Neural Networks 57 (2014) 29–38

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Using financial risk measures for analyzing generalization performance of machine learning models

Akiko Takeda [a,*], Takafumi Kanamori [b]

[a] Department of Mathematical Informatics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

[b] Department of Computer Science and Mathematical Informatics, Nagoya University, Furocho, Chikusa-ku, Nagoya-shi, Aichi 464-8603, Japan

ARTICLE INFO

Article history:
Received 26 July 2013
Received in revised form 11 February 2014
Accepted 12 May 2014
Available online 27 May 2014

Keywords:
Support vector machine
Minimax probability machine
Financial risk measure
Generalization performance

ABSTRACT

We propose a unified machine learning model (UMLM) for two-class classification, regression and outlier (or novelty) detection via a robust optimization approach. The model embraces various machine learning models such as support vector machine-based and minimax probability machine-based classification and regression models. The unified framework makes it possible to compare and contrast existing learning models and to explain their differences and similarities.

In this paper, after relating existing learning models to UMLM, we show some theoretical properties for UMLM. Concretely, we show an interpretation of UMLM as minimizing a well-known financial risk measure (worst-case value-at risk (VaR) or conditional VaR), derive generalization bounds for UMLM using such a risk measure, and prove that solving problems of UMLM leads to estimators with the minimized generalization bounds. Those theoretical properties are applicable to related existing learning models.

1. Introduction

Classification and regression are widely studied topics in machine learning and artificial intelligence, and various new learning models have been proposed so far. However, there are few works providing a unified framework to such independently derived learning models in one formulation and investigating the generalization performance altogether. The unified framework makes it possible to compare and contrast existing learning models and to explain their differences and similarities, while individual formulations for each model may be helpful when developing efficient algorithms.

We propose a unified machine learning model (UMLM) for two-class classification, regression and outlier (or novelty) detection. The model is formulated via the robust optimization approach (Ben-Tal, El Ghaoui, & Nemirovski, 2009). Robust optimization is an approach for modeling an optimization problem with uncertain data varying in a given set and for finding decisions with the best worst-case performance under such uncertainty. There are several applications of robust optimization to two-class classification. Some of them such as Trafalis and Gilbert (2006) and Xu, Caramanis, and Mannor (2009) dealt with uncertainties of observations by assuming bounded uncertainties for each of input samples. Other works such as Lanckriet, El Ghaoui, Bhattacharyya, and Jordan (2002) and Shivaswamy, Bhattacharyya, and Smola (2006) handled the issue of robustness with respect to estimation errors (in the means and covariances of the classes) for the class-conditional distributions by using robust optimization techniques.

Our model, UMLM, is formulated in the same line as a unified *robust classification model* (RCM) proposed by Takeda, Mitsugi, and Kanamori (2012). Takeda et al. (2012) considered a different setup from the above methods for applying robust optimization to two-class classification. Let $\boldsymbol{x}_+$ and $\boldsymbol{x}_-$ be representative points of each class (for example, means or medians of the data points). Assume that $\boldsymbol{x}_+$ and $\boldsymbol{x}_-$ are uncertain and that they have confidence intervals (so-called *uncertainty sets*) $\mathcal{U}_+$ and $\mathcal{U}_-$, respectively. RCM is formulated based on robust optimization as

$$\max_{\boldsymbol{w}:\|\boldsymbol{w}\|_2=1} \min_{\boldsymbol{x}_+\in\mathcal{U}_+,\boldsymbol{x}_-\in\mathcal{U}_-} (\boldsymbol{x}_+ - \boldsymbol{x}_-)^\top \boldsymbol{w},$$

or equivalently,

$$\min_{\boldsymbol{w}:\|\boldsymbol{w}\|_2=1} \max_{\boldsymbol{x}\in\mathcal{U}} -\boldsymbol{x}^\top \boldsymbol{w}, \tag{1}$$

where $\|\cdot\|_2$ is the Euclidean norm, and $\mathcal{U}$ is the Minkowski difference of $\mathcal{U}_+$ and $\mathcal{U}_-$, i.e.,

$$\mathcal{U} = \mathcal{U}_+ \ominus \mathcal{U}_- := \{\boldsymbol{x}_+ - \boldsymbol{x}_- \mid \boldsymbol{x}_+ \in \mathcal{U}_+,\ \boldsymbol{x}_- \in \mathcal{U}_-\}.$$

* Corresponding author. Tel.: +81 3 5841 6959; fax: +81 3 5841 6959.
E-mail addresses: takeda@mist.i.u-tokyo.ac.jp (A. Takeda), kanamori@is.nagoya-u.ac.jp (T. Kanamori).

http://dx.doi.org/10.1016/j.neunet.2014.05.006
0893-6080/

Table 1
Correspondence with existing models. Each cell shows the corresponding two-class classifier, one-class classifier and regressor. "–" indicates that the case does not happen and the connected cell of (i) implies that two cases with $K = 1$ and $K = 2$ give the same learning model. The colored background indicates that the corresponding models have an interpretation from financial risk-measure viewpoints. Let $M := \{1, \ldots, m\}$ and $\widetilde{M} := \{m + 1, \ldots, 2m\}$. The variable $\boldsymbol{v}$ can include b and/or 1 in addition to $\boldsymbol{w}$ ($\boldsymbol{w}$ can also consist of two subvectors: $\boldsymbol{w}_1, \boldsymbol{w}_2$). φ is for V of (3) in UMLM (2). η and L are inputs for a reduced convex hull (RCH) of (9), and $K, \bar{\boldsymbol{z}}_k$ and $\Sigma_k,\ k = 1, \ldots, K$, are inputs for an ellipsoidal set of (11).

$\mathcal{U}^\eta$	Two-class	One-class	Regression
RCH ($\eta = \frac{1}{\nu m}$)	(a) Eν-SVC(Perez-Cruz et al., 2003), ν-SVC (Schölkopf et al., 2000) $\boldsymbol{z}_i = \begin{pmatrix} y_i\boldsymbol{x}_i \\ y_i \end{pmatrix}, \quad \boldsymbol{v} = \begin{pmatrix} \boldsymbol{w} \\ b \end{pmatrix}$ $L = M, \quad \varphi = 1$	(c) OC-ν-SVC (Schölkopf et al., 2001) $\boldsymbol{z}_i = \boldsymbol{x}_i, \quad \boldsymbol{v} = \boldsymbol{w}$ $L = M, \quad \varphi = 1$	(e) ν-SVR (Schölkopf et al., 2000) $\boldsymbol{z}_i = \begin{pmatrix} \boldsymbol{x}_i \\ -y_i \\ 1 \end{pmatrix}, \quad \boldsymbol{v} = \begin{pmatrix} \boldsymbol{w} \\ 1 \\ b \end{pmatrix}$ $\boldsymbol{z}_{m+i} = -\boldsymbol{z}_i$ $L = M \cup \widetilde{M}, \quad \varphi = C$
RCH ($\eta = \infty$)	(b) Two-class L2-SVC (Tsang et al., 2005) $\boldsymbol{z}_i = \begin{pmatrix} y_i\boldsymbol{x}_i \\ \frac{1}{\sqrt{C}}\boldsymbol{e}_i \\ y_i \end{pmatrix}, \quad \boldsymbol{v} = \begin{pmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ b \end{pmatrix}$ $L = M, \quad \varphi = 1$	(d) One-class L2-SVC (Tsang et al., 2005) $\boldsymbol{z}_i = \begin{pmatrix} \boldsymbol{x}_i & \frac{1}{\sqrt{C}}\boldsymbol{e}_i \end{pmatrix}, \quad \boldsymbol{v} = \begin{pmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \end{pmatrix}$ $L = M, \quad \varphi = 1$	
Ellips. ($K = 1$)	(g) FDA (Fukunaga, 1990), FS-FD (Bhattacharyya, 2004) $\boldsymbol{z}_i = y_i\boldsymbol{x}_i, \quad \boldsymbol{v} = \boldsymbol{w}$ $\bar{\boldsymbol{z}}_1 = \bar{\boldsymbol{z}}_{M_+} + \bar{\boldsymbol{z}}_{M_-}$ $\Sigma_1 = \Sigma_{M_+} + \Sigma_{M_-}$ $\varphi = 1$	(h) OC-MPM (Lanckriet et al., 2003) $\boldsymbol{z}_i = \begin{pmatrix} \boldsymbol{x}_i \\ -1 \end{pmatrix}, \quad \boldsymbol{v} = \begin{pmatrix} \boldsymbol{w} \\ 1 \end{pmatrix}$ $\bar{\boldsymbol{z}}_1 = \bar{\boldsymbol{z}}_M$ $\Sigma_1 = \Sigma_M$ $\varphi = 1$	(i) MPM regression (Strohmann & Grudic, 2002) $\boldsymbol{z}_i = \begin{pmatrix} \boldsymbol{x}_i \\ y_i + \epsilon \end{pmatrix}, \quad \boldsymbol{v} = \begin{pmatrix} \boldsymbol{w}_1 \\ w_2 \end{pmatrix}$ $\boldsymbol{z}_{m+i} = \begin{pmatrix} -\boldsymbol{x}_i \\ -y_i + \epsilon \end{pmatrix}$ $\bar{\boldsymbol{z}}_1 = \bar{\boldsymbol{z}}_M + \bar{\boldsymbol{z}}_{\widetilde{M}}$ $\Sigma_1 = \Sigma_M + \Sigma_{\widetilde{M}}$ $\varphi = 1$
Ellips. ($K = 2$)	(f) MPM (Lanckriet et al., 2002), MM-MPM (Nath & Bhattacharyya, 2007) $\boldsymbol{z}_i = y_i\boldsymbol{x}_i,\ \boldsymbol{v} = \boldsymbol{w}$ $\bar{\boldsymbol{z}}_1 = \bar{\boldsymbol{z}}_{M_+}, \quad \Sigma_1 = \Sigma_{M_+}$ $\bar{\boldsymbol{z}}_2 = \bar{\boldsymbol{z}}_{M_-}, \quad \Sigma_2 = \Sigma_{M_-}$ $\varphi = 1$	–	$\bar{\boldsymbol{z}}_1 = \bar{\boldsymbol{z}}_M, \quad \Sigma_1 = \Sigma_M$ $\bar{\boldsymbol{z}}_2 = \bar{\boldsymbol{z}}_{\widetilde{M}}, \quad \Sigma_2 = \Sigma_{\widetilde{M}}$ $\varphi = 1$

When $\mathcal{U}_+$ and $\mathcal{U}_-$ are convex sets and have interior points, their Minkowski difference $\mathcal{U}$ is convex and has a nonempty interior. Takeda et al. (2012) showed several examples of $\mathcal{U}_+$ and $\mathcal{U}_-$, leading to existing learning models.

In this paper, we propose a unified formulation, UMLM, similar to (1) not only for two-class classification but for regression and outlier detection, and relate existing learning models to UMLM. We show financial risk interpretation and generalization performance for UMLM, which are also applicable to related existing learning models. The purpose of this paper is as follows:

- to present a unified machine learning model (UMLM) not only for two-class classification but also for regression and outlier detection,
- to show an interpretation of UMLM as minimizing a well-known financial risk, worst-case value-at risk (VaR) or conditional VaR, and
- to derive generalization bounds for UMLM using financial risk measures and show that minimizing such risk measures, i.e., solving problems of UMLM, leads to estimators with the minimized generalization bounds.

The unified framework of UMLM makes it possible to find new models by examining the differences and similarities of existing models, though we focus on theoretical analysis of UMLM in this paper.

For Eν-support vector classification (SVC) (Perez-Cruz, Weston, Hermann, & Schölkopf, 2003), ν-SVC (Schölkopf, Smola, Williamson, & Bartlett, 2000) and ν-support vector regression (SVR) (Schölkopf et al., 2000) (i.e., existing models shown by colored background in Table 1), our previous studies (Takeda, Gotoh, & Sugiyama, 2010; Takeda & Sugiyama, 2008) showed their generalization bounds in addition to their interpretations of minimizing financial risk measures. This paper shows that UMLM admits similar generalization bound analysis and interpretations based on risk measures. Below, we summarize UMLM-based description of existing learning models shown in the paper.

Two-class classification: besides ν-SVC and Eν-SVC, we show that two-class L2-SVC (Tsang, Kwok, & Cheung, 2005), Fisher's discriminant analysis (FDA) (Fukunaga, 1990), and minimax probability machine (MPM) (Lanckriet et al., 2002) are related to UMLM when uncertainty sets are properly chosen. Their variants using the maximum margin criterion such as maximum-margin MPM (Nath & Bhattacharyya, 2007) are also presented as a sort of UMLM.

One-class classification: one-class (OC) classification methods are used for outlier detection, and it is closely related to the standard two-class classification. OC-ν-SVC (Schölkopf, Platt, Shawe-Taylor, Smola, & Williamson, 2001), OC L2-SVC (Tsang et al., 2005) and OC-MPM (Lanckriet, El Ghaoui, & Jordan, 2003) correspond to one-class variant of ν-SVC, L2-SVC, and MPM, respectively. We show that a modification of UMLMs of two-class classification methods immediately yields these learning methods.

Regression: in regression problems, the main task is to predict real-valued outputs for given multi-dimensional input vectors, and numerous number of works on statistical inference of regression problems have been published so far. In machine learning community, two-class classifiers such as ν-SVC and MPM are often employed to regression problems by replacing the margin loss with the other penalty on residuals to model fit. We show that ν-SVR (Schölkopf et al., 2000) and MPM regression (Strohmann & Grudic, 2002) are derived from UMLM by choosing appropriate uncertainty sets. Besides the above learning