Check for updates

Robust Optimization and Data Classification for Characterization of Huntington Disease Onset via Duality Methods

Daniel Woolnough[1] · Niroshan Jeyakumar[2] · Guoyin Li[1] · Clement T Loy[3,4] · Vaithilingam Jeyakumar[1]

© The Author(s) 2021

Abstract

The features that characterize the onset of Huntington disease (HD) are poorly understood yet have significant implications for research and clinical practice. Motivated by the need to address this issue, and the fact that there may be inaccuracies in clinical HD data, we apply robust optimization and duality techniques to study support vector machine (SVM) classifiers in the face of uncertainty in feature data. We present readily numerically solvable semi-definite program reformulations via conic duality for a broad class of robust SVM classification problems under a general spectrahedron uncertainty set that covers the most commonly used uncertainty sets of robust optimization models, such as boxes, balls, and ellipsoids. In the case of the box-uncertainty model, we also provide a new simple quadratic program reformulation, via Lagrangian duality, leading to a very efficient iterative scheme for robust classifiers. Computational results on a range of datasets indicate that these robust classification methods allow for greater classification accuracies than conventional support vector machines in addition to selecting groups of highly correlated features. The conic duality-based robust SVMs were also successfully applied to a new, large HD dataset, achieving classification accuracies of over 95% and providing important information about the features that characterize HD onset.

This paper is dedicated to Franco Giannessi on his 85th birthday.

Communicated by Boris S. Mordukhovich.

✉ Guoyin Li
g.li@unsw.edu.au

Extended author information available on the last page of the article

🍀 Springer

1 Introduction

Support vector machines (SVMs) are optimization-based numerical methods for data classification problems [9,24] that are generally formulated as linear or convex optimization problems. SVMs have become one of the most widely used methods for binary classification, which separates data into two desired groups, and have found applications in numerous fields of science [9], engineering [15] and medicine [12,14,27,46]. These methods are inherently performed in the face of data uncertainty due to the presence of noise in the training data.

In recent years, robust optimization, which was pioneered in the 1970s for treating uncertain linear programming problems, has now emerged as a powerful approach for decision-making in the face of data uncertainty. It treats uncertainty as deterministic but does not limit data values to point estimates. Two decades since the advent of robust optimization, in the late 1990s, Ben-Tal et al. [4,5] provided a highly successful computationally tractable treatment of the robust optimization approach for linear as well as nonlinear optimization problems under data uncertainty [21,28,45].

In this framework, one associates with the uncertain SVM classification problem with feature uncertainty its robust counterpart, where the uncertain constraints are enforced for every possible value of the data within their prescribed uncertainty sets [6,7,29,41]. In this paper, we consider a broad class of robust SVM classification problems under general spectrahedron uncertainty sets [10,42]. The spectrahedron uncertainty set covers the most commonly used uncertainty sets of numerically solvable robust optimization models, such as boxes, balls and ellipsoids [4,6,7]. The robust counterpart, in general, is a hard nonlinear optimization problem with infinitely many constraints, and we reformulate it as a numerically tractable equivalent conic linear program using conic duality [3,4] and a support function technique [3,18]. We show that the robust counterpart reduces to the second-order cone programs for the cases where the uncertainty sets are ellipsoids, balls or boxes. These second-order cone programs can be solved more efficiently.

In the case of the box-uncertainty model, employing Lagrangian duality [17,19], we also provide a new robust SVM classifier by transforming the robust counterpart into a convex quadratic program with non-negative variables, leading to a very efficient computational scheme via a simple iterative algorithm. This approach, which was inspired by the Lagrangian support vector machine developed by Mangasarian et al [16,34], extends the pq-SVM developed by Dunbar et al [14] to robust SVMs. Computational results on a range of datasets indicate that our methods allow for greater classification accuracy than conventional SVMs in addition to selecting smaller groups of highly correlated features.

The conic duality-based robust SVM methods were also applied to a new dataset, Enroll-HD, which contains 36,953 sets of observations on 32 physical features from subjects with, or at risk of, Huntington disease (HD). HD is a neurodegenerative movement disorder with motor (relating to movement), cognitive and psychiatric man-

ifestations caused by an inherited mutation in the *Huntingtin (HTT)* gene [1,11,33]. Characterizing the onset of the disease in subjects harbouring the causative mutation in terms of its associated features is of significant clinical and research importance. Our robust SVM methods also performed well on the Enroll-HD dataset, achieving accuracies of over 95% and selecting meaningful features for classifying subjects as having manifest (post-onset) or non-manifest HD.

The outline of the paper is as follows. Section 2 develops robust SVM data classification models. Section 3 presents equivalent conic program reformulations of these robust classification models for various classes of uncertainty sets. Section 4 provides a robust classification scheme in the case of box uncertainty and gives a simple iterative algorithm to find robust classifiers. Section 5 provides results on the computational experiments on three publicly available datasets. Section 6 describes the Enroll-HD dataset, the performance of conic duality-based robust methods on this dataset and the implications of these results on the characterization of HD onset. Section 7 concludes with brief discussion on further work. The appendix provides additional technical details on spectrahedra, lists the features contained in the Enroll-HD dataset and the proof of (linear) convergence of our iterative algorithm.

2 Robust Optimization-Based Data Classification

In this section, we introduce the SVM formulation and describe the so-called robust SVM formulations. We begin by fixing the notations that will be used later in the paper. Given a vector $x \in \mathbb{R}^n$, $|x|$ denotes the vector consisting of the absolute value of each component x_i for $i = 1, 2, \ldots, n$. The zero vector in $\mathbb{R}^n$ is denoted by 0_n. For a vector $x \in \mathbb{R}^n$, $x \geq 0_n$ if every component, $x_i \geq 0$ for $i = 1, 2, \ldots, n$. The $n \times n$ identity matrix is denoted by I_n (or $I_{n \times n}$). The $n \times n$ matrix of zeros is denoted by $0_{n \times n}$ (or simply 0 if the dimension is clear). The vector of all ones in $\mathbb{R}^n$ is denoted by e_n. We denote by $\mathbb{S}^n$ the space of all real-valued $n \times n$ symmetric matrices. For a vector $x \in \mathbb{R}^n$, the p-norm, for $1 \leq p < \infty$ is defined as: $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}$ with $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$. For convenience, we also write $\|x\|_2 \equiv \|x\|$. For a matrix $A \in \mathbb{R}^{m \times n}$, its norm (or 2-norm) is denoted by $\|A\|$, and is given by $\|A\| = \delta_{max}(A)$ where $\delta_{max}(A)$ is the largest singular value of A. This corresponds to the magnitude of the largest eigenvalue of A, $|\lambda_{max}(A)|$ if $A \in \mathbb{R}^{n \times n}$ and symmetric. For a vector $x \in \mathbb{R}^n$, $\mathrm{diag}(x)$ denotes a diagonal matrix in $\mathbb{R}^{n \times n}$ whose entries consist of the elements of x. The gradient of the scalar function $f : \mathbb{R}^n \to \mathbb{R}$ with respect to the vector x is denoted by $\nabla_x f(x)$. The closed unit ball of $\mathbb{R}^n$ is denoted by $\mathbb{B}_n$.

Consider two sets of data $\mathcal{A}$ and $\mathcal{B}$ whose elements are vectors in $\mathbb{R}^s$. The SVM classifier distinguishes between these two datasets by attempting to separate the m data points into one of two open halfspaces with minimal error —each halfspace containing only those datapoints that correspond to the set $\mathcal{A}$ or $\mathcal{B}$, respectively, where m is the cardinality of $\mathcal{A} \cup \mathcal{B}$. Each datapoint u_i, $i = 1, \ldots, m$, has a corresponding class label $\alpha_i \in \{-1, 1\}$ according to the set $\mathcal{A}$ or $\mathcal{B}$ in which it is contained. The classifier used in the standard (linear) SVM formulation is a hyperplane of the form:

$$u^T w + \gamma = 0, \tag{2.1}$$

where w is the normal to the surface of the hyperplane and γ determines the location of the hyperplane relative to the origin. To construct the SVM classifier, the margin (denoted by $M = 2/\|w\|$) between the planes

$$u^T w + \gamma = +1 \tag{2.2}$$
$$u^T w + \gamma = -1, \tag{2.3}$$

is maximized, subject to the condition that each plane bounds one of the sets (the so-called "hard-margin" case). The optimal classifier lies midway between these two bounding planes.

Often the data are not linearly separable, and so the data cannot be correctly classified by linear bounding hyperplanes. This situation results in the following "soft-margin" SVM formulation with the tuning parameter λ (see [4,9]):

$$(SVM) \quad \min_{(w,\gamma,\xi)\in\mathbb{R}^s\times\mathbb{R}\times\mathbb{R}^m} \lambda\|w\|_2^2 + \sum_{i=1}^m \xi_i$$
$$\text{s.t.} \qquad \xi_i \geq 0, \ \alpha_i(u_i^T w + \gamma) + \xi_i \geq 1, \ i = 1,\ldots,m,$$

where $(u_i, \alpha_i) \in \mathbb{R}^s \times \{-1, 1\}$ are the given training data, α_i is the class label for each data u_i, and the number of nonzero entries in the slack vector ξ is the number of errors the classifier makes on the training data.

The soft-margin classification via a doubly regularized support vector machine (DrSVM) examined in [4, Section 12.1.1][14] can be formulated as:

$$(DrSVM) \quad \min_{(w,\gamma,\xi)\in\mathbb{R}^s\times\mathbb{R}\times\mathbb{R}^m} \lambda_1\|w\|_1 + \lambda_2\|w\|_2^2 + \sum_{i=1}^m \xi_i$$
$$\text{s.t.} \qquad \xi_i \geq 0, \ \alpha_i(u_i^T w + \gamma) + \xi_i \geq 1, \ i = 1,\ldots,m,$$

where $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ are tuning parameters. The $(DrSVM)$ formulation that incorporates 1-norm is known to generate sparse solutions. When a linear classifier is used, solution sparsity implies that the separating hyperplane depends on few input features. This makes the doubly regularized approach a very effective tool for feature selection in classification problems [8,14,16,24].

The soft-margin SVM model (SVM) is a convex quadratic optimization problem with finitely many linear inequality constraints. By introducing auxiliary variables, the doubly regularized support vector machine (DrSVM) can also be equivalently reformulated as a convex quadratic optimization problem with finitely many linear inequality constraints. Noting that the feasible regions of these optimization models are nonempty, the celebrated Frank–Wolfe theorem [2] ensures that the optimal solutions always exist for these two optimization models.

In practice, the given data u_i, $i = 1,\ldots,m$, are often uncertain. We assume that these data are subject to the following spectrahedral data uncertainty parameterized

with the radius parameter $r_i \geq 0$:

$$u_i \in \mathcal{U}_i(r_i) = \bar{u}_i + r_i \mathcal{V}_i,$$

where each $\mathcal{V}_i$, $i = 1, \ldots, m$, is a bounded spectrahedron given by

$$\mathcal{V}_i = \{(v_{i1}, \ldots, v_{is}) \in \mathbb{R}^s : A_0^{(i)} + \sum_{l=1}^{s} v_{il} A_l^{(i)} \succeq 0\},$$

with $A_l^{(i)}$, $l = 1, \ldots, s$, being symmetric $(p \times p)$ matrices. The spectrahedral uncertainty encompasses many important commonly used uncertainty sets such as polyhedral uncertainty sets (where all $A_l^{(i)}$ are diagonal matrices), ball uncertainty sets, ellipsoidal uncertainty sets and their intersections. *We assume that the label α_i is free of uncertainty.*

Let $r = (r_1, \ldots, r_m)$. Then, the robust support vector machine can be stated as:

$$(RSVM_r) \quad \min_{(w,\gamma,\xi) \in \mathbb{R}^s \times \mathbb{R} \times \mathbb{R}^m} \lambda_1 \|w\|_1 + \lambda_2 \|w\|_2^2 + \sum_{i=1}^{m} \xi_i$$

$$\text{s.t.} \quad \xi_i \geq 0, \ \alpha_i(u_i^T w + \gamma) + \xi_i \geq 1, \ \forall u_i \in \mathcal{U}_i(r_i), \ i = 1, \ldots, m.$$

Note that a robust support vector machine model problem is, in general, a semi-infinite convex optimization problem. Note also that an optimal solution exists for $(RSVM_r)$ whenever the robust feasible set F is nonempty where $F = \{(w, \gamma, \xi) : \xi_i \geq 0, \ \alpha_i(u_i^T w + \gamma) + \xi_i \geq 1, \ \forall u_i \in \mathcal{U}_i(r_i), \ i = 1, \ldots, m\}$, and the label sets $\mathcal{I}_\mathcal{A}$ and $\mathcal{I}_\mathcal{B}$ are both nonempty, where $\mathcal{I}_\mathcal{A} = \{1 \leq i \leq m : \alpha_i = 1\}$ and $\mathcal{I}_\mathcal{B} = \{1 \leq i \leq m : \alpha_i = -1\}$. To see this, denote the objective function of $(RSVM_r)$ by f and let the optimal value of $(RSVM_r)$ be $\inf(RSVM_r)$. As the robust feasible set is nonempty and the objective function f is always bounded below by 0, and so, $\inf(RSVM_r)$ is a non-negative real number. Let (w^k, γ^k, ξ^k) be a minimizing sequence, that is, $(w^k, \gamma^k, \xi^k) \in F$ and $f(w^k, \gamma^k, \xi^k) \to \inf(RSVM_r)$. From the definitions of f and the fact that $\xi^k \in \mathbb{R}_+^m$, we see that $\{w^k\}$ and $\{\xi^k\}$ are bounded sequences. Now, $(w^k, \gamma^k, \xi^k) \in F$ shows that $\min_{u_i \in \mathcal{U}_i(r_i)}\{\alpha_i u_i^T w^k\} + \alpha_i \gamma^k + \xi_i^k \geq 1$, $i = 1, \ldots, m$. Take $i \in \mathcal{I}_\mathcal{A}$ and $i' \in \mathcal{I}_\mathcal{B}$. Then, $\gamma^k = \alpha_i \gamma^k$ and $-\gamma^k = \alpha_{i'} \gamma^k$ are both bounded below. So, $\{\gamma^k\}$ is bounded, and hence $\{(w^k, \gamma^k, \xi^k)\}$ is also a bounded sequence in F. As F is a closed set and f is a continuous function, it follows that an optimal solution exists for $(RSVM_r)$.

Figure 1 presents an illustration for both robust and non-robust SVM classifiers. On the left, we see the separating hyperplane and two bounding hyperplanes found by solving a standard (non-robust) SVM; on the right, we see the corresponding hyperplanes found by solving a robust SVM with box uncertainty: $\mathcal{V}_i = \{v \in \mathbb{R}^2 : \|v\|_\infty \leq 1\}$.

In the next section, we will turn our attention to reformulating the robust SVM $(RSVM_r)$ into a numerically tractable optimization problem. Without loss of generality throughout this paper, we always assume that an optimal solution for the robust SVM model $(RSVM_r)$ exists.
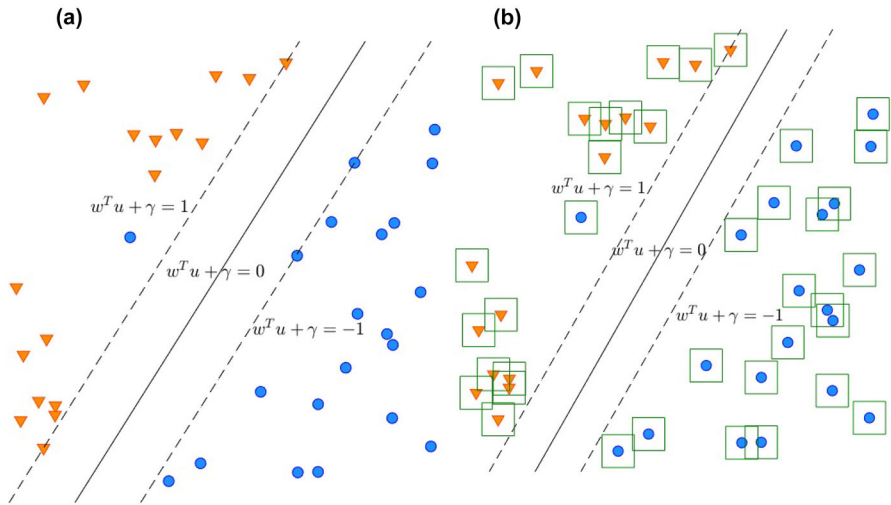
Fig. 1 **a** Classifier determined by the standard SVM. **b** Classifier determined by the robust SVM with box uncertainty. The uncertainty set around each datapoint is shown

3 SDP Formulations for Robust SVM via Conic Duality

In this section, we first show that the robust support vector machine problem under more general spectrahedron uncertainty can be equivalently reformulated as a semi-definite programming problem via a support function technique and conic duality [3,4]. We then derive simple numerically tractable formulations for the cases where the uncertainty sets are ellipsoids, balls and boxes.

We begin by establishing a simple lemma which shows that the robust support vector machine problem is equivalent to a nonsmooth convex optimization problem with finitely many inequality constraints. As we see later in the section, this lemma allows us to easily achieve an equivalent semi-definite programming reformulation for $(RSVM_r)$. To do this, we define the support function of a closed convex and bounded set $C \subset \mathbb{R}^s$ by

$$\sigma_C(x) = \max\{x^T z : z \in C\}.$$

Then, the support function, $\sigma_C(\cdot)$, is a convex function and closed-form formulae for the support function are known for various cases of C, such as balls and boxes. For instance, if $C = \{z \in \mathbb{R}^s : \|z\|_p \leq 1\}$, then $\sigma_C(x) = \|x\|_q$, where $1/p + 1/q = 1$ and $p > 1$. When $p = 1$, $\sigma_C(x) = \|x\|_\infty$. For details, see [3,4].

Consider the following nonsmooth convex optimization problem which we associate with $(RSVM_r)$:

$$(AP_r) \quad \min_{(w,\gamma,t,\mu,\xi) \in \mathbb{R}^s \times \mathbb{R} \times \mathbb{R}^s \times \mathbb{R} \times \mathbb{R}^m} \quad \lambda_1 \sum_{l=1}^{s} t_l + \lambda_2 \mu + \sum_{i=1}^{m} \xi_i$$

$$\text{s.t.} \qquad \|(w, \frac{1-\mu}{2})\|_2 \le \frac{1+\mu}{2}, \ \ w_l \le t_l, \ -w_l \le t_l, \ l = 1, \dots, s$$

$$\xi_i \ge 0, \ \sigma_{\mathcal{U}_i(r_i)}(-\alpha_i w) - \alpha_i \gamma - \xi_i + 1 \le 0, \ i = 1, \dots, m,$$

where for each $i = 1, 2, \dots, m$, the support function $\sigma_{\mathcal{U}_i(r_i)}(\cdot)$ is a convex function.

Lemma 3.1 *(Equivalent nonsmooth convex program) Consider the robust support vector machine problem* $(RSVM_r)$ *with a bounded spectrahedron uncertainty set. Then,*

$$\min(RSVM_r) = \min(AP_r).$$

Moreover, $(w, \gamma, \xi) \in \mathbb{R}^s \times \mathbb{R} \times \mathbb{R}^m$ *is a solution for* $(RSVM_r)$ *if and only if there exist* $t \in \mathbb{R}^s$ *and* $\mu \in \mathbb{R}$ *such that* $(w, \gamma, t, \mu, \xi) \in \mathbb{R}^s \times \mathbb{R} \times \mathbb{R}^s \times \mathbb{R} \times \mathbb{R}^m$ *is a solution for* (AP_r).

Proof The robust SVM problem can be equivalently rewritten as follows:

$$\min_{(w,\gamma,t,\mu,\xi) \in \mathbb{R}^s \times \mathbb{R} \times \mathbb{R}^s \times \mathbb{R} \times \mathbb{R}^m} \lambda_1 \sum_{l=1}^{s} t_l + \lambda_2 \mu + \sum_{i=1}^{m} \xi_i$$

$$\text{s.t.} \qquad \|w\|_2^2 \le \mu, \ \ |w_l| \le t_l, \ l = 1, \dots, s$$

$$\xi_i \ge 0, \ -\alpha_i(w^T u_i + \gamma) - \xi_i + 1 \le 0, \ \forall u_i \in \mathcal{U}_i(r_i), \ i = 1, \dots, m.$$

Note that $|w_l| \le t_l$ is equivalent to $-w_l \le t_l$ and $w_l \le t_l$ for all $l = 1, \dots, s$. Moreover, $\|w\|_2^2 \le \mu$ can be equivalently rewritten in terms of conic constraints as follows.

$$\|w\|_2^2 \le \mu \ \Leftrightarrow \ \|(w, \frac{1-\mu}{2})\|_2 \le \frac{1+\mu}{2}.$$

To finish the proof, we only need to show that for all $i = 1, \dots, m$,

$$\alpha_i(u_i^T w + \gamma) + \xi_i \ge 1, \ \forall u_i \in \mathcal{U}_i(r_i) \ \Leftrightarrow \ \sigma_{\mathcal{U}_i(r_i)}(-\alpha_i w) - \alpha_i \gamma - \xi_i + 1 \le 0. \quad (3.1)$$

To see this, observe that $\alpha_i(u_i^T w + \gamma) + \xi_i \ge 1, \ \forall u_i \in \mathcal{U}_i(r_i)$ is equivalent to the system,

$$1 - \alpha_i u_i^T w - \alpha_i \gamma - \xi_i \le 0, \ \forall u_i \in \mathcal{U}_i(r_i),$$

which is in turn equivalent to

$$0 \ge \max_{u_i \in \mathcal{U}_i(r_i)} \{1 - \alpha_i u_i^T w - \alpha_i \gamma - \xi_i\}$$

$$= 1 - \alpha_i \gamma - \xi_i + \max\{u_i^T(-\alpha_i w) : u_i \in \mathcal{U}_i(r_i)\}$$

$$= 1 - \alpha_i \gamma - \xi_i + \sigma_{\mathcal{U}_i(r_i)}(-\alpha_i w).$$

This means that (3.1) holds, and so, the conclusion follows. $\qquad \square$

We now consider the following semi-definite program which can easily be shown to be equivalent to $(RSVM_r)$ by conic duality [3,4].

$$(SDP_r) \quad \min_{\substack{(w,\gamma,t,\mu,\xi) \in \mathbb{R}^s \times \mathbb{R} \times \mathbb{R}^s \times \mathbb{R} \times \mathbb{R}^m \\ W_i \in \mathbb{S}^p,\ i=1,\ldots,m}} \lambda_1 \sum_{l=1}^{s} t_l + \lambda_2 \mu + \sum_{i=1}^{m} \xi_i$$

$$\text{s.t.} \quad \|(w, \frac{1-\mu}{2})\|_2 \leq \frac{1+\mu}{2}, \quad w_l \leq t_l, \quad -w_l \leq t_l, \ l = 1,\ldots,s$$

$$W_i \succeq 0, \quad -\alpha_i(w^T \overline{u}_i + \gamma)$$

$$-\xi_i + 1 + \text{Trace}(W_i A_0^{(i)}) \leq 0, \ i = 1,\ldots,m$$

$$-r_i \alpha_i w_l + \text{Trace}(W_i A_l^{(i)}) = 0, \ l = 1,\ldots,s, \ i = 1,\ldots,m$$

$$\xi_i \geq 0, \ i = 1,\ldots,m$$

Theorem 3.1 *(Spectrahedral uncertainty: semi-definite program) Consider the robust support vector machine problem $(RSVM_r)$ with spectrahedral uncertainty, and its associated semi-definite program problem (SDP_r). Suppose that for each $i = 1,\ldots,m$, the interior of the spectrahedron $\mathcal{V}_i$ is nonempty, i.e., there exists $\overline{v}_i = (\overline{v}_{i1},\ldots,\overline{v}_{is}) \in \mathbb{R}^s$ such that $A_0^{(i)} + \sum_{l=1}^{p} \overline{v}_{il} A_l^{(i)} \succ 0, i = 1,\ldots,m$. Then,*

$$\min(RSVM_r) = \min(SDP_r).$$

Moreover, (w,γ,ξ) is a solution of $(RSVM_r)$ if and only if there exist $W_i \succeq 0$, $i = 1,\ldots,m$, $t \in \mathbb{R}^s$ and $\mu \in \mathbb{R}$ such that $(w,\gamma,t,\mu,\xi,W_1,\ldots,W_m)$ is a solution of (SDP_r).

Proof By lemma 3.1, the robust SVM problem is equivalent to

$$(AP_r) \quad \min_{(w,\gamma,t,\mu,\xi) \in \mathbb{R}^s \times \mathbb{R} \times \mathbb{R}^s \times \mathbb{R} \times \mathbb{R}^m} \lambda_1 \sum_{l=1}^{s} t_l + \lambda_2 \mu + \sum_{i=1}^{m} \xi_i$$

$$\text{s.t.} \quad \|(w, \frac{1-\mu}{2})\|_2 \leq \frac{1+\mu}{2}, \quad w_l \leq t_l,$$

$$-w_l \leq t_l, \ l = 1,\ldots,s$$

$$\sigma_{\mathcal{U}_i(r_i)}(-\alpha_i w) - \alpha_i \gamma - \xi_i + 1 \leq 0, \ i = 1,\ldots,m$$

$$\xi_i \geq 0, \ i = 1,\ldots,m.$$

Recall that $\mathcal{U}_i(r_i) = \overline{u}_i + r_i \mathcal{V}_i$ where $\mathcal{V}_i = \{(v_{i1},\ldots,v_{is}) \in \mathbb{R}^s : A_0^{(i)} + \sum_{l=1}^{s} v_{il} A_l^{(i)} \succeq 0\}$, with $A_l^{(i)}$, $l = 1,\ldots,s$, being symmetric $(p \times p)$ matrices. To see the conclusion, it suffices to show that for each $i = 1,\ldots,m$,

$$\sigma_{\mathcal{U}_i(r_i)}(-\alpha_i w) - \alpha_i \gamma - \xi_i + 1 \leq 0 \tag{3.2}$$

is equivalent to the existence of $W_i \succeq 0$ such that

$$\begin{cases} -\alpha_i(w^T \overline{u}_i + \gamma) - \xi_i + 1 + \text{Trace}(W_i A_0^{(i)}) \leq 0 \\ -r_i \alpha_i w_l + \text{Trace}(W_i A_l^{(i)}) = 0, \ l = 1,\ldots,s. \end{cases} \tag{3.3}$$

Suppose that (3.2) holds. Then, the following implication holds:

$$A_0^{(i)} + \sum_{l=1}^{s} v_{il} A_l^{(i)} \succeq 0 \implies (-\alpha_i w)^T (\overline{u}_i + r_i v_i) - \alpha_i \gamma - \xi_i + 1 \leq 0. \quad (3.4)$$

As the interior point condition holds, it follows from the conic duality theorem [3,4] that there exists $W_i \succeq 0$ such that

$$(-\alpha_i w)^T (\overline{u}_i + r_i v_i) - \alpha_i \gamma - \xi_i + 1 + \text{Trace}\Big(W_i (A_0^{(i)} + \sum_{l=1}^{s} v_{il} A_l^{(i)})\Big) \leq 0,$$

$$\forall\, v_i = (v_{i1}, \ldots, v_{is}) \in \mathbb{R}^s. \quad (3.5)$$

Note that the validity of the affine inequality, $a^T x + b \leq 0$ for all $x \in \mathbb{R}^s$, with $a \in \mathbb{R}^s$ and $b \in \mathbb{R}$, means that $a = 0_s$ and $b \leq 0$. Thus, we see that (3.5) is equivalent to (3.3).

Conversely, suppose that for each $i = 1, \ldots, m$, there exists $W_i \succeq 0$ such that (3.3) holds. Then, (3.5) holds. Consequently,

$$\sigma_{\mathcal{U}_i(r_i)}(-\alpha_i w) - \alpha_i \gamma - \xi_i + 1 = \max_{v_i \in \mathcal{V}_i} (-\alpha_i w)^T (\overline{u}_i + r_i v_i) - \alpha_i \gamma - \xi_i + 1$$

$$\leq \max_{v_i \in \mathcal{V}_i} \{-\text{Trace}\Big(W_i (A_0^{(i)} + \sum_{l=1}^{s} v_{il} A_l^{(i)})\Big)\}$$

$$\leq 0,$$

where the last inequality follows from $W_i \succeq 0$ and $v_i \in \mathcal{V}_i$ (and so, $A_0^{(i)} + \sum_{l=1}^{s} v_{il} A_l^{(i)} \succeq 0$). So, (3.2) holds.

Hence, (3.2) is equivalent to the existence of $W_i \succeq 0$ such that (3.3) holds and the conclusion follows. □

We now derive numerically tractable formulations for $(RSVM_r)$ in terms of the second-order cone programs, under the uncertainty sets that take the form of an ellipsoid, ball or box. Although these equivalent formulations and the associated duality results may be derived from (SDP_r) and Theorem 3.1, respectively, by appropriately choosing the matrices $A_l^{(i)}$, $l = 1, \ldots, s$, $i = 1, \ldots, m$, of the spectrahedron $\mathcal{V}_i$, in the interest of simplicity, we present the results from the model (AP_r) and Lemma 3.1. Related special cases for the standard robust SVM models, where $\lambda_1 = 0$ or $\lambda_2 = 0$, can be found in [4,6].

Ellipsoidal Uncertainty Consider the case where the uncertainty sets $\mathcal{V}_i$ are ellipsoids in the sense that

$$\mathcal{V}_i = \{v_i \in \mathbb{R}^s : v_i^T M_i^{-1} v_i \leq 1\} \quad (3.6)$$

for some $M_i \succ 0$. Let $M_i = L_i L_i^T$ with L_i being an invertible matrix. We associate with this case the following second-order cone program:

$$(SOCP_{r,E}) \quad \min_{(w,\gamma,\xi,t,\mu)\in\, \mathbb{R}^s \times \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}} \lambda_1 \sum_{j=1}^{s} t_j + \lambda_2 \mu + \sum_{i=1}^{m} \xi_i$$

$$\text{subject to} \qquad \|(w, \frac{1-\mu}{2})\|_2 \leq \frac{1+\mu}{2}, \ w_j \leq t_j,$$
$$-w_j \leq t_j, \ j = 1, \ldots, s$$
$$-\alpha_i(w^T \bar{u}_i + \gamma) - \xi_i + 1$$
$$+r_i \|L_i^T w\|_2 \leq 0, \ i = 1, \ldots, m$$
$$\xi_i \geq 0, \ i = 1, \ldots, m$$

Proposition 3.1 *(Ellipsoidal uncertainty: second-order cone program) For the robust support vector machine problem ($RSVM_r$) under ellipsoidal uncertainty, as defined in (3.6), and its associated second-order cone problem ($SOCP_{r,E}$), it holds that*

$$\min(RSVM_r) = \min(SOCP_{r,E}).$$

Moreover, (w, γ, ξ) is a solution of ($RSVM_r$) under ellipsoidal uncertainty if and only if there exists $t \in \mathbb{R}^s$ and $\mu \in \mathbb{R}$ such that (w, γ, ξ, t, μ) is a solution of ($SOCP_{r,E}$).

Proof In the case of ellipsoidal uncertainty as defined in (3.6), the support function $\sigma_{\mathcal{U}_i(r_i)}(-\alpha_i w)$ can be expressed as:

$$\begin{aligned}
\sigma_{\mathcal{U}_i(r_i)}(-\alpha_i w) &= -\alpha_i w^T \bar{u}_i + r_i \max\{(-\alpha_i w)^T v_i : v_i^T M_i^{-1} v_i \leq 1\} \\
&= -\alpha_i w^T \bar{u}_i + r_i \max\{(-\alpha_i w)^T v_i : \|L_i^{-1} v_i\|_2 \leq 1\} \\
&= -\alpha_i w^T \bar{u}_i + r_i \max\{(-\alpha_i)(L_i^T w)^T z_i : \|z_i\|_2 \leq 1\} \\
&= -\alpha_i w^T \bar{u}_i + r_i \|(-\alpha_i) L_i^T w\|_2 \\
&= -\alpha_i w^T \bar{u}_i + r_i \|L_i^T w\|_2,
\end{aligned}$$

where the last two equalities follow from the support function formula and $\alpha_i \in \{-1, 1\}$, respectively. Thus, the conclusion follows from Lemma 3.1. □

Ball Uncertainty We now consider the case where the perturbation sets $\mathcal{V}_i$ are unit balls:

$$\mathcal{V}_i = \{v_i \in \mathbb{R}^s : \|v_i\|_2^2 \leq 1\} \qquad (3.7)$$

In this case, we consider the following second-order cone program:

$$(SOCP_{r,B}) \quad \min_{(w,\gamma,\xi,t,\mu) \in \mathbb{R}^s \times \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}} \lambda_1 \sum_{l=1}^{s} t_l + \lambda_2 \mu + \sum_{i=1}^{m} \xi_i$$

$$\text{subject to} \qquad \|(w, \frac{1-\mu}{2})\|_2 \leq \frac{1+\mu}{2}, \ w_l \leq t_l,$$
$$-w_l \leq t_l, \ l = l, \ldots, s$$
$$-\alpha_i(w^T \bar{u}_i + \gamma) - \xi_i + 1$$
$$+r_i \|w\|_2 \leq 0, \ i = 1, \ldots, m$$
$$\xi_i \geq 0, \ i = 1, \ldots, m.$$

Corollary 3.1 *(Ball uncertainty: simple second-order cone program) For the robust support vector machine problem ($RSVM_r$) under ball uncertainty, as defined in (3.7), and its associated second-order cone problem ($SOCP_{r,B}$), it holds that*

$$\min(RSVM_r) = \min(SOCP_{r,B}).$$

Moreover, (w, γ, ξ) is a solution of ($RSVM_r$) under ball uncertainty if and only there exist $t \in \mathbb{R}^s$ and $\mu \in \mathbb{R}$ such that (w, γ, ξ, t, μ) is a solution of ($SOCP_{r,B}$).

Proof The result follows immediately from Proposition 3.1, since $\|v_i\|_2^2 = v_i^T I_s^{-1} v_i$, and so $L_i = I_s$. ☐

Box Uncertainty Finally, we consider the case where the perturbation sets $\mathcal{V}_i$ are unit boxes:

$$\mathcal{V}_i = \{v_i \in \mathbb{R}^s : \|v_i\|_\infty \le 1\} \tag{3.8}$$

We associate with this case the following second-order cone program (see [6]):

$$
\begin{aligned}
(SOCP_{r,\infty}) \quad \min_{(w,\gamma,\xi,t,\mu) \,\in\, \mathbb{R}^s \times \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}} \quad & \lambda_1 \sum_{l=1}^{s} t_l + \lambda_2 \mu + \sum_{i=1}^{m} \xi_i \\
\text{subject to} \quad & \|(w, \frac{1-\mu}{2})\|_2 \le \frac{1+\mu}{2}, \ w_l \le t_l, \\
& -w_l \le t_l, \ l = 1, \ldots, s \\
& -\alpha_i(w^T \overline{u}_i + \gamma) - \xi_i + 1 \\
& + r_i \|w\|_1 \le 0, \ i = 1, \ldots, m \\
& \xi_i \ge 0, \ i = 1, \ldots, m.
\end{aligned}
$$

Proposition 3.2 *(**Robust SVM under Box Uncertainty**) For the robust support vector machine problem ($RSVM_r$) under box uncertainty, as defined in (3.8), and its associated second-order cone program problem ($SOCP_{r,\infty}$), it holds that*

$$\min(RSVM_r) = \min(SOCP_{r,\infty}).$$

Moreover, (w, γ, ξ) is a solution of ($RSVM_r$) under polytope uncertainty if and only if there exist $t \in \mathbb{R}^s$ and $\mu \in \mathbb{R}$ such that (w, γ, ξ, t, μ) is a solution of ($SOCP_{r,\infty}$).

Proof In the case of box uncertainty (that is, $\mathcal{U}_i(r_i) = \overline{u}_i + r_i \mathcal{V}_i$ with $\mathcal{V}_i = \{v_i \in \mathbb{R}^s : \|v_i\|_\infty \le 1\}$), the support function $\sigma_{\mathcal{U}_i(r_i)}(-\alpha_i w)$ can be expressed as:

$$
\begin{aligned}
\sigma_{\mathcal{U}_i(r_i)}(-\alpha_i w) &= -\alpha_i w^T \overline{u}_i + r_i \max\{(-\alpha_i w)^T v_i : \|v_i\|_\infty \le 1\} \\
&= -\alpha_i w^T \overline{u}_i + r_i \|(-\alpha_i)w\|_1 \\
&= -\alpha_i w^T \overline{u}_i + r_i \|w\|_1.
\end{aligned}
$$

Thus, the conclusion follows from Lemma 3.1. ☐

4 A New Robust *pq*-SVM for Efficient Classification

In this section, we derive an efficient scheme for finding a robust classifier under *box uncertainty* by extending the approach in [14,16,27] and using a variable transformation and Lagrangian duality [18] to reformulate the robust SVM model ($SOCP_{r,\infty}$) into a simple non-negative quadratic program.

4.1 QP Reformulation via Lagrangian Duality

Recall that the problem ($SOCP_{r,\infty}$) can be equivalently rewritten as:

$$
\min_{(w,\gamma,\xi)\in\mathbb{R}^s\times\mathbb{R}\times\mathbb{R}^m} \lambda_1\|w\|_1 + \lambda_2\|w\|_2^2 + e_m^T\xi
$$
$$
\text{subject to} \quad \alpha_i(\bar{u}_i^T w + \gamma) + \xi_i - r_i\|w\|_1 \geq 1, \, i = 1, 2, \ldots, m
$$
$$
\xi \geq 0_m. \tag{4.1}
$$

Define vectors $p, q \in \mathbb{R}_+^s$ by

$$
p_i = \begin{cases} 0, & \text{for } w_i \leq 0 \\ w_i, & \text{for } w_i > 0, \end{cases} \quad q_i = \begin{cases} -w_i, & \text{for } w_i \leq 0 \\ 0, & \text{for } w_i > 0, \end{cases} \tag{4.2}
$$

for $i = 1, \ldots, s$. Then, it is easy to see that

$$
w = p - q, \quad p, q \geq 0_s \text{ and } p^T q = 0.
$$

Consequently, we can rewrite $\|w\|_2^2$ and $\|w\|_1$ as $\|w\|_2^2 = \|p\|_2^2 + \|q\|_2^2$ and $\|w\|_1 = e_s^T(p+q)$. So, the problem ($SOCP_{r,\infty}$) can be defined as:

$$
\min_{(p,q,\gamma,\xi)\in\mathbb{R}^s\times\mathbb{R}^s\times\mathbb{R}\times\mathbb{R}^m} \lambda_1 e_s^T(p+q) + \lambda_2\|p\|_2^2 + \lambda_2\|q\|_2^2 + e_m^T\xi
$$
$$
\text{subject to} \quad \alpha_i(\bar{u}_i^T(p-q) + \gamma) + \xi_i - r_i e_s^T(p+q) \geq 1, \, i = 1, 2, \ldots, m
$$
$$
p, q \geq 0_s, \, \xi \geq 0_m.
$$

Now, we define:

$$
y = \begin{bmatrix} \xi \\ p \\ q \end{bmatrix} \in \mathbb{R}^{m+2s}, \quad C = \begin{bmatrix} 0_{m\times m} & & \\ & \lambda_2 I_s & \\ & & \lambda_2 I_s \end{bmatrix}, \text{ and } b = \begin{bmatrix} \frac{1}{\lambda_1}e_m \\ e_s \\ e_s \end{bmatrix}.
$$

Let

$$
\hat{D} = \begin{bmatrix} I_m & 0_{m\times 2s} \\ 0_{2s\times m} & 0_{2s\times 2s} \end{bmatrix} \tag{4.3}
$$

and

$$\hat{U} = \begin{bmatrix} 0_{m\times m} & U_1 & U_2 \\ 0_{2s\times m} & 0_{2s\times s} & 0_{2s\times s} \end{bmatrix} \tag{4.4}$$

where

$$U_1 = \begin{bmatrix} \alpha_1\bar{u}_1^T - r_1 e_s^T \\ \vdots \\ \alpha_m\bar{u}_m^T - r_m e_s^T \end{bmatrix} \in \mathbb{R}^{m\times s} \text{ and } U_2 = \begin{bmatrix} -\alpha_1\bar{u}_1^T - r_1 e_s^T \\ \vdots \\ -\alpha_m\bar{u}_m^T - r_m e_s^T \end{bmatrix} \in \mathbb{R}^{m\times s}.$$

Define further that

$$A = \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_m \end{bmatrix} \in \mathbb{R}^{m\times m}, \quad \hat{A} = \begin{bmatrix} A^{-1} & \\ & 0_{2s\times 2s} \end{bmatrix} \in \mathbb{R}^{(m+2s)\times(m+2s)} \text{ and }$$

$$\hat{e} = \begin{bmatrix} Ae_m \\ 0_s \\ 0_s \end{bmatrix} \in \mathbb{R}^{m+2s}.$$

Then, problem $(SOCP_{r,\infty})$ can be rewritten as the following convex quadratic programming problem:

$$\min_{(y,\gamma)\in\mathbb{R}^{m+2s}\times\mathbb{R}} y^T C y + \lambda_1 b^T y$$
$$\text{subject to} \quad \hat{D}(\hat{U}y + \hat{e}\gamma) + y \geq \hat{A}\hat{e}$$
$$y \geq 0_{m+2s},$$

where $b \geq 0$. By removing the linear term of the objective function via regularization, as in [14,16,27] and by regularizing γ, we arrive at the regularized problem

$$\min_{(y,\gamma)\in\mathbb{R}^{m+2s}\times\mathbb{R}} \frac{1}{2}y^T C y + \frac{1}{2}\gamma^2 + \frac{\nu}{2}y^T y$$
$$\text{subject to} \quad \hat{D}(\hat{U}y + \hat{e}\gamma) + y \geq \hat{A}\hat{e},$$

where $\nu > 0$. Note that the regularization makes the non-negativity condition $y \geq 0_{m+2s}$ redundant. The Lagrangian dual of the regularized problem is given by

$$\max_{(y,\gamma,u)\in\mathbb{R}^{m+2s}\times\mathbb{R}\times\mathbb{R}^{m+2s}} \mathcal{L}(y,\gamma,u) \text{ subject to } \nabla_{y,\gamma}\mathcal{L}(y,\gamma,u) = 0, \ u \geq 0_{m+2s},$$

where

$$\mathcal{L}(y,\gamma,u) = \frac{1}{2}y^T(C + \nu I_{m+2s})y + \frac{1}{2}\gamma^2 - u^T(\hat{D}(\hat{U}y + \hat{e}\gamma) + y - \hat{A}\hat{e}),$$

and

$$\nabla_y \mathcal{L}(y, \gamma, u) = (C + \nu I_{m+2s})y - ((\hat{D}\hat{U})^T + I_{m+2s})u \implies y$$
$$= (C + \nu I_{m+2s})^{-1}(\hat{D}\hat{U} + I_{m+2s})^T u \qquad (4.5)$$
$$\nabla_\gamma \mathcal{L}(y, \gamma, u) = \gamma - (\hat{D}\hat{e})^T u = 0 \implies \gamma = \hat{e}^T \hat{D}^T u. \qquad (4.6)$$

Eliminating the equality constraints, the dual can be written as:

$$\min_{u \in \mathbb{R}^{m+2s}} \quad \frac{1}{2} u^T \left((\hat{D}\hat{U} + I)(C + \nu I_{m+2s})^{-1}(\hat{D}\hat{U} + I_{m+2s})^T + \hat{D}\hat{e}\hat{e}^T \hat{D}^T \right) u - (\hat{A}\hat{e})^T u$$
$$\text{subject to} \quad u \geq 0_{m+2s}.$$

Define the matrix $Q \in \mathbb{R}^{(m+2s) \times (m+2s)}$ and the vector $\eta \in \mathbb{R}^{m+2s}$ by

$$Q = (\hat{D}\hat{U} + I_{m+2s})(C + \nu I_{m+2s})^{-1}(\hat{D}\hat{U} + I_{m+2s})^T + \hat{D}\hat{e}\hat{e}^T \hat{D}, \quad \eta = \hat{A}\hat{e}. \quad (4.7)$$

Then, a direct verification shows that $(\hat{D}\hat{U} + I_{m+2s})^T d = 0_{m+2s}$ if and only if $d = 0_{m+2s}$, and so, Q is positive definite. So, we arrive at a simple strictly convex quadratic programming problem over non-negative orthant:

$$(QP) \quad \min_{u \in \mathbb{R}^{m+2s}} \frac{1}{2} u^T Q u - \eta^T u \text{ subject to } u \geq 0_{m+2s}. \qquad (4.8)$$

Notice that (QP) no longer has the hyperparameter λ_1 in its formulation. Having found a solution u of (4.8), we can then retrieve a solution to our original problem, via the dual equality constraints (4.5) and (4.6), i.e.,

$$y = (C + \nu I_{m+2s})^{-1}(\hat{D}\hat{U} + I_{m+2s})^T u, \quad p = \begin{bmatrix} y_{m+1} \\ \vdots \\ y_{m+s} \end{bmatrix}, \quad q = \begin{bmatrix} y_{m+s+1} \\ \vdots \\ y_{m+2s} \end{bmatrix},$$

$$w = p - q, \quad \gamma = \hat{e}^T \hat{D}u.$$

4.2 Efficient Iterative Scheme

To solve (QP), we propose a variation of the **LSVM algorithm** put forth in [34]. Recall from [34] that the point u is an optimal solution for (QP) if and only if

$$0_{m+2s} \leq u \perp (Qu - \eta) \geq 0_{m+2s}.$$

Notice that

$$0_{m+2s} \leq a \perp b \geq 0_{m+2s}, \ a, b \in \mathbb{R}^{m+2s} \iff \exists \alpha > 0 \text{ such that } a = \max(a - \alpha b, 0_{m+2s}),$$
$$(4.9)$$

where the kth-component of $\max(z, 0_{m+2s}) = \max(z_k, 0), k = 1, \ldots, m+2s$. Unlike the **LSVM Algorithm** where it is taken that $a = Qu - \eta$ and $b = u$, we propose instead to take $a = u$ and $b = Qu - \eta$. Therefore, we arrive at the optimality condition

$$u = \max(u - \alpha(Qu - \eta), 0_{m+2s}), \ \alpha > 0$$

from which we derive the simple iterative scheme:

$$u^{(i+1)} = \max(u^{(i)} - \alpha(Qu^{(i)} - \eta), 0_{m+2s}), \ i = 0, 1, \ldots \qquad (4.10)$$

with a starting point given by $u^{(0)} = Q^{-1}\eta$.

The difference between our proposed iterative scheme and the **LSVM Algorithm** is that we require only the inversion of the matrix Q once, to define the starting point for the iteration, whereas the **LSVM Algorithm** requires solving a linear system in Q at each step of the iteration. We also only require a single matrix multiplication per iteration.

The complete iterative algorithm required to return the optimal solution u^* of (QP) is given below, and its proof of (linear) convergence is given in Appendix C.

Algorithm 1: Box-pq-SVM Algorithm

$u \leftarrow Q^{-1}\eta$;
it $\leftarrow 0$;
$u_{\text{prev}} \leftarrow u + \eta$;
while it $<$ maxiter and $\|u_{\text{prev}} - u\| >$ tol **do**
    $u_{\text{prev}} \leftarrow u$;
    $u \leftarrow \max(u - \alpha(Qu - \eta), 0_{m+2s})$;
    it $\leftarrow$ it $+1$;
end
return u

Notice that maxiter is a threshold for the maximum number of iterations, tol is for the convergence tolerance of the algorithm, and $\alpha > 0$ is a pre-selected constant.

5 Experiments with Real-World Data Sets

In this section, we evaluate several different SVM models, derived from our above results, against some real-world datasets, which are all available from the UCI Machine Learning repository [13]. The aim is to compare the models in terms of accuracy, computational expense and feature selection.

5.1 Experimental Setup

Three datasets are used for the comparison of our models:

- The Wisconsin Breast Cancer (Diagnostic) (WBCD) [13]: this dataset describes the (binary) classification problem of labelling tumours as either malignant or

benign. The dataset contains 569 instances, 30 features and a 63/37 split of the two classes.

- Cylinder Bands (CYLBN) [13]: this dataset describes the problem of mitigating delays known as "cylinder bands" in rotogravure printing. It contains 541 instances, 33 features and a 58/42 class split.
- Pima Indians Diabetes (PIMA) [39]: this dataset describes the problem of classifying diabetes in patients. It contains 768 instances, 8 features and a 35/65 class split.

Each of these datasets had their features standardized and then split into 80%/20% training and test sets. For each of our models, we then perform the following. We tune the model hyperparameters in their cross-validation range via fivefold cross-validation on the training set. The model is then fitted to the full training set, to obtain optimal solution w^*, γ^*, and the training accuracy is recorded. We next determine the features selected by the model. This is done by considering the significance (weighting) of each feature in w^*. More precisely, feature k is considered significant by a model if

$$\left| \frac{w_k^*}{\max\limits_{1 \le j \le s} w_j^*} \right| > 0.05. \tag{5.1}$$

Having done this, we set the value of w_k^* to zero for each insignificant feature. Note that this is a stricter version of the feature selection methods employed in [14,27]. Finally, the model predicts the classification of all datapoints u_i in the test set: $\alpha_i := \text{sign}(w^T u_i + \gamma)$, and we record the test accuracy.

5.2 Classification Methods

We will apply the following classification methods to each of the datasets. For robust models, we will assume that the radius of robustness is constant for each datapoint: $r_i = r, i = 1, \ldots, m$.

- The standard SVM model (SVM). This method does not consider uncertainty in the datapoints. We refer to this method as **SVM**.
- The ball uncertainty robust SVM model ($SOCP_{r,B}$). We refer to this method as **Ball-SVM**.
- The box-uncertainty robust SVM model ($SOCP_{r,\infty}$). We refer to this method as **Box-SVM**.
- The pq-robust SVM over box-uncertainty (QP), which we solved by **Algorithm 1**. We set `maxiter` $= 1000$, `tol` $= 10^{-4}$ and α tuned from 10^{-16} to 10^{-4}. We refer to this method as **Box-pq-SVM**.

For our experiments, the cross-validation range for each hyperparameter was determined as follows. For the first three methods, λ_1 and λ_2 were tuned over values 2^k, $k \in \{-10, \ldots, 4\}$. (Note that **SVM** does not tune the parameter λ_1 which is set to 0.) For the pq-SVM, ν and λ_2 were tuned over the range 2^j, $j \in \{-10, 9, \ldots, 10\}$. This is due to the sensitivity of the pq-SVM to hyperparameters.

For all of our robust methods (**Ball-SVM**, **Box-SVM**, and **Box-pq-SVM**), the radius of robustness was tuned as follows. We set a lower bound of 2^{-20}, and an upper bound given by a simple heuristic: for each datapoint in one class, we calculate the maximum distance from it to any point in the other class. We then take the minimum of these distances and finally use as our upper bound the maximum over the two classes. Formally, we can define our upper bound U as:

$$U = \max \left\{ \min_{u \in \mathcal{A}} \max_{v \in \mathcal{B}} \|u - v\|_2, \min_{v \in \mathcal{B}} \max_{u \in \mathcal{A}} \|u - v\|_2 \right\}.$$

Having obtained our upper bound on the radius, we then tuned our radius of robustness within an exponential range over this lower and upper bound.

Our choice of heuristic for the upper bound is justified as follows: for any radius of robustness larger than this value, consider a (non-trivial) true classifier. It is certain that every datapoint would simultaneously have a point in its uncertainty set on one side of the classifier, and another point in its uncertainty set on the other side. In this case, it is clear that the robust SVM will select a trivial solution, i.e., a majority class prediction.

5.3 Results

All computations were performed using a 3.2GHz Intel(R) Core(TM) i7-8700 and 16GB of RAM, equipped with MATLAB R2019B. All optimization problems were solved via the MOSEK software [36], handled through the YALMIP interface [31].

Table 1 shows the results for each of the classification methods on each of the datasets. The columns are interpreted as follows:

- Dataset: as written.
- Instances: number of instances in the dataset.
- Model: the classification method used.
- Type: either nominal (non-robust) or robust.
- Train Acc: Accuracy obtained on the training set, as a percentage.
- Test Acc: Accuracy obtained on the test set, as a percentage.
- Fit Time (s): CPU time (seconds) taken to solve the final optimization problem, after cross validation.
- Features: the number of selected features from each model, out of the total number of features in the dataset.

5.4 Discussion

Overall, it is apparent from Table 1 that on all three datasets, for both testing accuracy and number of selected features, the standard SVM is outperformed by both Box-SVM and Ball-SVM robust methods.

Box-SVM–Best performing robust method Regarding best performance in terms of testing accuracy, this method performs best in this area, producing the highest testing

Table 1 Results for each classification method on each dataset

Dataset	Instances	Model	Type	Train Acc. (%)	Test Acc. (%)	Fit Time (s)	Features
WBCD	569	**SVM**	Nominal	**99.34**	92.11	0.09	28/30
		Ball-SVM	Robust	98.46	92.98	1.13	27/30
		Box-SVM	Robust	98.68	**95.61**	2.16	**21/30**
		Box-pq-SVM	Robust	98.90	94.74	**0.03**	29/30
CYLBN	541	**SVM**	Nominal	79.40	66.97	0.11	31/33
		Ball-SVM	Robust	79.63	68.81	2.33	30/33
		Box-SVM	Robust	**79.86**	**70.64**	2.32	**25/33**
		Box-pq-SVM	Robust	73.38	69.72	**0.04**	27/33
PIMA	768	**SVM**	Nominal	**77.20**	77.27	0.11	7/8
		Ball-SVM	Robust	77.04	**78.57**	1.49	7/8
		Box-SVM	Robust	**77.20**	**78.57**	2.15	6/8
		Box-pq-SVM	Robust	76.71	75.97	**0.02**	**3/8**

Columns defined above. The best result(s) for each dataset in each category is given in bold

accuracy on all three datasets. It is also consistently economical in its selection of features. We see the method as the most consistent of the four.

Box-pq-SVM–Faster robust method This method is remarkably efficient for reasonable sized datasets, and much faster than our other two robust methods.

Robust methods in applications Regarding the choice of which of the three robust methods to utilize in an application, one should consider which subset of accuracy, feature selection and computational time is most desired. No one of our robust methods outperforms any other on all three counts. We do note that the **Box-pq-SVM** method requires storage of an $(m + 2s) \times (m + 2s)$ size matrix which, even if defined sparsely, can consume significant amounts of physical memory in computation for very large datasets.

Finally, we can also compare our results to others in the recent literature on robust optimization methods for SVMs, namely [6]. Our methods achieve near identical accuracies compared on WBCD and PIMA to the feature-robust method presented in [6] (which also is designed for box-uncertainty), with a slightly higher accuracy achieved by our methods on PIMA.

We do note that whilst the heuristic we use for defining an upper bound on the radius of robustness tuning range is simple, it is both effective and computationally efficient even for very large datasets. Characterizing an upper bound on the radius within a mathematical framework would need to be investigated in a further study.

6 Application to Characterization of Huntington Disease Onset

HD is a neurodegenerative movement disorder caused by a mutation in the *HTT* gene with motor, cognitive and psychiatric manifestations [1,11,33]. It is one of the most common monogenic neurological diseases in the developed world [1]. From its onset,

typically in the fourth decade of life, the signs and symptoms of HD progress inexorably until certain death. No treatment currently available can alter this course.

Defining the onset of HD is of significant clinical and research importance. Studies tracking the progression of the motor manifestations of HD over time have shown that there is an acceleration in motor decline around the time of onset diagnosis [30,32]. Thus, making a clinical diagnosis of HD onset heralds a poor prognosis for the patient and, in addition to carrying significant emotional weight, can have important implications for key life decisions such as family planning. Secondly, clearly defining disease onset is paramount for establishing endpoints in clinical trials where the efficacy of putative disease-modifying therapies in the future could be measured by their ability to delay onset in subjects harbouring the *HTT* mutation [25].

The current formal diagnosis of HD onset or "manifest HD" (mHD) is based on the motor manifestations of the disease according to the motor component of the Unified Huntington Disease Rating Scale (UHDRS). The UHDRS has been shown to have a high degree of internal consistency and interrater reliability [26]. It contains 31 items relating to characteristic motor abnormalities in HD, which are scored from 0 (normal) to 4 (significantly abnormal), and their sum, the total motor score (TMS). Based on these scores, the clinician assigns a "diagnostic confidence level" (DCL) between 0 (normal) and 4 ($\geq$ 99% confidence that the motor abnormalities are unequivocal signs of HD), representing their confidence that the subject's motor abnormalities are due to mHD. A diagnosis of mHD is made when the DCL is 4 [33]. However, there are currently no rules relating the scores from the 31 items and the TMS to the DCL and so the DCL rating relies on the clinician's expertise. Given the significance of making a diagnosis of HD onset, a standardized and objective means of arriving at the diagnosis from the motor assessment is needed [11,33].

6.1 Conic duality Methods

We applied the conic duality-based robust SVM methods to data from the Enroll-HD study consisting of 36,953 sets of motor scores and corresponding DCLs from patients with HD around the world. Subjects with a DCL of 4 were defined as having mHD, whilst those with a DCL less than 4 (i.e. 0, 1, 2 and 3) were defined as having non-manifest HD (nHD). There were 19303 (52%) cases of mHD and 17650 (48%) cases of nHD defined in this way, a roughly even split between the two groups, minimizing classification bias. Data from the 31 motor items from the UHDRS and the TMS (32 features in total) for these subjects were used as potential features to predict classification by conic duality-based robust SVM models.

Due to the physical memory limitations of our hardware, and the size of the dataset, **Box-pq-SVM** was not applied to the Enroll-HD dataset.

6.2 Results

Both box and ball robust SVM models achieved similarly high accuracies of over 95% in both the training and testing phases (Table 2).

Table 2 Results for each classification method on Enroll-HD dataset

Dataset	Model	Train acc. (%)	Test acc. (%)	Features
Enroll-HD	**Ball-SVM**	95.55	95.28	30/32
	Box-SVM	95.54	95.21	22/32

Table 3 Motor features selected by the classifier found by each method

Dataset	Model	Selected motor features
Enroll-HD	**Ball-SVM**	Ocularh, ocularv, sacinith, sacinitv, sacvelh, dysarth, tongue, fingtapr, fingtapl, prosupr, prosupl, luria, rigarml, brady, dysttrnk, dystrue, dystlue, dystrle, dystlle, chorface, chorbol, chortrnk, chorrue, chorlue, chorrle, chorlle, gait, tandem, retropls, motscore.
	Box-SVM	Ocularh, ocularv, sacinitv, sacvelh, dysarth, tongue, fingtapr, fingtapl, prosupr, prosupl, luria, rigarml, brady, chorface, chorbol, chortrnk, chorrue, chorlue, chorrle, chorlle, tandem, retropls.

The features that were selected by each model for each problem are given in Table 3 above. The descriptions of all features are given later in "Section 8.2 Appendix B".

6.3 Discussion

In this application, we have used conic duality-based robust SVM methods to establish a highly accurate classification of HD disease status based solely on UHDRS motor scores.

Not all features were selected by the models despite achieving very high classification accuracies, suggesting that the feature selection aspect of these models successfully eliminated unnecessary variables that may interfere with prediction. The **Ball-SVM** achieved a marginally higher classification accuracy than the **Box-SVM** at the expense of selecting more features. All of the features selected by the **Box-SVM** were also selected by the **Ball-SVM**.

The features that were not selected by both models, vertical saccade velocity and right-sided arm rigidity, are members of two pairs of features examining identical aspects of motor function on different sides of the body (left/right) or in different planes (horizontal/vertical). It would not be surprising if only one feature from each pair was sufficient and more efficient for prediction. Alternatively, this may reflect an inherent asymmetry in HD [37,44].

Notably, none of the five features relating to dystonia (involuntary muscle contractions) were selected by the **Box-SVM** model. This may reflect the particular disease phenotype or disease stage of the study population. In typical adult-onset HD, choreiform ("dance-like") movements, which were features that were selected by both

models, dominate early in the disease course, whereas dystonia is not prominent until the later stages [1].

Similarly, gait abnormality, which is a composite effect of multiple other motor abnormalities including late features like dystonia, was not selected by the **Box-SVM** model [43]. TMS was also not selected by this model, suggesting that it does not strongly influence classification when its component items are used as features.

7 Conclusion and Further Work

In this paper, by employing a support function approach and conic duality of convex optimization, we first presented a readily numerically solvable semi-definite program reformulation for a general robust SVM data classification problem, where the uncertainty sets are spectrahedra. A spectrahedron, which is an important generalization of a polyhedron, has played a key role in a range of fields, including algebraic geometry and semi-definite optimization. It also encompasses the most commonly used uncertainty sets of robust optimization models, such as boxes, balls and ellipsoids. We have shown that the conic duality-based robust SVMs with box and ball uncertainty sets achieved classification accuracies of over 95% on the large Enroll-HD dataset and provided important information about the features that characterize HD onset.

As an alternative to the second-order cone program reformulation of the robust SVM with box-uncertainty sets, we also presented a new efficient iterative scheme, **Box-pq-SVM**, for solving the robust SVM by reformulating it as a simple convex quadratic optimization problem via Lagrangian duality. We have demonstrated through computational studies on a range of datasets that these robust classification methods allow for greater classification accuracies than conventional support vector machines in addition to selecting groups of highly correlated features.

Further work is planned to examine the generalizability of the Enroll-HD robust classifier on other HD datasets with the aim of producing a simple and reliable clinical decision support tool to aid in the identification of patients with manifest HD. It would also be of interest to study how the **Box-pq-SVM** approach can be extended to treating large-scale data sets.

8 Appendix

8.1 Appendix A: Spectrahedra

A *spectrahedron* $\Omega \subset \mathbb{R}^m$ is the intersection of the cone of positive semi-definite matrices with an affine-linear space and is represented algebraically by $\Omega := \{(x_1, \ldots, x_m) \in \mathbb{R}^m : A_0 + \sum_{j=1}^m x_j A_j \succeq 0\}$ for some symmetric matrices $A_j, j = 0, 1, \ldots, m$. If A_j's are diagonal matrices, then Ω is known as a polyhedron; thus, a polyhedron is a spectrahedron and the converse is not true. For instance, a closed unit ball $\{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}$ is a bounded spectrahedron as it can be written as:

$$\{(x_1, x_2) \in \mathbb{R}^2 : \begin{pmatrix} 1 - x_1 & x_2 \\ x_2 & 1 + x_1 \end{pmatrix} \succeq 0\}.$$

A spectrahedron is an important extension of a polyhedron, and it covers many convex infinite sets arising in a range of applications (see [42]). The reader is directed to [38] for more details on the links between polyhedra, spectrahedra and semi-definite linear programming. It is worth noting that in general, a bounded spectrahedron can have infinitely many faces, whilst a bounded polyhedron has only finitely many faces (see Fig. 2 for an illustration).

A three-dimensional example of a bounded spectrahedron which has infinitely many faces is the three-dimensional elliptope [42] which is given by $E := \{(x_1, x_2, x_3) \in \mathbb{R}^3 : \begin{pmatrix} 1 & x_1 & x_2 \\ x_1 & 1 & x_3 \\ x_2 & x_3 & 1 \end{pmatrix} \succeq 0\}$.

**Fig. 2** An illustration of polytope
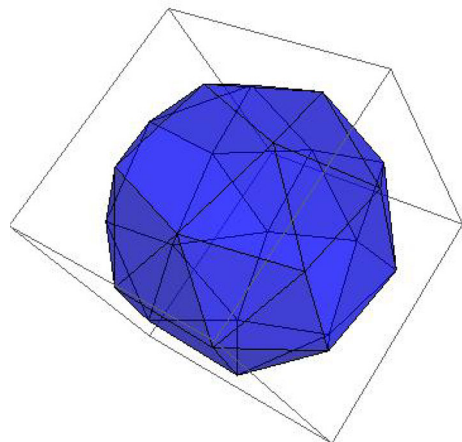
8.2 Appendix B: Motor features of Huntington disease

Table 4 Table caption

Abbreviation	Description
Ocularh	Ocular pursuit—horizontal
Ocularv	Ocular pursuit—vertical
Sacinith	Saccade initiation—horizontal
Sacinitv	Saccade initiation—vertical
Sacvelh	Saccade velocity—horizontal
Sacvelv	Saccade velocity—vertical
Dysarth	Dysarthria
Tongue	Tongue protrusion
Fingtapr	Finger taps—right
Fingtapl	Finger taps—left
Prosupr	Pronation/supination—right
Prosupl	Pronation/supination—left
Luria	Luria
Rigarmr	Arm rigidity—right
Rigarml	Arm rigidity—left
Brady	Bradykinesia—body
Dysttrnk	Maximal dystonia—trunk
Dystrue	Maximal dystonia—right upper extremity
Dystlue	Maximal dystonia—left upper extremity
Dystrle	Maximal dystonia—right lower extremity
Dystlle	Maximal dystonia—left lower extremity
Chorface	Maximal chorea—face
Chorbol	Maximal chorea—mouth
Chortrnk	Maximal chorea—trunk
Chorrue	Maximal chorea—right upper extremity
Chorlue	Chorea—left upper extremity
Chorrle	Maximal chorea—right lower extremity
Chorlle	Maximal chorea—left lower extremity
Gait	Gait
Tandem	Tandem walking
Retropls	Retropulsion
Motscore	Total motor score

Fig. 3 The three-dimensional elliptope E
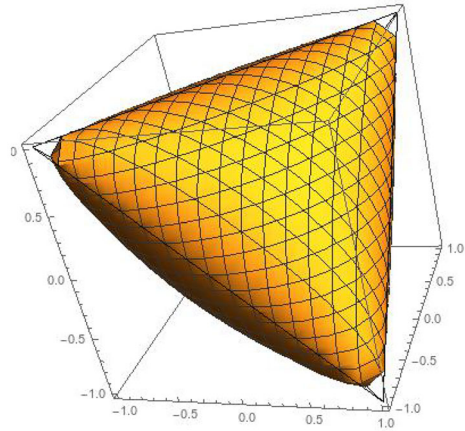
8.3 Appendix C: Convergence of Iterative Scheme (4.10)

Consider the iterative scheme in (4.10)

$$u^{(i+1)} = \max(u^{(i)} - \alpha(Qu^{(i)} - \eta), 0_{m+2s}), \ i = 0, 1, \ldots$$

for solving the quadratic programming problem over non-negative orthant (QP):

$$(QP) \quad \min_{u \in \mathbb{R}^{m+2s}} \frac{1}{2} u^T Q u - \eta^T u \ \text{ subject to } \ u \geq 0_{m+2s}. \tag{8.1}$$

where

$$Q = (\hat{D}\hat{U} + I_{m+2s})(C + \nu I_{m+2s})^{-1}(\hat{D}\hat{U} + I_{m+2s})^T + \hat{D}\hat{e}\hat{e}^T \hat{D}, \quad \eta = \hat{A}\hat{e},$$

with $\nu > 0$, $\hat{D}$, $\hat{U}$, $\hat{A}$ and $\hat{e}$ are given as in (4.3) and (4.4), and I_{m+2s} is the $(m + 2s) \times (m + 2s)$ identity matrix. As (QP) is a convex quadratic program with strictly convex objective function, it has a unique optimal solution. We now have the following linear convergence result for (4.10) whose proof makes use of similar arguments employed in [27,34].

Proposition 8.1 *Let* $u^* \in \mathbb{R}^{m+2s}$ *be the solution for (QP). If* $0 < \alpha < \frac{2\nu}{\|\hat{U}^T\hat{D}+I_{m+2s}\|^2+\nu\|\hat{D}\hat{e}\|^2}$, *then there exists* $r \in (0, 1)$ *such that*

$$\|u^{(i+1)} - u^*\| \leq r\|u^{(i)} - u^*\|.$$

Proof As u^* is the optimal solution for (QP), $0_{m+2s} \leq u^* \perp (Qu^* - \eta) \geq 0_{m+2s}$. This together with (4.9) shows that

$$u^* = \max(u^* - \alpha(Qu^* - \eta), 0_{m+2s}),$$

and so,

$$
\begin{aligned}
\|u^{(i+1)} - u^*\| &= \| \max(u^{(i)} - \alpha(Qu^{(i)} - \eta), 0_{m+2s}) - \max(u^* - \alpha(Qu^* - \eta), 0_{m+2s})\| \\
&\leq \| \left(u^{(i)} - \alpha(Qu^{(i)} - \eta)\right) - \left(u^* - \alpha(Qu^* - \eta)\right) \| \\
&= \|(I_{m+2s} - \alpha Q)(u^{(i)} - u^*)\| \\
&\leq \|I_{m+2s} - \alpha Q\| \cdot \|u^{(i)} - u^*\|,
\end{aligned}
$$

where the first inequality is from the non-expansive property for the projection mapping onto the non-negative orthant. To finish the proof, it suffices to show that $\|I_{m+2s} - \alpha Q\| < 1$. Now, let $d \in \mathbb{R}^{m+2s}$ with $d \neq 0_{m+2s}$ and let $w = (\hat{D}\hat{U} + I_{m+2s})^T d = (\hat{U}^T \hat{D} + I_{m+2s})d$ (as $\hat{D}$ is symmetric). Then, from the definitions of $\hat{D}$ and $\hat{U}$, a direct verification shows that $w \neq 0_{m+2s}$. Moreover,

$$
d^T(I_{m+2s} - \alpha Q)d = \|d\|^2 - \alpha d^T Qd = \|d\|^2 - \alpha \left(w^T(C + \nu I_{m+2s})^{-1} w + \|\hat{e}^T \hat{D}d\|^2 \right).
$$

This gives us that

$$
\|d\|^2 > d^T(I_{m+2s} - \alpha Q)d \geq \|d\|^2 - \alpha \left(\lambda_{\max}\left((C + \nu I_{m+2s})^{-1}\right)\|w\|^2 + \|\hat{e}^T \hat{D}d\|^2 \right).
\tag{8.2}
$$

As C is positive semi-definite, $\lambda_{\max}\left((C + \nu I)^{-1}\right) \leq \lambda_{\max}\left((\nu I_{m+2s})^{-1}\right) = \frac{1}{\nu}$. Moreover, one has $\|w\| \leq \|\hat{U}^T \hat{D} + I_{m+2s}\|\|d\|$ and $\|\hat{e}^T \hat{D}d\|^2 \leq \|\hat{D}^T \hat{e}\|^2\|d\|^2 = \|\hat{D}\hat{e}\|^2\|d\|^2$. It then follows that

$$
\begin{aligned}
&\lambda_{\max}\left((C + \nu I_{m+2s})^{-1}\right)\|w\|^2 + \|\hat{e}^T \hat{D}d\|^2 \\
&\leq \left(\frac{\|\hat{U}^T \hat{D} + I_{m+2s}\|^2}{\nu} + \|\hat{D}\hat{e}\|^2 \right) \|d\|^2 < \frac{2}{\alpha}\|d\|^2,
\end{aligned}
$$

where the last strict inequality follows by our choice of α and the fact that $d \neq 0_{m+2s}$. So, (8.2) implies that $|d^T(I_{m+2s} - \alpha Q)d| < \|d\|^2$ for all $d \neq 0_{m+2s}$. Hence, all the eigenvalues of $I_{m+2s} - \alpha Q$ have absolute value strictly less than one, and so, $\|I_{m+2s} - \alpha Q\| < 1$. Thus, the conclusion follows. □

References

1. Bates, G.P., et al.: Huntington disease. Nat. Rev. Dis. Primers **1**, 15005 (2015)
2. Belousov, E.G., Klatte, D.: A Frank-Wolfe type theorem for convex polynomial optimization problems. Comp. Optim. Appl. **22**, 37–48 (2002)
3. Ben-Tal, A., Nemirovski, A.: Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications. SIAM, Philadelphia (2001)
4. Ben-Tal, A., Ghaoui, L.E., Nemirovski, A.: *Robust Optimization*, Princeton Series in Applied Mathematics (2009)
5. Bertsimas, D., Brown, D.B., Caramanis, C.: Theory and applications of robust optimization. SIAM Rev. **53**, 464–501 (2011)
6. Bertsimas, D., Dunn, J., Pawlowski, C., Zhuoa, Y.D.: Robust classification. Inf. J. Optim. **1**, 2–34 (2019)

7. Bi, J., Zhang, Z.: Support vector classification with input data uncertainty. Adv. Neural Inform. Process. Syst. **17**, 161–168 (2005)

8. Bradley, P., Mangasarian, O.L.: Feature selection via concave minimization and support vector machines. In: Shavlik, J. (ed.) Machine Learning Proceedings of the Fifteenth International Conference (ICML '98), pp. 82-90. Morgan Kaufmann, San Francisco (1998)

9. Cervantes, J., et al.: A comprehensive survey on support vector machine classification: applications, challenges and trends. Neurocomputing **408**, 189–215 (2020)

10. Chuong, T.D., Jeyakumar, V.: An exact formula for radius of robust feasibility of uncertain linear programs. J. Optim. Theor. Appl. **173**, 203–226 (2017)

11. Conners, M.H., Teixeira-Pinto, A., Loy, C.T.: Psychosis and longitudinal outcomes in Huntington disease: the COHORT Study. Neural Neurosourg. Psychiarty **91**, 15–20 (2020)

12. Cysiqued, L.A., Murray, J.M., Dunbar, M., Jeyakumar, V., Brew, B.J.: A screening algorithm for HIV-associated neurocognitive disorders. HIV Med. **11**, 642–649 (2010)

13. Dua, D., Graff, C.: UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine, CA http://archive.ics.uci.edu/ml (2019)

14. Dunbar, M., Murray, J.M., Cysiqued, L.A., Brew, B.J., Jeyakumar, V.: Simultaneous classification and feature selection via convex quadratic programming with application to HIV-associated neurocognitive disorder assessment. Eur. J. Oper. Res. **206**, 470–478 (2010)

15. Feng, J., Liu, L., Wu, D., Li, G., Beer, M.: Dynamic reliability analysis using the extended support vector regression (X-SVR). Mech. Syst. Signal Proc. **126**, 368–391 (2019)

16. Fung, M., Mangasarian, O.L.: A feature selection Newton method for support vector machine classification. Comp. Optim. Appl. **28**, 185–202 (2004)

17. Giannessi, F.: On the theory of Lagrangian duality. Optim. Lett. **1**, 9–20 (2007)

18. Giannessi, F.: Constrained Optimization and Image Space Analysis, vol. 1: Separation of Sets and Optimality Conditions. Springer, New York (2005)

19. Giannessi, F.: Separation theorems and classification problems. Rend. Studi Econ. Quant. **7–27**, (2005)

20. Gil, D., Girela, J.L., De Juan, J., Gomez-Torres, M.J., Johnsson, M.: Predicting seminal quality with artificial intelligence methods. Expert Sys. Appl. **39**, 12564–12573 (2012)

21. Goberna, M.A., Jeyakumar, V., Li, G., Vicente-Perez, J.: Robust solutions to multi-objective linear programs with uncertain data. Eur. J. Oper. Res. **242**, 730–743 (2015)

22. Goberna, M.A., Jeyakumar, V., Li, G., Linh, N.: Radius of robust feasibility formulas for classes of convex programs with uncertain polynomial constraints. Oper. Res. Lett. **44**, 67–73 (2016)

23. Grant, M., Boyd, S.: CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx (2014)

24. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edn. Springer, New York, NY (2017)

25. Hogarth, P., et al.: Interrater agreement in the assessment of motor manifestations of Huntington's disease. Mov. Disord. **20**, 293–297 (2005)

26. Huntington Study Group: Unified Huntington's Disease Rating Scale: Reliability and Consistency. Mov. Disord. **11**, 136–142 (1996)

27. Jeyakumar, V., Li, G., Suthaharan, S.: Support vector machine classifiers with uncertain knowledge sets via robust optimization. Optimization **63**, 1099–1116 (2014)

28. Jeyakumar, V., Li, G., Vicente-Perez, J.: Robust SOS-convex polynomial programs: exact SDP relaxations. Optim. Lett. **9**, 1–18 (2015)

29. Lee, G.M.: On support vector machine classifiers with uncertain knowledge sets. Linear Nonlinear Anal. **5**, 105–111 (2019)

30. Lee, U., et al.: Estimating disease onset from change points of markers measured with error. Biostatistics **1–17**, (2020)

31. Lofberg, J.: YALMIP: A toolbox for modeling and optimization in MATLAB. In: Proceedings of the CACSD Conference, Taipei, Taiwan (2004)

32. Long, J.D., et al.: Tracking motor impairments in the progression of Huntington's disease. Mov. Disord. **29**, 311–319 (2014)

33. Loy, C.T., McCusker, E.A.: Is a Motor Criterion Essential for the Diagnosis of Clinical Huntington Disease? PLoS Curr **5**, 1 (2013)

34. Mangasarian, O.L., Musicant, D.R.: Lagrangian Support Vectors. J. Mach. Learn. Res. **1**, 161–177 (2001)

35. McCusker, E.A., Loy, C.T.: Huntington disease: the complexities of making and disclosing a clinical diagnosis after premanifest genetic testing. Tremor Other Hyperkinet. Mov **7**, 467 (2017)
36. MOSEK ApS: The MOSEK optimization toolbox for MATLAB manual. Version 9.0. http://docs.mosek.com/9.0/toolbox/index.html (2019)
37. Muhlau, M., et al.: Striatal gray matter loss in Huntington's disease is leftward biased. Mov. Disord. **22**, 1169–73 (2007)
38. Ramana, M., Goldman, A.J.: Some geometric results in semi-definite programming. J. Glob. Optim. **7**, 33–50 (1995)
39. Rossi, R.A., Ahmed, N.K.: The network data repository with interactive graph analytics and visualization. http://networkrepository.com/ (2015)
40. Sikora, M., Wrobel, L.: Application of rule induction algorithms for analysis of data collected by seismic hazard monitoring systems in coal mines. Arch Mining Sciences, 55, 91-114 (2010)
41. Trafalis, T.B., Gilbert, R.C.: Robust support vector machines for classification and computational issues. Optim. Methods Softw. **22**, 187–198 (2007)
42. Vinzant, C.: What is a spectrahedron? Not. Am. Math. Soc. **61**, 492–494 (2014)
43. Vuong, K., Canning, C., Menant, J., Loy, C.: Gait, balance and falls in Huntington disease. Handb. Clin. Neurol **159**, 251–260 (2018)
44. Wang, S., et al.: Markedly asymmetrical parkinsonism as a leading feature of adult-onset Huntington's disease. Mov. Disord. **19**, 854–856 (2004)
45. Woolnough, D., Jeyakumar, V., Li, G.: Exact conic programming reformulations of two-stage adjustable robust linear programs with new quadratic decision rules. Optim Lett **15**, 25–44 (2021)
46. Zhang, S., Poon, S., Vuong, K., Sneddon, A., Loy, C.T., MEDINFO.: A deep learning-based approach for gait analysis in Huntington disease. In: Ohno-Machado, L., Seroussi, B. (Eds.) Health and Welbeing e-Networks for All. IMIA and IOS Press, pp. 477–481

Authors and Affiliations

Daniel Woolnough[1] · Niroshan Jeyakumar[2] · Guoyin Li[1] · Clement T Loy[3,4] · Vaithilingam Jeyakumar[1]

✉ Guoyin Li
g.li@unsw.edu.au

Daniel Woolnough
daniel.woolnough@unsw.edu.au

Niroshan Jeyakumar
niroshan.jeyakumar@health.nsw.gov.au

Clement T Loy
clement.loy@sydney.edu.au

Vaithilingam Jeyakumar
v.jeyakumar@unsw.edu.au

[1] Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia

[2] Westmead Hospital, Westmead, NSW 2145, Australia

[3] Brain and Mind Centre, the University of Sydney, Sydney, Australia

[4] the Garvan Institute of Medical Research, Sydney, Australia