

A Survey of Support Vector Machines with Uncertainties

Ximing Wang · Panos M. Pardalos

Received: 10 October 2014 / Revised: 1 November 2014 / Accepted: 10 December 2014 /
Published online: 17 January 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Support Vector Machines (SVM) is one of the well known supervised classes of learning algorithms. SVM have wide applications to many fields in recent years and also many algorithmic and modeling variations. Basic SVM models are dealing with the situation where the exact values of the data points are known. This paper presents a survey of SVM when the data points are uncertain. When a direct model cannot guarantee a generally good performance on the uncertainty set, robust optimization is introduced to deal with the worst case scenario and still guarantee an optimal performance. The data uncertainty could be an additive noise which is bounded by norm, where some efficient linear programming models are presented under certain conditions; or could be intervals with support and extremum values; or a more general case of polyhedral uncertainties with formulations presented. Another field of the uncertainty analysis is chance constrained SVM which is used to ensure the small probability of misclassification for the uncertain data. The multivariate Chebyshev inequality and Bernstein bounding schemes have been used to transform the chance constraints through robust optimization. The Chebyshev based model employs moment information of the uncertain training points. The Bernstein bounds can be less conservative than the Chebyshev bounds since it employs both support and moment information, but it also makes a strong assumption that all the elements in the data set are independent.

Keywords Support vector machines · Robust optimization · Bounded norm · Polyhedral uncertainties · Chance constraints

X. Wang (✉) · P. M. Pardalos
Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, USA
e-mail: x.wang@ufl.edu

P. M. Pardalos
e-mail: pardalos@ufl.edu

1 Introduction

As one of the well known supervised learning algorithms, Support Vector Machines (SVM) are gaining more and more attention. It was proposed by Vapnik [1,2] as a maximum-margin classifier, and tutorials on SVM could refer to [3–6]. In recent years, SVM have been applied to many fields and have many algorithmic and modeling variations. In the biomedical field, SVM have been used to identify physical diseases [7–10] as well as psychological diseases [11]. Electroencephalography (EEG) signals can also be analyzed using SVM [12–14]. Besides these, SVM also applied to protein prediction [15–19] and medical images [20–22]. Computer vision includes many applications of SVM like person identification [23], hand gesture detection [24], face recognition [25] and background subtraction [26]. In geosciences, SVM have been applied to remote sensing analysis [27–29], land cover change [30–32], landslide susceptibility [33–36] and hydrology [37,38]. In power systems, SVM was used for transient status prediction [39], power load forecasting [40], electricity consumption prediction [41] and wind power forecasting [42]. Stock price forecasting [43–45] and business administration [46] can also use SVM. Other applications of SVM include agriculture plant disease detection [47], condition monitoring [48], network security [49] and electronics [50,51]. When basic SVM models cannot satisfy the application requirement, different modeling variations of SVM can be found in [52].

In this paper, a survey of SVM with uncertainties is presented. Basic SVM models are dealing with the situation that the exact values of the data points are known. When the data points are uncertain, different models have been proposed to formulate the SVM with uncertainties. Bi and Zhang [53] assumed the data points are subject to an additive noise which is bounded by the norm and proposed a very direct model. However, this model cannot guarantee a generally good performance on the uncertainty set. To guarantee an optimal performance when the worst case scenario constraints are still satisfied, robust optimization is utilized. Trafalis et al. [54–58] proposed a robust optimization model when the perturbation of the uncertain data is bounded by norm. Ghaoui et al. [59] derived a robust model when the uncertainty is expressed as intervals. Fan et al. [60] studied a more general case for polyhedral uncertainties. Robust optimization is also used when the constraint is a chance constraint which is to ensure the small probability of misclassification for the uncertain data. The chance constraints are transformed by different bounding inequalities, for example multivariate Chebyshev inequality [61,62] and Bernstein bounding schemes [63].

The organization of this paper is as follows: Sect. 2 gives an introduction to the basic SVM models. Section 3 presents the SVM with uncertainties, stating both the robust SVM with bounded uncertainty and chance constrained SVM through robust optimization. Section 4 presents concluding remarks and suggesting for further research.

2 Basic SVM Models

Support Vector Machines construct maximum-margin classifiers, such that small perturbations in data are least likely to cause misclassification. Empirically, SVM works really well and are well known supervised learning algorithms proposed by Vap-

nik [1,2]. Suppose we have a two-class dataset of m data points $\{\mathbf{x}_i, y_i\}_{i=1}^m$ with n -dimensional features $\mathbf{x}_i \in \mathbb{R}^n$ and respective class labels $y_i \in \{+1, -1\}$. For linearly separable datasets, there exists a hyperplane $\mathbf{w}^\top \mathbf{x} + b = 0$ to separate the two classes and the corresponding classification rule is based on the sign($\mathbf{w}^\top \mathbf{x} + b$). If this value is positive, \mathbf{x} is classified to be in $+1$ class; otherwise, -1 class.

The datapoints that the margin pushes up against are called support vectors. A maximum-margin hyperplane is one that maximizes the distance between the hyperplane and the support vectors. For the separating hyperplane $\mathbf{w}^\top \mathbf{x} + b = 0$, \mathbf{w} and b could be normalized so that $\mathbf{w}^\top \mathbf{x} + b = +1$ goes through support vectors of $+1$ class, and $\mathbf{w}^\top \mathbf{x} + b = -1$ goes through support vectors of -1 class. The distance between these two hyperplane, i.e., the margin width, is $\frac{2}{\|\mathbf{w}\|_2}$, therefore, maximization of the margin can be performed as minimization of $\frac{1}{2}\|\mathbf{w}\|_2^2$ subject to separation constraints. This can be expressed as the following quadratic optimization problem:

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|_2^2 \tag{1a}$$

$$\text{s.t. } y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \quad i = 1, \dots, m \tag{1b}$$

Introducing Lagrange multipliers $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_m]$, the above constrained problem can be expressed as:

$$\min_{\mathbf{w}, b} \max_{\boldsymbol{\alpha} \geq 0} \mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} \|\mathbf{w}\|_2^2 - \sum_{i=1}^m \alpha_i [y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1] \tag{2}$$

Take the derivatives with respect to \mathbf{w} and b , and set to zero:

$$\frac{\partial \mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i \tag{3a}$$

$$\frac{\partial \mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial b} = 0 \Rightarrow \sum_{i=1}^m \alpha_i y_i = 0 \tag{3b}$$

Substituting into $\mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha})$:

$$\mathcal{L}(\boldsymbol{\alpha}) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j \tag{4}$$

Then the dual of the original SVM problem is also a convex quadratic problem:

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j \tag{5a}$$

$$\text{s.t. } \sum_{i=1}^m \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad i = 1, \dots, m \tag{5b}$$

Since only the α_i corresponding to support vectors can be nonzero, this dramatically simplifies solving the dual problem.

The above is in the case that the two classes are linearly separable. When they are not, mislabeled samples need to be allowed where soft margin SVM arises. Soft margin SVM introduces non-negative slack variables ξ_i to measure the distance of within-margin or misclassified data \mathbf{x}_i to the hyperplane with the correct label, and $\xi_i = \max\{0, 1 - y_i(\mathbf{w}^\top \mathbf{x}_i + b)\}$. When $0 < \xi_i < 1$, the data is within margin but correctly classified; when $\xi_i > 1$, the data is misclassified. The objective function is then adding a term that penalizes these slack variables, and the optimization is a trade off between a large margin and a small error penalty. The soft margin SVM formulation with L_1 regularization [64] is:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^m \xi_i \tag{6a}$$

$$\text{s.t. } y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, m \tag{6b}$$

where C is a trade-off parameter.

Similarly, the Lagrange of the soft margin SVM is:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} \max_{\alpha, \beta \geq 0} \mathcal{L}(\mathbf{w}, b, \xi, \alpha, \beta) &= \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^m \xi_i \\ &\quad - \sum_{i=1}^m \alpha_i [y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^m \beta_i \xi_i \end{aligned} \tag{7}$$

Take the derivative with respect to ξ_i and set to zero:

$$\frac{\partial \mathcal{L}(\mathbf{w}, b, \xi, \alpha, \beta)}{\partial \xi_i} = 0 \Rightarrow C - \alpha_i - \beta_i = 0 \tag{8}$$

Then $\alpha_i = C - \beta_i$. Since $\beta_i \geq 0$, it indicates that $\alpha_i \leq C$.

The derivatives with respect to \mathbf{w} and b are the same as before, substituting into $\mathcal{L}(\mathbf{w}, b, \xi, \alpha, \beta)$ and get the dual of the soft margin SVM:

$$\max_{\alpha} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j \tag{9a}$$

$$\text{s.t. } \sum_{i=1}^m \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \dots, m \tag{9b}$$

The only difference is that the dual variables α_i now have upper bounds C . The advantage of the L_1 regularization (linear penalty function) is that in the dual problem, the slack variables ξ_i vanish and the constant C is just an additional constraint on the

Lagrange multipliers α_i . Because of this nice property and its huge impact in practice, L_1 is the most widely used regularization term.

Besides the linear kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\top \mathbf{x}_j$, nonlinear kernels are also introduced into SVM to create nonlinear classifiers. The maximum-margin hyperplane is constructed in a high-dimensional transformed feature space with a possible nonlinear transformation, therefore, it could be nonlinear in the original feature space. A widely used nonlinear kernel is the Gaussian radial basis function $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|_2^2)$. It corresponds to a Hilbert space of infinite dimensions.

3 SVM with Uncertainties

Given m training data points in \mathbb{R}^n , use $X_i = [X_{i1}, \dots, X_{in}]^\top \in \mathbb{R}^n, i = 1, \dots, m$ to denote the uncertain training data points and $y_i \in \{+1, -1\}, i = 1, \dots, m$ to denote the respective class labels. The soft margin SVM with uncertainty is as following:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^m \xi_i \tag{10a}$$

$$\text{s.t. } y_i(\mathbf{w}^\top X_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, m \tag{10b}$$

When the training data points X_i are random vectors, the model needs to be modified to consider the uncertainties. The simplest model is to just employ the means of the uncertain data points, $\mu_i = \mathbf{E}[X_i]$. The formulation would become:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^m \xi_i \tag{11a}$$

$$\text{s.t. } y_i(\mathbf{w}^\top \mu_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, m \tag{11b}$$

The above model is equivalent to a soft margin SVM on data points fixed on the means, therefore does not take into account the uncertainties of the data. Bi and Zhang [53] assumed the data points are subject to an additive noise, $X_i = \bar{\mathbf{x}}_i + \Delta \mathbf{x}_i$ and the noise is bounded by $\|\Delta \mathbf{x}_i\|_2 \leq \delta_i$. Then they proposed the model as:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^m \xi_i \tag{12a}$$

$$\text{s.t. } y_i(\mathbf{w}^\top (\bar{\mathbf{x}}_i + \Delta \mathbf{x}_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, m \tag{12b}$$

$$\|\Delta \mathbf{x}_i\|_2 \leq \delta_i, \quad i = 1, \dots, m \tag{12c}$$

In this model, the uncertain data X_i is free in the circle centered at $\bar{\mathbf{x}}_i$ with radius equal to δ_i , i.e., X_i could move toward any direction in the uncertainty set. A drawback of this model is that it cannot guarantee a generally good performance on the uncertainty set since the direction of how the data points are perturbed is not constrained in this model. It is highly possible and already presented in this paper that a data point

with a perturbation making it move far away from the separation hyperplane could be used as the support vector. Then considering the original uncertainty set of this data point, it would be mostly lie within the margin and the constraint would not be satisfied any more. To guarantee a better performance under most conditions or with higher probability, robust optimization is introduced to solve the SVM with uncertainty.

3.1 Robust SVM with Bounded Uncertainty

Robust optimization is to guarantee an optimal performance under the worst case scenario. Given different information of the uncertain data, several models have been proposed. Trafalis et al. [54–58] proposed a model when the perturbation of the uncertain data is bounded by norm. The uncertain data could be expressed as $X_i = \bar{x}_i + \sigma_i$, the mean vector \bar{x}_i plus the additional perturbation σ_i which is bounded by the L_p norm with $\|\sigma_i\|_p \leq \eta_i$, for all $i = 1, \dots, m$. Robust optimization is to deal with the worst case perturbation, and this would be:

$$\min_{\|\sigma_i\|_p \leq \eta_i} y_i(\mathbf{w}^\top \bar{x}_i + b) + y_i \mathbf{w}^\top \sigma_i \geq 1 - \xi_i, \quad i = 1, \dots, m \tag{13}$$

To solve the robust SVM, the following subproblem needs to be solved first:

$$\min_{\sigma_i} y_i \mathbf{w}^\top \sigma_i \tag{14a}$$

$$\text{s.t. } \|\sigma_i\|_p \leq \eta_i \tag{14b}$$

Hölder’s inequality says that for a pair of dual norms L_p and L_q with $p, q \in [1, \infty]$ and $1/p + 1/q = 1$, the following inequality holds:

$$\|fg\|_1 \leq \|f\|_p \|g\|_q \tag{15}$$

Therefore

$$|y_i \mathbf{w}^\top \sigma_i| \leq \|\sigma_i\|_p \|\mathbf{w}\|_q \leq \eta_i \|\mathbf{w}\|_q \tag{16}$$

A lower bound of $y_i \mathbf{w}^\top \sigma_i$ is $-\eta_i \|\mathbf{w}\|_q$, substituting into the original problem will get the following formulation:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^m \xi_i \tag{17a}$$

$$\text{s.t. } y_i(\mathbf{w}^\top \bar{x}_i + b) - \eta_i \|\mathbf{w}\|_q \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, m \tag{17b}$$

The above formulation depends on the norm L_p . When $p = q = 2$, a conic program of the above formulation can be obtained:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^m \xi_i \tag{18a}$$

$$\text{s.t. } y_i(\mathbf{w}^\top \bar{\mathbf{x}}_i + b) - \eta_i \|\mathbf{w}\|_2 \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, m \tag{18b}$$

An interesting property of the norm transformation is that for L_1 and L_∞ norms, with the objective function $\frac{1}{2} \|\mathbf{w}\|_q + C \sum_{i=1}^m \xi_i$, the problem can be transformed into a linear programming (LP) problem.

The dual of L_1 norm is L_∞ norm. When $p = 1$, the formulation becomes:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|_\infty + C \sum_{i=1}^m \xi_i \tag{19a}$$

$$\text{s.t. } y_i(\mathbf{w}^\top \bar{\mathbf{x}}_i + b) - \eta_i \|\mathbf{w}\|_\infty \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, m \tag{19b}$$

Introducing an auxiliary variable $\alpha = \|\mathbf{w}\|_\infty$, then the above formulation can be written as a LP problem:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \alpha + C \sum_{i=1}^m \xi_i \tag{20a}$$

$$\text{s.t. } y_i(\mathbf{w}^\top \bar{\mathbf{x}}_i + b) - \eta_i \alpha \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, m \tag{20b}$$

$$\alpha \geq -w_j, \quad \alpha \geq w_j, \quad j = 1, \dots, n \tag{20c}$$

When the L_∞ norm is chosen to express the perturbation, then the formulation becomes:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|_1 + C \sum_{i=1}^m \xi_i \tag{21a}$$

$$\text{s.t. } y_i(\mathbf{w}^\top \bar{\mathbf{x}}_i + b) - \eta_i \|\mathbf{w}\|_1 \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, m \tag{21b}$$

Introducing an auxiliary vector α with $\alpha_j = |w_j|$, the resulting optimization problem is also LP:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \sum_{j=1}^n \alpha_j + C \sum_{i=1}^m \xi_i \tag{22a}$$

$$\text{s.t. } y_i(\mathbf{w}^\top \bar{\mathbf{x}}_i + b) - \eta_i \sum_{j=1}^n \alpha_j \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, m \tag{22b}$$

$$\alpha_j \geq -w_j, \quad \alpha_j \geq w_j, \quad j = 1, \dots, n \tag{22c}$$

Ghaoui et al. [59] derived a robust model when the uncertainty is expressed as intervals (also known as support or extremum values). Suppose the extremum values of the uncertain data points are known as $l_{ij} \leq X_{ij} \leq u_{ij}$, then each training data point X_i

is lying in a hyper-rectangle $\mathcal{R}_i = \{\mathbf{x}_i = [x_{i1}, \dots, x_{in}]^\top \in \mathbb{R}^n \mid l_{ij} \leq x_{ij} \leq u_{ij}, j = 1, \dots, n\}$ and the robust optimization requires that all points in the hyper-rectangle should satisfy $y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \forall \mathbf{x}_i \in \mathcal{R}_i$. The geometric center of the hyper-rectangle \mathcal{R}_i is $\mathbf{c}_i = [c_{i1}, \dots, c_{in}]^\top \in \mathbb{R}^n$ where $c_{ij} = (l_{ij} + u_{ij})/2, j = 1, \dots, n$. The semi-lengths of the sides of the hyper-rectangle \mathcal{R}_i is $s_{ij} = (u_{ij} - l_{ij})/2, j = 1, \dots, n$. Then the worst case with these interval information would be:

$$y_i(\mathbf{w}^\top \mathbf{c}_i + b) \geq 1 - \xi_i + \sum_{j=1}^n s_{ij}|w_j| \tag{23}$$

Then the SVM model with support information can be written as:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^m \xi_i \tag{24a}$$

$$\text{s.t. } y_i(\mathbf{w}^\top \mathbf{c}_i + b) \geq 1 - \xi_i + \|\mathbf{S}_i \mathbf{w}\|_1, \quad \xi_i \geq 0, \quad i = 1, \dots, m \tag{24b}$$

where \mathbf{S}_i is a diagonal matrix with entries s_{ij} .

The interval uncertainty is a special case of polyhedral uncertainty [60]. The polyhedral uncertainty can be expressed as $\mathbf{D}_i \mathbf{x}_i \leq \mathbf{d}_i$, where the matrix $\mathbf{D}_i \in \mathbb{R}^{q \times n}$ and the vector $\mathbf{d}_i \in \mathbb{R}^q$. And since zero vectors could be added to obtain the same number q of inequalities for all data points, q is the largest dimension of the uncertainties of all the points. The robust SVM with polyhedral uncertainty is:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^m \xi_i \tag{25a}$$

$$\text{s.t. } \min_{\{\mathbf{x}_i: \mathbf{D}_i \mathbf{x}_i \leq \mathbf{d}_i\}} y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, m \tag{25b}$$

Since $\min_{\{\mathbf{x}_i: \mathbf{D}_i \mathbf{x}_i \leq \mathbf{d}_i\}} y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i$ is equivalent to:

$$\max_{\{\mathbf{x}_i: \mathbf{D}_i \mathbf{x}_i \leq \mathbf{d}_i\}} (-y_i \mathbf{w}^\top \mathbf{x}_i) - y_i b \leq -1 + \xi_i \tag{26}$$

To solve

$$\max \quad -y_i \mathbf{w}^\top \mathbf{x}_i \tag{27a}$$

$$\text{s.t. } \mathbf{D}_i \mathbf{x}_i \leq \mathbf{d}_i \tag{27b}$$

The dual is:

$$\min \mathbf{d}_i^\top \mathbf{z}_i \tag{28a}$$

$$\text{s.t. } \mathbf{D}_i^\top \mathbf{z}_i = -y_i \mathbf{w} \tag{28b}$$

$$\mathbf{z}_i = (z_{i1}, \dots, z_{iq})^\top \geq 0 \tag{28c}$$

Strong duality would guarantee that the objective values of the dual and primal are equal. Therefore, the robust SVM with polyhedral uncertainty formulation is equivalent to:

$$\min_{\mathbf{w}, b, \xi_i, \mathbf{z}} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^m \xi_i \tag{29a}$$

$$\text{s.t. } \mathbf{d}_i^\top \mathbf{z}_i - y_i b \leq -1 + \xi_i, \quad \xi_i \geq 0 \tag{29b}$$

$$\mathbf{D}_i^\top \mathbf{z}_i + y_i \mathbf{w} = 0, \quad \mathbf{z}_i = (z_{i1}, \dots, z_{iq})^\top \tag{29c}$$

$$z_{ij} \geq 0, \quad i = 1, \dots, m, \quad j = 1, \dots, q \tag{29d}$$

The authors also proved that for the hard margin SVM (i.e., when there is no ξ_i), the dual of the above formulation is:

$$\min_{\lambda, \mu} \sum_{i=1}^m \lambda_i - \frac{1}{2} \sum_{k=1}^n \left(\sum_{i=1}^m y_i \mu_{ik} \right)^2 \tag{30a}$$

$$\text{s.t. } \lambda_i d_{ij} + \sum_{k=1}^n \mu_{ik} D_{ijk} = 0, \quad i = 1, \dots, m, \quad j = 1, \dots, q \tag{30b}$$

$$\sum_{i=1}^m \lambda_i y_i = 0 \tag{30c}$$

$$\lambda_i \geq 0, \quad i = 1, \dots, m \tag{30d}$$

The interval uncertainty $[\mathbf{x}_i^0 - \delta_i, \mathbf{x}_i^0 + \delta_i]$ is a special case of polyhedral uncertainty since when defining

$$\mathbf{D}_i = \begin{pmatrix} I \\ -I \end{pmatrix}, \quad \mathbf{d}_i = \begin{pmatrix} \mathbf{x}_i^0 + \delta_i \\ -\mathbf{x}_i^0 + \delta_i \end{pmatrix} \tag{31}$$

$\{\mathbf{x}_i : \mathbf{x}_i \in [\mathbf{x}_i^0 - \delta_i, \mathbf{x}_i^0 + \delta_i]\}$ and $\{\mathbf{x}_i : \mathbf{D}_i \mathbf{x}_i \leq \mathbf{d}_i\}$ are equivalent. The authors of [60] also proposed probabilistic bounds on constraint violation in this case.

3.2 Chance Constrained SVM through Robust Optimization

The chance-constrained program (CCP) is used to ensure the small probability of misclassification for the uncertain data. The chance-constrained SVM formulation is:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^m \xi_i \tag{32a}$$

$$\text{s.t. } \text{Prob}\left\{y_i(\mathbf{w}^\top X_i + b) \leq 1 - \xi_i\right\} \leq \varepsilon, \quad \xi_i \geq 0, \quad i = 1, \dots, m \tag{32b}$$

where $0 < \varepsilon \leq 1$ is a parameter given by the user and close to 0. This model ensures an upper bound on the misclassification probability, but the chance constraints are typically non-convex so the problem is very hard to solve.

The work so far to deal with the chance constraint is to transform them by different bounding inequalities. When the mean and covariance matrix are known, the multivariate Chebyshev bound via robust optimization can be used to express the chance constraints above [61,62].

Markov’s inequality states that if X is a nonnegative random variable and $a > 0$, then

$$\text{Prob}\{X \geq a\} \leq \frac{\mathbf{E}[X]}{a} \tag{33}$$

Consider the random variable $(X - \mathbf{E}[X])^2$. Since $\mathbf{Var}(X) = \mathbf{E}[(X - \mathbf{E}[X])^2]$, then

$$\text{Prob}\{(X - \mathbf{E}[X])^2 \geq a^2\} \leq \frac{\mathbf{Var}(X)}{a^2} \tag{34}$$

which yields the Chebyshev’s inequality

$$\text{Prob}\{|X - \mathbf{E}[X]| \geq a\} \leq \frac{\mathbf{Var}(X)}{a^2} \tag{35}$$

Let $\mathbf{x} \sim (\mu, \Sigma)$ denote the random vector \mathbf{x} with mean μ and covariance matrix Σ . The multivariate Chebyshev inequality [65,66] states that for an arbitrary closed convex set S , the supremum of the probability that \mathbf{x} takes a value in S is

$$\sup_{\mathbf{x} \sim (\mu, \Sigma)} \text{Prob}\{\mathbf{x} \in S\} = \frac{1}{1 + d^2} \tag{36a}$$

$$d^2 = \inf_{\mathbf{x} \in S} (\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu) \tag{36b}$$

For the constraint $\text{Prob}\{\mathbf{w}^\top \mathbf{x} + b \leq 0\} \leq \varepsilon$, it could be derived that:

$$\mathbf{w}^\top \mu + b \geq \kappa_C \|\Sigma^{\frac{1}{2}} \mathbf{w}\|_2 \tag{37}$$

where $\kappa_C = \sqrt{(1 - \varepsilon)/\varepsilon}$.

Applying the above result to the chance constrained SVM, the Chebyshev based reformulation utilizing the means μ_i and covariance matrix Σ_i of each uncertain training point X_i can be obtained as the following robust model [61,62]:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^m \xi_i \tag{38a}$$

$$\text{s.t. } y_i (\mathbf{w}^\top \mu_i + b) \geq 1 - \xi_i + \kappa_C \|\Sigma_i^{\frac{1}{2}} \mathbf{w}\|_2, \quad \xi_i \geq 0, \quad i = 1, \dots, m \tag{38b}$$

Another approach to study SVM with chance constraints is to use Bernstein approximation schemes [67–69]. Ben-Tal et al. [63] employed Bernstein bounding schemes for the CCP relaxation and transformed the problem as a convex second order cone program with robust set constraints to guarantee the satisfaction of the chance constraints and can be solved efficiently using interior point solvers.

The Bernstein based relaxation utilized both the support (bounds, i.e. extremum values of the data points) and moment information (mean and variance). For random data point $X_i = [X_{i1}, \dots, X_{in}]^T$ and its label y_i , support information is the bounds of the data points $l_{ij} \leq X_{ij} \leq u_{ij}$, i.e. $X_i \in \mathcal{R}_i = \{x_i = [x_{i1}, \dots, x_{in}]^T \in \mathbb{R}^n \mid l_{ij} \leq x_{ij} \leq u_{ij}, j = 1, \dots, n\}$, 1st moment information is the bounds on the means of the data points $\mu_i^- = [\mu_{i1}^-, \dots, \mu_{in}^-]^T \leq \mu_i = \mathbf{E}[X_i] = [\mathbf{E}[X_{i1}], \dots, \mathbf{E}[X_{in}]]^T \leq \mu_i^+ = [\mu_{i1}^+, \dots, \mu_{in}^+]^T$, and 2nd moment information is the bounds on the second-moments of the data points $0 \leq \mathbf{E}[X_{ij}^2] \leq \sigma_{ij}^2$.

The Bernstein based relaxation is to derive convex constraints so that when these convex constraints are satisfied then the chance-constraints are guaranteed to be satisfied. They proved that with the information of independent random variable X_{ij} , i.e. support $l_{ij} \leq X_{ij} \leq u_{ij}$, bounds on the first-moment $\mu_{ij}^- \leq \mu_{ij} = \mathbf{E}[X_{ij}] \leq \mu_{ij}^+$, and bounds on the second-moment $0 \leq \mathbf{E}[X_{ij}^2] \leq \sigma_{ij}^2$, the chance-constraint in SVM is satisfied if the following convex constraint holds:

$$1 - \xi_i - y_i b + \sum_j \left(\max[-y_i \mu_{ij}^- w_j, -y_i \mu_{ij}^+ w_j] \right) + \kappa_B \|\Sigma_i \mathbf{w}\|_2 \leq 0 \tag{39}$$

where $\kappa_B = \sqrt{2 \log(1/\epsilon)}$, and the diagonal matrix

$$\Sigma_i = \text{diag} \left(s_{i1} v(\mu_{i1}^-, \mu_{i1}^+, \sigma_{i1}), \dots, s_{in} v(\mu_{in}^-, \mu_{in}^+, \sigma_{in}) \right) \tag{40}$$

where $s_{ij} = \frac{u_{ij} - l_{ij}}{2}$ and the function $v(\mu_{ij}^-, \mu_{ij}^+, \sigma_{ij})$ is defined by normalizing $\hat{X}_{ij} = \frac{X_{ij} - c_{ij}}{s_{ij}}$, where $c_{ij} = \frac{l_{ij} + u_{ij}}{2}$ and $s_{ij} = \frac{u_{ij} - l_{ij}}{2}$. Using the information of X_{ij} , one can easily compute the moment information of \hat{X}_{ij} , which are denoted by $\hat{\mu}_{ij}^- \leq \hat{\mu}_{ij} = \mathbf{E}[\hat{X}_{ij}] \leq \hat{\mu}_{ij}^+$ and $0 \leq \mathbf{E}[\hat{X}_{ij}^2] \leq \hat{\sigma}_{ij}^2$. They proved that

$$\mathbf{E} \left[\exp\{\tilde{t} \hat{X}_{ij}\} \right] \leq g_{\hat{\mu}_{ij}, \hat{\sigma}_{ij}}(\tilde{t}) = \begin{cases} \frac{(1 - \hat{\mu}_{ij})^2 \exp\left\{\tilde{t} \frac{\hat{\mu}_{ij} - \hat{\sigma}_{ij}^2}{1 - \hat{\mu}_{ij}}\right\} + (\hat{\sigma}_{ij}^2 - \hat{\mu}_{ij}^2) \exp\{\tilde{t}\}}{1 - 2\hat{\mu}_{ij} + \hat{\sigma}_{ij}^2}, & \tilde{t} \geq 0 \\ \frac{(1 + \hat{\mu}_{ij})^2 \exp\left\{\tilde{t} \frac{\hat{\mu}_{ij} + \hat{\sigma}_{ij}^2}{1 + \hat{\mu}_{ij}}\right\} + (\hat{\sigma}_{ij}^2 - \hat{\mu}_{ij}^2) \exp\{-\tilde{t}\}}{1 + 2\hat{\mu}_{ij} + \hat{\sigma}_{ij}^2}, & \tilde{t} \leq 0 \end{cases} \tag{41}$$

They defined $h_{\hat{\mu}_{ij}, \hat{\sigma}_{ij}}(\tilde{t}) = \log g_{\hat{\mu}_{ij}, \hat{\sigma}_{ij}}(\tilde{t})$, and the function $v(\mu^-, \mu^+, \sigma)$ is defined as:

$$\begin{aligned}
 v(\mu^-, \mu^+, \sigma) = \min \left\{ k \geq 0 : h_{\hat{\mu}, \hat{\sigma}}(\tilde{t}) \leq \max[\hat{\mu}^- \tilde{t}, \hat{\mu}^+ \tilde{t}] + \frac{k^2}{2} \tilde{t}^2, \right. \\
 \left. \forall \hat{\mu} \in [\hat{\mu}^-, \hat{\mu}^+], \tilde{t} \right\} \tag{42}
 \end{aligned}$$

This value can be calculated numerically. Under the condition that $\mu_{ij}^- \leq c_{ij} \leq \mu_{ij}^+$, this value can be computed analytically by $v(\mu^-, \mu^+, \sigma) = \sqrt{1 - (\hat{\mu}^{\min})^2}$, where $\hat{\mu}^{\min} = \min(-\hat{\mu}^-, \hat{\mu}^+)$.

Replacing the chance-constraints in SVM by the convex constraint derived above, the problem is transformed into a convex second order cone program:

$$\min_{\mathbf{w}, b, \xi_i, z_{ij}} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^m \xi_i \tag{43a}$$

$$\text{s.t. } 1 - \xi_i - y_i b + \sum_j z_{ij} + \kappa_B \|\Sigma_i \mathbf{w}\|_2 \leq 0 \tag{43b}$$

$$z_{ij} \geq -y_i \mu_{ij}^- w_j, \quad z_{ij} \geq -y_i \mu_{ij}^+ w_j \tag{43c}$$

$$\xi_i \geq 0, \quad i = 1, \dots, m \tag{43d}$$

which can be solved efficiently using cone programming solvers.

The geometrical interpretation of this convex constraint is that $y_i(\mathbf{w}^\top \mathbf{x} + b) \geq 1 - \xi_i$ is satisfied for all \mathbf{x} belonging to the union of ellipsoids $\mathcal{E}(\mu_i, \kappa_B \Sigma_i) = \{\mathbf{x} = \mu_i + \kappa_B \Sigma_i \mathbf{a} : \|\mathbf{a}\|_2 \leq 1\}$ with center μ_i , shape and size $\kappa_B \Sigma_i$, and the union is over $\mu_i \in [\mu_i^-, \mu_i^+]$, i.e.,

$$y_i(\mathbf{w}^\top \mathbf{x} + b) \geq 1 - \xi_i, \quad \forall \mathbf{x} \in \cup_{\mu_i \in [\mu_i^-, \mu_i^+]} \mathcal{E}(\mu_i, \kappa_B \Sigma_i) \tag{44}$$

Therefore, this constraint is defining an uncertainty set $\cup_{\mu_i \in [\mu_i^-, \mu_i^+]} \mathcal{E}(\mu_i, \kappa_B \Sigma_i)$ for each uncertain training data point X_i . If all the points in the uncertainty set satisfy $y_i(\mathbf{w}^\top \mathbf{x} + b) \geq 1 - \xi_i$, then the chance-constraint is guaranteed to be satisfied. This transforms the CCP into a robust optimization problem over the uncertainty set.

Since the size of the uncertainty set depend on κ_B , and $\kappa_B = \sqrt{2 \log(1/\varepsilon)}$, when the upperbound of misclassification error ε decreases, the size of the uncertainty set increases. When ε is very small, the uncertainty set would become huge so the constraint would be too conservative. As the support information provides with the bounding hyper-rectangle \mathcal{R}_i where the true training data point X_i would always lie in, a less conservative classifier can be obtained by taking the intersection of $\cup_{\mu_i \in [\mu_i^-, \mu_i^+]} \mathcal{E}(\mu_i, \kappa_B \Sigma_i)$ and \mathcal{R}_i as the new uncertainty set.

The authors proved that when the uncertainty set is the intersection, i.e.,

$$y_i(\mathbf{w}^\top \mathbf{x} + b) \geq 1 - \xi_i, \quad \forall \mathbf{x} \in \left(\cup_{\mu_i \in [\mu_i^-, \mu_i^+]} \mathcal{E}(\mu_i, \kappa_B \Sigma_i) \right) \cap \mathcal{R}_i \tag{45}$$

The above constraint is satisfied if and only if the following convex constraint holds:

$$\sum_j \left(\max[-l_{ij}(y_i w_j + a_{ij}), -u_{ij}(y_i w_j + a_{ij})] + \max[\mu_{ij}^- a_{ij}, \mu_{ij}^+ a_{ij}] \right) + 1 - \xi_i - y_i b + \kappa_B \|\Sigma_i \mathbf{a}_i\|_2 \leq 0 \tag{46}$$

Replacing the chance-constraints in SVM by the robust but less conservative convex constraint above, the problem is transformed into the following SOCP:

$$\min_{\mathbf{w}, b, \xi_i, z_{ij}, \tilde{z}_{ij}, \mathbf{a}_i} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^m \xi_i \tag{47a}$$

$$\text{s.t. } 1 - \xi_i - y_i b + \sum_j \tilde{z}_{ij} + \sum_j z_{ij} + \kappa_B \|\Sigma_i \mathbf{a}_i\|_2 \leq 0 \tag{47b}$$

$$z_{ij} \geq \mu_{ij}^- a_{ij}, \quad z_{ij} \geq \mu_{ij}^+ a_{ij} \tag{47c}$$

$$\tilde{z}_{ij} \geq -l_{ij}(y_i w_j + a_{ij}), \quad \tilde{z}_{ij} \geq -u_{ij}(y_i w_j + a_{ij}) \tag{47d}$$

$$\xi_i \geq 0, \quad i = 1, \dots, m \tag{47e}$$

The Bernstein based formulations (43) and (47) are robust to the moment estimation errors in addition to the uncertainty in data, since they are using the bounds on mean (μ_{ij}^-, μ_{ij}^+) and bounds on second-moment (σ_{ij}^2) instead of the exact values of the moments which are often unknown.

Comparing the two approaches for the chance constrained SVM, both of them are robust to uncertainties in data and did not make assumptions to the underlying probability distribution. Chebyshev based schemes only employed moment information of the uncertain training points, while Bernstein bounds employed both support and moment information, therefore can be less conservative than Chebyshev bounds. The resulting classifier by Bernstein approach achieved larger classification margins and therefore better generalization ability according to the structural risk minimization principle of Vapnik [1]. A drawback of Bernstein based formulation is that it assumes each element X_{ij} is independent with each other, while Chebyshev based formulation allows the covariance matrix Σ_i of uncertain training point X_i .

4 Concluding Remarks

This paper presented a survey on SVM with uncertainties. When direct model cannot guarantee a generally good performance on the uncertainty set, robust optimization is utilized to obtain an optimal performance under the worst case scenario. The perturbation of the uncertain data could be bounded by the norm, or expressed as intervals and polyhedrons. When the constraint is a chance constraint, different bounding schemes like multivariate Chebyshev inequality and Bernstein bounding schemes are used to ensure the small probability of misclassification for the uncertain data.

The models in the literature are generally processing the linear SVM. A big part of the power of SVM lies in the powerful representation of nonlinear kernel in SVM mod-

els, which is to generate nonlinear classification boundaries. Therefore, it is suggested that more study could be conducted to explore how to deal with nonlinear kernels. And more schemes could be explored to represent the robust regions of the uncertain data and formulate the models as convex solvable problems.

Acknowledgments Research is supported by RSF Grant 14-41-00039.

References

- Vapnik VN (1998) *Statistical learning theory*. Wiley, New York
- Vapnik VN (1999) An overview of statistical learning theory. *IEEE Trans Neural Netw* 10(5):988–999
- Burges CJ (1998) A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery* 2(2):121–167
- Abe S (2010) *Support vector machines for pattern classification*. Springer, Heidelberg
- Ben-Hur A, Weston J (2010) A users guide to support vector machines. In: Carugo O, Eisenhaber F (eds) *Data mining techniques for the life sciences*. Springer, Berlin, pp 223–239
- Chang CC, Lin CJ (2011) Libsvm: a library for support vector machines. *ACM Trans Intell Syst Technol (TIST)* 2(3):27
- Mueller A, Candrian G, Grane VA, Kropotov JD, Ponomarev VA, Baschera GM (2011) Discriminating between adhd adults and controls using independent erp components and a support vector machine: a validation study. *Nonlinear Biomed Phys* 5(1):5
- Xie J, Wang C (2011) Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases. *Expert Syst Appl* 38(5):5809–5815
- Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A (2012) Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev* 36(4):1140–1152
- Ramírez J, Górriz J, Salas-Gonzalez D, Romero A, López M, Álvarez I, Gómez-Río M (2013) Computer-aided diagnosis of Alzheimers type dementia combining support vector machines and discriminant set of features. *Inf Sci* 237:59–72
- Mourao-Miranda J, Reinders A, Rocha-Rego V, Lappin J, Rondina J, Morgan C, Morgan KD, Fearon P, Jones PB, Doody GA et al.(2012) Individualized prediction of illness course at the first psychotic episode: a support vector machine mri study. *Psychol Med* 42(05):1037–1047
- Subasi A, Ismail Gursoy M (2010) EEG signal classification using PCA, ICA, IDA and support vector machines. *Expert Syst Appl* 37(12):8659–8666
- Übeyli ED (2010) Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals. *Expert Syst Appl* 37(1):233–239
- Nicolaou N, Georgiou J (2012) Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. *Expert Syst Appl* 39(1):202–209
- Qiu JD, Huang JH, Shi SP, Liang RP (2010) Using the concept of chou's pseudo amino acid composition to predict enzyme family classes: an approach with support vector machine based on discrete wavelet transform. *Protein Pept Lett* 17(6):715–722
- Kumar Kandaswamy K, Pugalenti G, Moller S, Hartmann E, Uwe Kalies K, N Suganthan P, Martinetz T (2010) Prediction of apoptosis protein locations with genetic algorithms and support vector machines through a new mode of pseudo amino acid composition. *Protein Pept Lett* 17(12):1473–1479
- Bikadi Z, Hazai I, Malik D, Jemnitz K, Veres Z, Hari P, Ni Z, Loo TW, Clarke DM, Hazai E et al.(2011) Predicting p-glycoprotein-mediated drug transport based on support vector machine and three-dimensional crystal structure of p-glycoprotein. *PLoS One* 6(10):e25–815
- Lise S, Buchan D, Pontil M, Jones DT (2011) Predictions of hot spot residues at protein-protein interfaces using support vector machines. *PLoS One* 6(2):e16,774
- Mohabatkar H, Mohammad Beigi M, Esmaeili A (2011) Prediction of gaba receptor proteins using the concept of Chou's pseudo-amino acid composition and support vector machine. *J Theor Biol* 281(1):18–23
- Morra JH, Tu Z, Apostolova LG, Green AE, Toga AW, Thompson PM (2010) Comparison of adaboost and support vector machines for detecting Alzheimer's disease through automated hippocampal segmentation. *IEEE Trans Med Imaging* 29(1):30–43

21. Bauer S, Nolte LP, Reyes M (2011) Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In: Fichtinger G, Martel A, Peters T (eds) *Proceedings of Medical Image Computing and Computer-Assisted Intervention-MICCAI 2011*. Springer, Berlin, pp 354–361
22. Yao J, Dwyer A, Summers RM, Mollura DJ (2011) Computer-aided diagnosis of pulmonary infections using texture analysis and support vector machine classification. *Acad Radiol* 18(3):306–314
23. Prosser B, Zheng WS, Gong S, Xiang T, Mary Q (2010) Person re-identification by support vector ranking. *BMVC* 1:5
24. Dardas NH, Georganas ND (2011) Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. *IEEE Trans Instrum Meas* 60(11):3592–3607
25. Wei J, Jian-qi Z, Xiang Z (2011) Face recognition method based on support vector machine and particle swarm optimization. *Expert Syst Appl* 38(4):4390–4393
26. Han B, Davis LS (2012) Density-based multifeature background subtraction with support vector machine. *IEEE Trans Pattern Anal Mach Intell* 34(5):1017–1023
27. Waske B, van der Linden S, Benediktsson JA, Rabe A, Hostert P (2010) Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Trans Geosci Remote Sens* 48(7):2880–2889
28. Mountrakis G, Im J, Ogole C (2011) Support vector machines in remote sensing: a review. *ISPRS J Photogramm Remote Sens* 66(3):247–259
29. Li CH, Kuo BC, Lin CT, Huang CS (2012) A spatial-contextual support vector machine for remotely sensed image classification. *IEEE Trans Geosci Remote Sens* 50(3):784–799
30. Otukey J, Blaschke T (2010) Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *Int J Appl Earth Obs Geoinf* 12:527–531
31. Shao Y, Lunetta RS (2012) Comparison of support vector machine, neural network, and cart algorithms for the land-cover classification using limited training data points. *ISPRS J Photogramm Remote Sens* 70:78–87
32. Volpi M, Tuia D, Bovolo F, Kanevski M, Bruzzone L (2013) Supervised change detection in VHR images using contextual information and support vector machines. *Int J Appl Earth Obs Geoinf* 20:77–85
33. Yilmaz I (2010) Comparison of landslide susceptibility mapping methodologies for koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environ Earth Sci* 61(4):821–836
34. Tien Bui D, Pradhan B, Lofman O, Revhaug I (2012) Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and naive bayes models. *Math Problems Eng* 2012: Article ID 974638
35. Xu C, Dai F, Xu X, Lee YH (2012) Gis-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River Watershed, China. *Geomorphology* 145:70–80
36. Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput Geosci* 51:350–365
37. Kisi O, Cimen M (2011) A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *J Hydrol* 399(1):132–140
38. Yoon H, Jun SC, Hyun Y, Bae GO, Lee KK (2011) A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J Hydrol* 396(1):128–138
39. Gomez FR, Rajapakse AD, Annakkage UD, Fernando IT (2011) Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements. *IEEE Trans Power Syst* 26(3):1474–1483
40. Niu D, Wang Y, Wu DD (2010) Power load forecasting using support vector machine and ant colony optimization. *Expert Syst Appl* 37(3):2531–2539
41. Kavaklioglu K (2011) Modeling and prediction of Turkey's electricity consumption using support vector regression. *Appl Energy* 88(1):368–375
42. Zhou J, Shi J, Li G (2011) Fine tuning support vector machines for short-term wind speed forecasting. *Energy Convers Manag* 52(4):1990–1998
43. Kara Y, Acar Boyacioglu M, Baykan ÖK (2011) Predicting direction of stock price index movement using artificial neural networks and support vector machines: the sample of the istanbul stock exchange. *Expert Syst Appl* 38(5):5311–5319

44. Yeh CY, Huang CW, Lee SJ (2011) A multiple-kernel support vector regression approach for stock market price forecasting. *Expert Syst Appl* 38(3):2177–2186
45. Huang CF (2012) A hybrid stock selection model using genetic algorithms and support vector regression. *Appl Soft Comput* 12(2):807–818
46. Yang XS, Deb S, Fong S (2011) Accelerated particle swarm optimization and support vector machine for business optimization and applications. In: Fong S (ed) *Networked digital technologies*. Springer, Berlin, pp 53–66
47. Rumpf T, Mahlein AK, Steiner U, Oerke EC, Dehne HW, Plümer L (2010) Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Comput Electron Agric* 74(1):91–99
48. Konar P, Chattopadhyay P (2011) Bearing fault detection of induction motor using wavelet and support vector machines (SVMS). *Appl Soft Comput* 11(6):4203–4211
49. Horng SJ, Su MY, Chen YH, Kao TW, Chen RJ, Lai JL, Perkasa CD (2011) A novel intrusion detection system based on hierarchical clustering and support vector machines. *Expert Syst Appl* 38(1):306–313
50. Wong PK, Xu Q, Vong CM, Wong HC (2012) Rate-dependent hysteresis modeling and control of a piezostage using online support vector machine and relevance vector machine. *IEEE Trans Ind Electron* 59(4):1988–2001
51. Cui J, Wang Y (2011) A novel approach of analog circuit fault diagnosis using support vector machines classifier. *Measurement* 44(1):281–289
52. Tian Y, Shi Y, Liu X (2012) Recent advances on support vector machines research. *Technol Econ Dev Econ* 18(1):5–33
53. Bi J, Zhang T (2004) Support vector classification with input data uncertainty. *Adv Neural Inf Process Syst* 17:161–168
54. Trafalis TB, Gilbert RC (2006) Robust classification and regression using support vector machines. *Eur J Op Res* 173(3):893–909
55. Trafalis TB, Gilbert RC (2007) Robust support vector machines for classification and computational issues. *Optim Methods Softw* 22(1):187–198
56. Trafalis TB, Alwazzy SA (2010) Support vector machine classification with noisy data: a second order cone programming approach. *Int J Gen Syst* 39(7):757–781
57. Pant R, Trafalis TB, Barker K (2011) Support vector machine classification of uncertain and imbalanced data using robust optimization. In: *Proceedings of the 15th WSEAS international conference on computers, World Scientific and Engineering Academy and Society (WSEAS)*, pp 369–374
58. Xanthopoulos P, Pardalos PM, Trafalis TB (2012) *Robust data mining*. Springer, New York
59. Ghaoui LE, Lanckriet GR, Natsoulis G (2003) Robust classification with interval data. Technical report UCB/CSD-03-1279, Computer Science Division, University of California, Berkeley
60. Fan N, Sadeghi E, Pardalos PM (2014) Robust support vector machines with polyhedral uncertainty of the input data. In: Pardalos PM, Resende MGC, Vogiatzis C, Walteros JL (eds) *Learning and intelligent optimization*. Springer, Berlin, pp 291–305
61. Bhattacharyya C, Grate LR, Jordan MI, El Ghaoui L, Mian IS (2004) Robust sparse hyperplane classifiers: application to uncertain molecular profiling data. *J Comput Biol* 11(6):1073–1089
62. Shivaswamy PK, Bhattacharyya C, Smola AJ (2006) Second order cone programming approaches for handling missing and uncertain data. *J Mach Learn Res* 7:1283–1314
63. Ben-Tal A, Bhadra S, Bhattacharyya C, Nath JS (2011) Chance constrained uncertain classification via robust optimization. *Math Program* 127(1):145–173
64. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
65. Marshall AW, Olkin I (1960) Multivariate Chebyshev inequalities. *Ann Math Stat* 31(4):1001–1014
66. Bertsimas D, Popescu I (2005) Optimal inequalities in probability theory: a convex optimization approach. *Siam J Optim* 15(3):780–804
67. Ben-Tal A, Ghaoui LE, Nemirovski A (2009) *Robust optimization*. Princeton University Press, Princeton
68. Ben-Tal A, Nemirovski A (2008) Selected topics in robust convex optimization. *Math Program* 112(1):125–158
69. Nemirovski A, Shapiro A (2006) Convex approximations of chance constrained programs. *Siam J Optim* 17(4):969–996



Ximing Wang received her BS degree in Civil Engineering and MS degree in Geodesy and Survey Engineering from Tsinghua University, Beijing, China in 2008 and 2011. She is now working at her PhD degree in Industrial and Systems Engineering at the University of Florida, Gainesville, FL, USA. Her research interests include data mining, machine learning, stochastic programming, and robust optimization.



Panos M. Pardalos received his PhD degree in Computer and Information Sciences from the University of Minnesota, Minneapolis, MN, USA. He serves as Distinguished Professor of Industrial and Systems Engineering at the University of Florida. Additionally, he is the Paul and Heidi Brown Preeminent Professor in Industrial and Systems Engineering. He is also an affiliated faculty member of the Computer and Information Science Department, the Hellenic Studies Center, and the Biomedical Engineering Program. He is also the Director of the Center for Applied Optimization. Dr. Pardalos is a world leading expert in global and combinatorial optimization. His recent research interests include network design problems, optimization in telecommunications, e-commerce, data mining, biomedical applications, and massive computing.