uncertainty

In classification tasks, the perturbation of the features and labels could arise during the data collection, or/and preprocessing. Provided that left unaddressed, the trained model could be biased and severely influenced by uncertainties in the data. We introduce the perturbation in labels and given corresponding robust counterpart formulation.

Uncertainties in label: With the approaching by robust optimization, such perturbations are taken into consideration. To address this, we introduce supplement $\Delta y_i \in \{0, 1\}$, where 1 indicates that the labels were incorrect and flipped, and 0 is otherwise. Then, the uncertainty set would be:

$$\mathcal{U}_y = \left\{ \Delta y \in \{0, 1\}^{n \times p} \mid \sum_{i=1}^n \Delta y_i \leq \gamma \right\} \quad (4)$$

We introduce the parameter γ , an integer parameter, to control the total number of data points that we allow to be mislabeled. Then, the labels of training set as $y_i(1 - 2\Delta y_i)$, $i = 1, \dots, n$. In practice, the parameter γ is often modeled as the proportion of total number of data points and could be chosen via validation.

2.3.2. Robust optimization formulation

The robust counterpart of (1) against $\mathcal{U}_y$ is 

$$\min_{\omega, b} \max_{\Delta y \in \mathcal{U}_y} \frac{1}{2} \|\omega\|_2^2 + C \sum_{i=1}^n \max\{1 - y_i(1 - 2\Delta y_i)(\omega^T x_i - b), 0\} \quad (5)$$

The robust formulation of (3) is

$$\min \frac{1}{2} \|\omega\|_2^2 + C \sum_{i=1}^n \xi_i + \gamma q + \sum_{i=1}^n r_i$$

s.t:

$$\begin{aligned} q + r_i &\geq C(\phi_i - \xi_i) & i = 1, \dots, n \\ \xi_i &\geq 1 - y_i(\omega^T x_i - b) & i = 1, \dots, n \\ \xi_i &\leq 1 - y_i(\omega^T x_i - b) + M(1 - s_i) & i = 1, \dots, n \\ \xi_i &\leq Ms_i & i = 1, \dots, n \\ \phi_i &\geq 1 + y_i(\omega^T x_i - b) & i = 1, \dots, n \\ \phi_i &\geq 1 + y_i(\omega^T x_i - b) + M(1 - t_i) & i = 1, \dots, n \\ \phi_i &\leq Mt_i & i = 1, \dots, n \\ r_i, \xi_i, \phi_i &\geq 0 & i = 1, \dots, n \\ q &\geq 0 \\ s, t &\in \{0, 1\}^{n \times p} \end{aligned}$$

The label-robust formulation of (1), which excluded the parameter C, has been shown by [7]. Problem (5) is a mixed-integer optimization problem which is practically solved by solvers.

Robust TSVMs

To obtain a label-robust reformulation for problem (3), we may seek:

$$\min_{w, b, y_j^*} \frac{1}{2} \|\omega\|_2^2 + C \sum_{i=1}^n \max\{1 - y_i(\omega^T x_i - b), 0\}$$

s.t:

$$\begin{aligned} y_j^*(\omega^T x_j^* - b) &\geq 1 & j = 1, \dots, m \\ y_j^* &\in \{-1, 1\} \end{aligned}$$

Then we substitute the label variables y_i ($1 - 2\Delta y_i$), $i = 1, \dots, n$

$$\min_{w, b, y_j^*} \max_{\Delta y \in \mathcal{U}_y} \frac{1}{2} \|\omega\|_2^2 + C \sum_{i=1}^n \max\{1 - y_i(1 - 2\Delta y_i)(\omega^T x_i - b), 0\} \quad (6)$$

s.t:

$$\begin{aligned} y_j^*(\omega^T x_j^* - b) &\geq 1 & j = 1, \dots, m \\ y_j^* &\in \{-1, 1\} \end{aligned}$$

Observe that the objective of problem (6) has the same value with problem (5).

Moreover, we can formulate the constraint $y_j^*(\omega^T x_j^* - b) \geq 1$ to seek a new formulation for (7). Since, y_j^* is unknown variables and $\omega = (\omega_1, \dots, \omega_p)^T$ is continuous, we use inequalities to reconstruct $y_j^* \omega_k$ by using a variable a_{jk} , $k = 1, \dots, p$.

$$\begin{cases} a_{jk} \geq \omega_k + \frac{l}{2}(1 - y_j^*) \\ a_{jk} \geq -\omega_k + \frac{l}{2}(1 + y_j^*) \\ a_{jk} \leq \omega_k + \frac{u}{2}(1 - y_j^*) \\ a_{jk} \leq -\omega_k + \frac{u}{2}(1 + y_j^*) \end{cases}$$

Where the lower bound $\frac{l}{2} < 0$ and the upper bound are the control for ω_k . Next, we denote $a_j = (a_{j1}, \dots, a_{jp})^T$.

Theorem 3: The complete formulation for (6) is followed:

$$\min \frac{1}{2} \|\omega\|_2^2 + C \sum_{i=1}^n \xi_i + \gamma q + \sum_{i=1}^n r_i \quad (7)$$

s.t:

$$\begin{aligned}
 q + r_i &\geq C(\phi_i - \xi_i) & i = 1, \dots, n \\
 \xi_i &\geq 1 - y_i(\omega^T \mathbf{x}_i - b) & i = 1, \dots, n \\
 \xi_i &\leq 1 - y_i(\omega^T \mathbf{x}_i - b) + M(1 - s_i) & i = 1, \dots, n \\
 \xi_i &\leq Ms_i & i = 1, \dots, n \\
 \phi_i &\geq 1 + y_i(\omega^T \mathbf{x}_i - b) & i = 1, \dots, n \\
 \phi_i &\geq 1 + y_i(\omega^T \mathbf{x}_i - b) + M(1 - t_i) & i = 1, \dots, n \\
 \phi_i &\leq Mt_i & i = 1, \dots, n \\
 r_i, \xi_i, \phi_i &\geq 0 & i = 1, \dots, n \\
 (a_j x_j^* - y_j^* b) &\geq 1 & j = 1, \dots, m \\
 a_j &= \begin{cases} a_{jk} \geq \omega_k + \frac{l}{2}(1 - y_j^*) \\ a_{jk} \geq -\omega_k + \frac{l}{2}(1 + y_j^*) \\ a_{jk} \leq \omega_k + \frac{u}{2}(1 - y_j^*) \\ a_{jk} \leq -\omega_k + \frac{u}{2}(1 + y_j^*) \end{cases} \\
 y_j^* &\in \{-1, 1\} \\
 q &\geq 0 \\
 s, t &\in \{0, 1\}^{n \times p}
 \end{aligned}$$

In this formulation, all constraints are linearized. This formulation is mixed-integer optimization problem which can be solved practically. Note that, we could use piecewise-linearized the quadratic terms in the objective in order to completely transform the problem into. For details of these reformulation, we prefer reader to [7].

3. NUMERICAL EXPERIMENT

3.1. Experiment setup

In order to report the performance of the given robust SVMs, we tested the accuracy of RSVMs on the well-known dataset, which is Wisconsin Breast Cancer dataset from UCI Machine Learning Repository. To seek the result, one-versus-rest was considered. The dataset was split into the training set (60%), validation set (20%) and testing set (20%). The training set was used to train the model with variety of input parameters. Next, the misclassification

error was kept track on validation set, hence, used for hyperparameter tuning to seek the best combination of input parameters. Finally, with the best input parameters, the misclassification on testing set was reported. Since, the RSVMs in this report is mixed-integer optimization problem which could be solved by commercial solver practically. Note that, in the objective of RSVMs, there is a quadratic term which can be reformulated by piecewise formulation to transforms the nature of problem to mixed-integer linear optimization problems. There are many specific modelling languages that can be used on this problem. We chose CVXPY, a convex discipline programming language on Python. The experiment was conducted on selflaptop which contained i7-4510U(2.2 GHz), 8GB RAM.

3.2. Computational result

For comparison between SPO+ and LS loss, we generally observe that SPO+ provides superior result to LS, which emphasizes the roles of incorporating optimization information to prediction tasks. As long as the number of assets becomes greater, the SPO+ method performs significantly compare to LS loss. As for the θ values and N scenarios grow up which elevate the complexity of downstream optimization, we observe that SPO+ still show improvement over LS loss. Especially for the case of 50 assets, SPO+ method tends to offer improvement over LS method in all sectors. However, one cannot claim that SPO+ outperforms LS due to the fact that the improvement of SPO+ over LS is not considerable large, whereby most of them are lower than 5%. Finally, SPO+, which combine the information of downstream optimization problems into prediction task, is better approach than LS in purpose of integrating prediction and optimization task. The detailed result is shown as:



Table 1. Computation result

Metric	Support Vector Machines		
	RSVM	SVM	Robust Improvement
Accuracy	98.20%	97.20%	+1%
Speed	6.44s	8s	1.56s

4. CONCLUSIONS

In this paper, we provide the formulation for chance constrained portfolio selection when the uncertainty affects to the decision variables using Wasserstein ambiguity sets, which implies that we handle the unknown parameters in constraints. We also provide the SPO and SPO+ frameworks for developing prediction model under predict-then-optimize paradigm. By using SPO and SPO+, we could include the contextual information to optimization, which usually ignore in practice. However, there is still some drawbacks and possible new research opportunities in our approach. ♦

References:

- [1]. I. Steinwart and A. Christmann; “*Support vector machines*”. Springer Science & Business Media, 2008.
- [2]. J. Bi and T. Zhang; “*Support vector classification with input data uncertainty*”, in *Advances in neural information processing systems*, 2005, pp. 161–168.
- [3]. T. B. Trafalis and R. C. Gilbert; “*Robust classification and regression using support vector machines*”. *Eur. J. Oper. Res.*, vol. 173, no. 3, pp. 893–909, 2006.
- [4]. T. B. Trafalis and R. C. Gilbert; “*Robust support vector machines for classification and computational issues*”. *Optim. Methods Softw.*, vol. 22, no. 1, pp. 187–198, 2007.
- [5]. L. El Ghaoui, G. R. G. Lanckriet, G. Natsoulis, and others; “*Robust classification with interval data*”, 2003.
- [6]. E. Niaf, R. Flamary, C. Lartizien, and S. Canu; “*Handling uncertainties in SVM classification*”, in *2011 IEEE Statistical Signal Processing Workshop (SSP)*, 2011, pp. 757–760.
- [7]. D. Bertsimas, J. Dunn, C. Pawlowski, and Y. D. Zhuo; “*Robust Classification*”. *INFORMS J. Optim.*, no. October, 2018.
- [8]. Z. Qi, Y. Tian, and Y. Shi; “*Robust twin support vector machine for pattern classification*”. *Pattern Recognit.*, vol. 46, no. 1, pp. 305–316, 2013.
- [9]. P. Xanthopoulos, M. R. Guarracino, and P. M. Pardalos; “*Robust generalized eigenvalue classifier with ellipsoidal uncertainty*”. *Ann. Oper. Res.*, vol. 216, no. 1, pp. 327–342, 2014.
- [10]. A. Ben-Tal, S. Bhadra, C. Bhattacharyya, and J. S. Nath; “*Chance constrained uncertain classification via robust optimization*”. *Math. Program.*, vol. 127, no. 1, pp. 145–173, 2011.
- [11]. A. Ben-Tal and -A. Nemirovski; “*Robust solutions of uncertain linear programs*”. *Oper. Res. Lett.*, vol. 25, no. 1, pp. 1–13, 1999.
- [12]. A. Ben-Tal and A. Nemirovski; “*Robust optimization - methodology and applications*”. *Math. Program.*, vol. 92, no. 3, pp. 453–480, 2002.
- [13]. D. Bertsimas and M. Sim; “*Robust discrete optimization and network flows*”. *Math. Program.*, vol. 98, no. 1–3, pp. 49–71, 2003.
- [14]. C. Cortes and V. Vapnik; “*Support-vector networks*”. *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [15]. M. Brunato and R. Battiti; “*Learning and intelligent optimization (LION)*”. *Proc. VLDB Endow.*, vol. 6, no. 11, pp. 1176–1177, 2014.
- [16]. S. Ding; “*An overview on semi-supervised support vector machine*”. *Neural Comput. Appl.*, 2015.