Machine Learning and Risk Management: SVDD Meets RQE *

Georges DIONNE † Gilles Boevi KOUMOU‡

Preliminary version: Friday 30th November, 2018

Abstract

This paper re-examines Rao's Quadratic Entropy (RQE) portfolio diversification in the light of the support vector data description (SVDD), an unsupervised machine learning algorithm developed for the one-class classification. It makes the link between RQE portfolio diversification and the SVDD. More specifically, we show that RQE portfolio diversification problem is equivalent to the dual representation of a hard-margin SVDD problem. This result demonstrates, on the one hand, that the SVDD and its rich set of extensions are relevant for risk management, in particular portfolio diversification. On the second hand, it provides new insights on RQE portfolio diversification approach in terms of interpretation, understanding, implementation, specification, and optimization. This strengthens the believe that machine learning can play an important role in risk management and RQE is an adequate framework to quantify and to manage portfolio diversification.

JEL Classifications : G11, D83

Keywords : Machine Learning, One-Class Classification, Support Vector Data Description, Rao's Quadratic Entropy, Portfolio Diversification.

*Corresponding author: Gilles Boevi KOUMOU, nettey.koumou@hec.ca.

†Canada Research Chair in Risk Management, Departement of Finance, HEC Montréal, 3000, chemin de la Côte-Sainte-Catherine, Montréal, Québec, H3T 2A7, Canada.

‡Canada Research Chair in Risk Management, Departement of Finance, HEC Montréal, 3000, chemin de la Côte-Sainte-Catherine, Montréal, Québec, H3T 2A7, Canada.

1 Introduction

Diversification is at the core of investment decision making as well as risk and return. Thus, its measurement and management is of fundamental importance in finance, and more generally in economics. This is why since Markowitz (1952)'s seminal work on portfolio selection, several measures of portfolio diversification were proposed (see Booth and Fama, 1992; Choueifaty and Coignard, 2008; Evans and Archer, 1968; Fernholz and Shay, 1982; Goetzmann et al., 2005; Meucci, 2009; Meucci et al., 2014; Rudin and Morgan, 2006; Sharpe, 1972; Statman and Scheid, 2005; Vermorken et al., 2012; Woerheide and Persson, 1993). Recently, Carmichael et al. (2015) (see also Koumou (2017) and Carmichael et al. (2018)) introduced a novel and powerful class of measures of portfolio diversification inspired from Rao'Quadratic Entropy (RQE).

RQE is a general framework for measuring population diversity developed by Rao (1982a). It has been used in fields such as statistics (see Nayak, 1986a,b; Rao, 1982a,b) and ecology (see Champely and Chessel, 2002; Pavoine, 2012; Pavoine and Bonsall, 2009; Pavoine et al., 2005; Ricotta and Szeidl, 2006; Zhao and Naik, 2012). It is also applied in energy policy (see Stirling, 2010) and economics (see Nayak and Gastwirth, 1989). Recently, Carmichael et al. (2015) extended and adapted its use to the portfolio theory as a novel class of measures of portfolio diversification. The authors argue that when RQE is adequately calibrated it becomes a valid class of portfolio diversification measures summarizing complex features of portfolio diversification in a simple manner, and provides at the same time a unified theory that includes many previous contributions. Using this new class of portfolio diversification measures, Carmichael et al. (2018) provide new formulations of the maximum diversification approach developed by Choueifaty and Coignard (2008) and currently used to manage a portfolio of 8 billion dollars U.S by the firm Think Out of the Box Asset Management (TOBAM).[1] These new formalizations clarify the investment problem behind the maximum diversification approach, helps identify the source of its optimal portfolios strong out-of-sample performance relative to other diversified portfolios, and suggest new directions along which its out-of-sample performance can be improved.

The purpose of this paper is to re-examine RQE portfolio diversification in link with support vector data description (SVDD), an unsupervised machine learning algorithm designed for

[1] https://www.tobam.fr/

the one-classes classification (OCC) problem.

In contrast to the normal classification problems where one tries to distinguish between two (or more) classes of objects, the OCC (see Khan and Madden, 2009, 2014; Tax, 2001) tries to describe one class of objects (called target class), and distinguishes it from all other possible class of objects (considered as outliers or novelty). This classification is applied when only one class of objects is well characterized and the other class has either no or very few instances. The OCC approach gained much popularity during the last two decades, and has been generally employed to solve the problem of novelty, outlier, intrusion, fault or anomaly detection, badly balanced data, credit scoring and data set comparison (see Bartkowiak, 2011; Burnaev and Smolyakov, 2016; Chen et al., 2016; Kennedy et al., 2009; Khan and Madden, 2009, 2014; Ma and Perkins, 2003; Manevitz and Yousef, 2001; Mourao-Miranda et al., 2011; Nader et al., 2016).

Several algorithms have been proposed to compute the OCC model, including the support vector data description (SVDD); see Khan and Madden (2009, 2014); Nanni (2006); Tax (2001); Wang et al. (2004) for a comprehensive review. The SVDD is one of the most successful of OCC's algorithms. It is a non-parametric algorithm developed by Tax (2001) and Tax and Duin (2004) inspired from the support vector machine and based on the smallest enclosing ball problems. In the SVDD, the classifier is a hyperspherical boundary, in the input space or a high-dimensional feature space using some kernel functions, which distinguishes target class from outliers or novelty.

The SVDD has been intensively and successfully applied in various domains, such as pattern denoising (Park et al., 2007), face recognition (Lee et al., 2006), anomaly detection (Feng et al., 2017), among others. Recently, with the growing interest in applying machine learning's algorithms in financial industries (see QF2, 2018; Abu-Mostafa and Atiya, 1996; Antunes et al., 2017; Atkins et al., 2018; Ban et al., 2016; Dunis et al., 2016; Fischer and Krauss, 2017; Gyorfi et al., 2012; Kim, 2003; Krauss et al., 2017; Lee, 2007; Li and Hoi, 2014; Li et al., 2016; Liew and Mayster, 2017; Lin et al., 2012; Liu and Xie, 2018; Nava et al., 2018; Tay and Cao, 2001; White, 1988), and in economics (see Exterkate, 2013; Exterkate et al., 2016; Mullainathan and Spiess, 2017), there are some applications of the SVDD in finance. These applications cover credit fraud detection in banking sector (Gangolf et al., 2014; Pang et al., 2014), for example, but there are still too few.

In this paper, we propose a novel and useful application of the SVDD to risk management through portfolio diversification. We make the link between RQE portfolio diversification and the SVDD. More specifically, we show that RQE portfolio diversification optimization problem is equivalent to the dual representation of a hard-margin SVDD problem. This result demonstrates, on the one hand, that the SVDD and its rich set of extensions are relevant for financial portfolio selection. It also provides new insights on RQE portfolio diversification approach in terms of interpretation, understanding, implementation, specification and optimization. This strengthens the believe that machine learning can play an important role in risk management and RQE is an adequate framework to quantify and to manage portfolio diversification.

The remainder of this paper is organized as follows. Section 2 reviews the general definition of RQE and its portfolio selection version. Section 3 reviews the SVDD algorithm. Section 4 establishes the link between the RQE portfolio diversification optimization problem and the SVDD algorithm. Section 5 discusses some implications of our result and concludes.

2 Rao's Quadratic Entropy

This section refreshes the general definition of Rao's Quadratic Entropy (RQE) and its portfolio selection version. We follow Rao (1982b) and Rao (1982a) for the general definition and Carmichael et al. (2015), Carmichael et al. (2018) and Koumou (2017) for the definition in the context of portfolio selection.

2.1 General Definition

RQE[2] is a general statistical tool to measure diversity of a population of individuals. It was developed by Rao (1982a,b). Given a population of individuals $\mathcal{P}$, it is defined as the average difference between two randomly drawn individuals from $\mathcal{P}$. More formally, suppose that each individual in $\mathcal{P}$ is characterized by a set of measurement X with P the associated probability distribution function. RQE of $\mathcal{P}$ is defined as

$$H_{\mathbf{D}}(P) = \int d(X_1, X_2)P(dX_1)P(dX_2), \quad (1)$$

[2] RQE is also referred to as Diversity Coefficient (Rao, 1982b) or Quadratic Entropy (Rao and Nayak, 1985).

where the non-negative, symmetric dissimilarity function $d(.,.)$ expresses the difference between two individuals from $\mathcal{P}$.

When X is a discrete random variable, $H_{\mathbf{D}}(P)$ becomes

$$H_{\mathbf{D}}(\mathbf{p}) = \mathbf{p}^{\top}\mathbf{D}\,\mathbf{p}, \tag{2}$$

where $\mathbf{p}$ is a column vector of probabilities with elements $p_i = P(X = x_i)$, $\forall\, i = 1, ..., N$ and $\mathbf{D} = (d_{ij})_{i,j=1}^{N}$ is the dissimilarity matrix, where N is a number of individuals i.e. the cardinal of $\mathcal{P}$.

The interpretation of RQE is straightforward: the higher is $H_{\mathbf{D}}(P)$, the higher is the diversity of individuals among the population $\mathcal{P}$.

RQE has been used extensively in various domains, such as statistics (see Nayak, 1986a,b; Rao, 1982a,b), ecology (see Champely and Chessel, 2002; Pavoine, 2012; Pavoine and Bonsall, 2009; Pavoine et al., 2005; Ricotta and Szeidl, 2006; Zhao and Naik, 2012), energy policy (see Stirling, 2010) and economics (see Nayak and Gastwirth, 1989). Recently, Carmichael et al. (2015) (see also Carmichael et al. (2018) and Koumou (2017)) extended and adapted its use to portfolio theory as a novel class of measures of portfolio diversification. To do so, the authors calibrated the magnitudes $\mathcal{P}$, X, P and $\mathbf{D}$ in the portfolio selection context. The next section reviews this extension.

2.2 Portfolio Definition

Consider a universe of N assets (risky or not) and denote by $\mathbf{w} = (w_i)_{i=1}^{N}$ a given, long-only portfolio associated to this universe of assets, with w_i the weight of asset i in $\mathbf{w}$. Carmichael et al. (2015) consider $\mathbf{w}$ as a population of assets. Next, the authors define the random variable X to take the finite values $1, ..., N$ (N assets) and its probability distribution function $P(X = i) = w_i$, $\forall\, i = 1, ..., N$, so that it is associated to the random experiment whereby assets are randomly selected (with replacement) from portfolio $\mathbf{w}$. Finally, the authors define RQE of a portfolio $\mathbf{w}$ as half of the mean difference between two randomly drawn (with replacement) assets from portfolio $\mathbf{w}$

$$H_{\mathbf{D}}(\mathbf{w}) = \frac{1}{2}\,\mathbf{w}^{\top}\mathbf{D}\,\mathbf{w}, \tag{3}$$

where $\mathbf{D} = (d_{ij})_{i,j=1}^{N}$ is the dissimilarity matrix between the various assets of the portfolio $\mathbf{w}$. The authors (see also Koumou (2017)) demonstrates that, when $\mathbf{D}$ is adequately specified, $H_{\mathbf{D}}(\mathbf{w})$ is a valid class of portfolio diversification measure, which i) meets ex-ante desirable properties of diversification; ii) unifies several existing measures; iii) at a core of the mean-variance utility functions; iv) and provides a flexible but formal approach for fund managers to develop new diversified portfolios.

The interpretation of $H_{\mathbf{D}}(\mathbf{w})$ is straightforward. All things being equal, the higher $H_{\mathbf{D}}(\mathbf{w})$ is, the more portfolio $\mathbf{w}$ is diversified, because the more dissimilar are assets, the less is the probability that they do poorly at the same time and in the same proportion. A well-diversified portfolio can therefore be obtained by maximizing (3)

$$\mathbf{w}^{RQE} \in \underset{\mathbf{w}\in\mathbb{W}}{\arg\operatorname{Max}}\, H_{\mathbf{D}}(\mathbf{w}), \tag{4}$$

where $\mathbb{W}$ is the set of long-only portfolios and $\mathbf{w}^{RQE}$ is the RQE portfolios or the well-diversified portfolios of RQE.

What makes RQE attractive comparatively to other entropy measures used in finance (see Zhou et al., 2013) is the dissimilarity matrix $\mathbf{D}$. First, $\mathbf{D}$ provides to RQE a real advantage over its competitors to measure and manage portfolio diversification. Second, it also gives an extreme flexibility to portfolio managers in the construction of their well-diversified portfolio. However, this attractiveness becomes a challenge when it comes the time of the implementation of RQE. Carmichael et al. (2015) (see also Koumou (2017); **?**) suggest guidelines for the choice of $\mathbf{D}$ and provide some examples. New directions are also suggested by the results of this paper, in which we re-examine RQE portfolio diversification in the light of the support vector data description (SVDD), an unsupervised machine learning algorithm designed for the one-class classification (OCC) problem. Specifically, we establish the link between RQE portfolio diversification optimization problem (4) and the SVDD algorithm by showing that the two are equivalent.

3 Support Vector Data Description

In this section, we introduce the optimization problems of the support vector data description (SVDD). We follow Tax (2001) and Tax and Duin (2004).

Before, we provide a brief description of the one-class classification (OCC)'s algorithms. Given a training set of objects $\mathcal{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$ with $\mathbf{x}_i \in \mathbb{R}^T$, $i = 1, ..., N$ and T an integer such that $T \geq 2$, all OCC's algorithms identify outliers or novelty objects based on two elements: discriminant function or classifier $f(\mathbf{x})$ and the threshold parameter θ; more formally

$$g(\mathbf{x}) = \mathbf{1}\left(f(\mathbf{x}) < \theta\right) = \begin{cases} 1 & \mathbf{x} \text{ is a target object} \\ 0 & \mathbf{x} \text{ is an outlier or a novelty.} \end{cases} \tag{5}$$

The discriminant function or classifier $f(\mathbf{x})$ is a measure for the distance or resemblance (or probability) and $\mathbf{1}(.)$ is an indicator function. The difference between the existing OCC's algorithms lie in the specification of $f(\mathbf{x})$ and the optimization of $f(\mathbf{x})$ and θ. In this paper, we focus on the SVDD, an OCC's algorithm developed by Tax and Duin (2004) (see also Tax (2001)) based on the smallest enclosing ball problem (see Elzinga and Hearn, 1972a,b; Gartner, 1999; Kallberg and Larsson, 2013; Larsson and Kallberg, 2013; Sylvester, 1857; Yildirim, 2008; Zhou et al., 2005) and inspired by Schiilkop et al. (1995)'s work. As shown in Tax and Duin (2004), the SVDD gives solutions similar to Scholkopf et al. (1999b)'s OCC approach which uses a hyperplane to describe the data. In what follows, we present the standard SVDD and its kernel extension (kernel SVDD).

3.1 Normal Data Description

Contrary to Scholkopf et al. (1999b)'s OCC approach which uses a hyperplane to describe the data, SVDD is an OCC's algorithm which uses a hypersphere approach based on the smallest enclosing ball. More formally, the SVDD minimizes the volume of the sphere by minimizing the square of its radius ζ^2 under the constraint that the sphere contains all training objects $\mathcal{X}$; more formally

$$\min_{(\overline{\mathbf{x}}, \zeta^2) \in \mathbb{R}^T \times \mathbb{R}_+} \zeta^2 \tag{6}$$

$$\text{s.t } \|\overline{\mathbf{x}} - \mathbf{x}_i\|_2^2 \leq \zeta^2, \ \ \forall\, i = 1, ..., N, \tag{7}$$

where $\overline{\mathbf{x}}$ is the center of the hypersphere. ζ^2 is also call the structural error function (see Tax, 2001). The problem (6)-(7) is a hard-margin formulation of the SVDD. This formulation assumes that the training set does not contain outliers or novelty. Therefore, the formulation is only useful for data description and cannot be used to outliers or novelty detection.

The Lagrangian of problem (6)-(7) is

$$L(\zeta, \boldsymbol{\nu}, \overline{\mathbf{x}}) = \zeta^2 - \sum_{i=1}^{N} \nu_i \left(\zeta^2 - \|\overline{\mathbf{x}} - \mathbf{x}_i\|_2^2\right) \tag{8}$$

with the Lagrange multipliers $\nu_i \geq 0,\ i = 1, ..., N$. $L(\zeta, \boldsymbol{\nu}, \overline{\mathbf{x}})$ has to be minimized with respect to ζ and $\overline{\mathbf{x}}$, and maximized with respect to $\boldsymbol{\nu}$. The first order conditions gives

$$\sum_{i=1}^{N} \nu_i = 1 \tag{9}$$

$$\overline{\mathbf{x}} = \sum_{i=1}^{N} \nu_i \, \mathbf{x}_i \tag{10}$$

Substituting (9)-(10) into $L(\zeta, \boldsymbol{\nu}, \overline{\mathbf{x}})$, the dual form of (6)-(7) is obtained

$$\max_{\boldsymbol{\nu}} \ \sum_{i=1}^{N} \nu_i \|\overline{\mathbf{x}} - \mathbf{x}_i\|_2^2 = \sum_{i=1}^{N} \nu_i \langle \mathbf{x}_i, \mathbf{x}_i \rangle - \sum_{i,j=1}^{N} \nu_i \, \nu_j \, \langle \mathbf{x}_i, \mathbf{x}_j \rangle \tag{11}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} \nu_i = 1 \tag{12}$$

$$\nu_i \geq 0 \ \ \forall \, i = 1, ..., N, \tag{13}$$

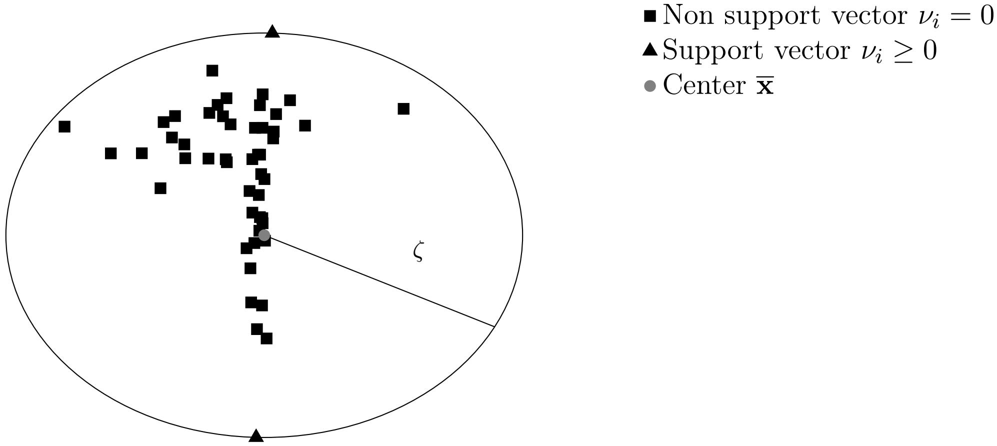
where $\langle ., . \rangle$ is an inner product on $\mathbb{R}^T$. At optimal, the objects for $\nu_i > 0$ are called support objects of the description or the support vectors. Only these objects are needed in the description of the data set, and they satisfy the equality that determine the value of ζ, $\|\overline{\mathbf{x}} - \mathbf{x}_i\|_2^2 = \zeta^2$. The classification rule is therefore

$$f(\mathbf{z}) = \mathbf{1}\left(\|\overline{\mathbf{x}} - \mathbf{z}\|_2^2 \leq \zeta^2\right) = \begin{cases} 1 & \mathbf{z} \text{ is a target object} \\ 0 & \mathbf{z} \text{ is an outlier or a novelty} \end{cases} \tag{14}$$

See Figure 1 for an illustration of the hard-margin SVDD.

To allow the possibility of outliers or empirical errors in $\mathcal{X}$, Tax (2001) and Tax and Duin

Figure 1: Hard-Margin SVDD



(2004) introduce slack variables $\xi_i \geq 0$ and the problem (6)-(7) becomes

$$\min_{(\overline{\mathbf{x}},\zeta^2,\boldsymbol{\xi})\in\mathbb{R}^T\times\mathbb{R}_+\times\mathbb{R}^N} \quad \zeta^2 + C\sum_{i=1}^{N}\xi_i \tag{15}$$

$$\text{s.t } \|\overline{\mathbf{x}} - \mathbf{x}_i\|_2^2 \leq \zeta^2 + \xi_i, \;\; \forall\, i = 1, ..., N \tag{16}$$

$$\xi_i \geq 0, \;\; \forall\, i = 1, ..., N. \tag{17}$$

The parameter C is the regularization parameter which controls the trade-off between the volume of the sphere and the errors. When $C \geq 1$ the problem (15)-(17) coincides with the problem (6)-(7). When $C < \frac{1}{N}$, it is straightforward to verify that the problem (15)-(17) is infeasible. Thus, the parameter C must be specified such as

$$\frac{1}{N} \leq C \leq 1.$$

Inspired from the φ-support vectors machine of Scholkopf et al. (1999b), Tax and Duin (2004) and Tax (2001) suggest

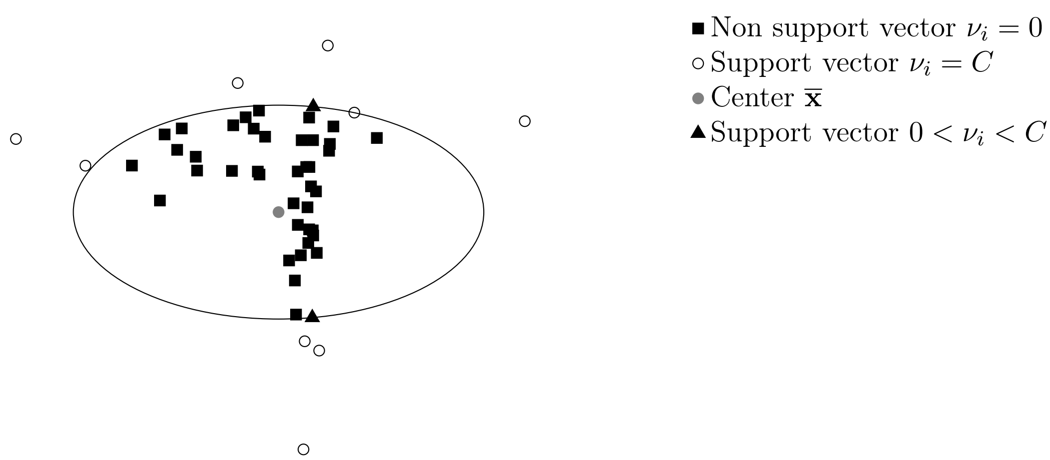
$$C = \frac{1}{N\varphi},$$

where $\varphi \in [0, 1]$ is the upper-bound of the proportion of outliers in the data.

The Lagrangian of the problem (15)-(17) is

$$L\left(\zeta, \boldsymbol{\gamma}, \boldsymbol{\nu}, \overline{\mathbf{x}}, \boldsymbol{\xi}\right) = \zeta^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\nu_i\left(\zeta^2 + \xi_i - \|\overline{\mathbf{x}} - \mathbf{x}_i\|_2^2\right) - \sum_{i=1}^{N}\gamma_i\xi_i \tag{18}$$

Figure 2: Soft-Margin SVDD



with the Lagrange multipliers $\nu_i \geq 0$ and $\gamma_i \geq 0$, $i = 1, ..., N$. $L(\zeta, \boldsymbol{\gamma}, \boldsymbol{\nu}, \overline{\mathbf{x}}, \boldsymbol{\xi})$ has to be minimized with respect to ζ, $\overline{\mathbf{x}}$ and $\boldsymbol{\xi}$, and maximized with respect to $\boldsymbol{\nu}$ and $\boldsymbol{\gamma}$. The first order conditions give

$$\sum_{i=1}^{N} \nu_i = 1 \tag{19}$$

$$\overline{\mathbf{x}} = \sum_{i=1}^{N} \nu_i \, \mathbf{x}_i \tag{20}$$

$$C - \nu_i - \gamma_i = 0, \;\; \forall \, i = 1, ..., N \tag{21}$$

Since $\gamma_i \geq 0$ and $\nu_i \geq 0$, (21) implies that

$$0 \leq \nu_i \leq C, \;\; \forall \, i = 1, ..., N. \tag{22}$$

Substituting (19)-(20) into $L(\zeta, \boldsymbol{\gamma}, \boldsymbol{\nu}, \overline{\mathbf{x}}, \boldsymbol{\xi})$, the dual form of (15)-(17) is obtained

$$\max_{\boldsymbol{\nu}} \;\; \sum_{i=1}^{N} \nu_i \|\overline{\mathbf{x}} - \mathbf{x}_i\|_2^2 = \sum_{i=1}^{N} \nu_i \langle \mathbf{x}_i, \mathbf{x}_i \rangle - \sum_{i,j=1}^{N} \nu_i \, \nu_j \, \langle \mathbf{x}_i, \mathbf{x}_j \rangle \tag{23}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} \nu_i = 1 \tag{24}$$

$$0 \leq \nu_i \leq C \;\; \forall \, i = 1, ..., N \tag{25}$$

At optimal, the objects with $\nu_i > 0$ are called support vectors or support objects. The objects with $0 < \nu_i < C$ are called margin support vectors. The objects with $\nu_i = C$ are regarded as outliers (or nonmargin support vectors), and will not be accepted by the data description. The objects with $\nu_i = 0$ are considered unnecessary for data description. It can be regarded as noise. The classification rule is therefore

$$f(\mathbf{z}) = \mathbf{1}\left(\|\overline{\mathbf{x}} - \mathbf{z}\|_2^2 \leq \zeta^2\right) = \begin{cases} 1 & \mathbf{z} \text{ is a target object} \\ 0 & \mathbf{z} \text{ is an outlier or a novelty.} \end{cases} \tag{26}$$

The square of the radius ζ^2 is given by $\zeta^2 = \|\overline{\mathbf{x}} - \mathbf{x}_i\|_2^2$, where $\mathbf{x}_i$ is such that $0 < \nu_i < C$ i.e. a support vector. See Figure 2 for an illustration of the soft-margin SVDD.

3.2 Non-Spherical Data Description: Kernel SVDD

In this section, we present an extension of the SVDD based on *kernel trick*. For other extensions, we refer readers to Cha et al. (2014); Chen et al. (2015); El Azami et al. (2014); Hamidzadeh et al. (2017); Lee et al. (2007); Liu et al. (2013, 2014); Mygdalis et al. (2017); Sadeghi and Hamidzadeh (2016); Tao et al. (2016); Xuanthanh et al. (2017); Yin and Huang (2011).

Because most of the data is not spherical distributed, the standard SVDD does not worked properly. It is not flexible enough to provide a tight description of the target class (see Figure 3 for example). To make a more accurate decision and more tight description, Tax and Duin (2004) and Tax (2001) suggest to transform the data to a higher dimensional feature space $\mathcal{H}$ by a function

$$\Phi : \mathbb{R}^T \to \mathcal{H} \tag{27}$$

$$\mathbf{x} \mapsto \Phi(\mathbf{x}). \tag{28}$$

The principal objective is to obtain a more spherical data $\Phi(\mathcal{X}) = \{\Phi(\mathbf{x}_1), ..., \Phi(\mathbf{x}_N)\}$ with $\Phi(\mathbf{x}_i) \in \mathcal{H}$, $i = 1, ..., N$ and to recover the true boundary in the input space through the inverse of Φ or the principle of preimage (Scholkopf et al., 1999a; Weston et al., 2004).

Applying the SVDD to $\Phi(\mathcal{X})$, we obtain

$$\max_{\boldsymbol{\nu}} \quad \sum_{i=1}^{N} \nu_i \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_i) \rangle_{\mathcal{H}} - \sum_{i,j=1}^{N} \nu_i \, \nu_j \, \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle_{\mathcal{H}} \tag{29}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} \nu_i = 1 \tag{30}$$

$$0 \leq \nu_i \leq C \;\; \forall \, i = 1, ..., N, \tag{31}$$

where $\langle ., . \rangle_{\mathcal{H}}$ is the inner product in the space $\mathcal{H}$. To avoid working in the potentially high dimensional space $\mathcal{H}$, the mapping Φ is generally defined implicitly through the *kernel trick*. The *kernel trick* is a technique developed by Vapnik (1998) which consists in replacing the inner products $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle_{\mathcal{H}}$ by a kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$ such that

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle_{\mathcal{H}}. \tag{32}$$

The definition of Φ via $k(\mathbf{x}_i, \mathbf{x}_j)$ is possible if and only if the kernel matrix $\mathbf{K} = (k(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1}^{N}$ is positive definite. It is also possible if and only if $\mathbf{K}$ is conditionally positive definite. In that case (32) becomes

$$\|\Phi(\mathbf{x}_i) - \Phi(\mathbf{x}_j)\|_{\mathcal{H}}^2 = -k(\mathbf{x}_i, \mathbf{x}_j) + \frac{1}{2} \left(k(\mathbf{x}_i, \mathbf{x}_i) + k(\mathbf{x}_j, \mathbf{x}_j) \right). \tag{33}$$

For the connection between the positive definite and conditionally positive definite kernels, we refer readers to Schölkopf et al. (2001). For a better understanding of the *kernel trick* let us consider the polynomial kernel function

$$k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^\top \mathbf{y} + \theta)^m, \tag{34}$$

where θ and m are the hyper parameters with m the degree of polynomial and θ a parameter controlling for the relative weightings of the different degree monomials. Increasing θ decreases the relative weighting of the higher order polynomials (see, Part III, Chapter 9 Shawe-Taylor and Cristianini, 2004). Assume that the degree $m = 2$, $\theta = 1$, $\mathbf{x} = (x_1, x_2)^\top$ and $\mathbf{y} = (y_1, y_2)^\top$. By expressing (34) in terms of monomials of various order, we obtain

$$k(\mathbf{x}, \mathbf{y}) = 1 + x_1^2 y_1^2 + 2x_1 x_2 y_1 y_2 + x_2^2 y_2^2 + 2x_1 y_1 + 2x_2 y_2. \tag{35}$$

Equation (35) can be factorized as follows

$$k(\mathbf{x},\mathbf{y}) = \langle (1, x_1^2, \sqrt{2}x_1x_2, x_2^2, \sqrt{2}x_i, \sqrt{2}x_2)^\top, (1, y_1^2, \sqrt{2}y_1y_2, y_2^2, \sqrt{2}y_i, \sqrt{2}y_2)^\top \rangle_{\mathcal{H}}. \tag{36}$$

From (36), it follows that the feature map induced by $k(\mathbf{x},\mathbf{y})$ is

$$\Phi(\mathbf{x}) = (1, x_1^2, \sqrt{2}x_1x_2, x_2^2, \sqrt{2}x_i, \sqrt{2}x_2)^\top. \tag{37}$$

However, it is difficult or impossible in general to express Φ explicitly, even for the polynomial kernel.

By exploiting the *kernel trick* relation (32), the dual form (29)-(31) can therefore be rewritten as follows

$$\max_{\boldsymbol{\nu}} \sum_{i=1}^{N} \nu_i \, k(\mathbf{x}_i, \mathbf{x}_i) - \sum_{i,j=1}^{N} \nu_i \, \nu_j \, k(\mathbf{x}_i, \mathbf{x}_j) \tag{38}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} \nu_i = 1 \tag{39}$$

$$0 \leq \nu_i \leq C \;\; \forall \, i = 1, ..., N \tag{40}$$

The classification rule is

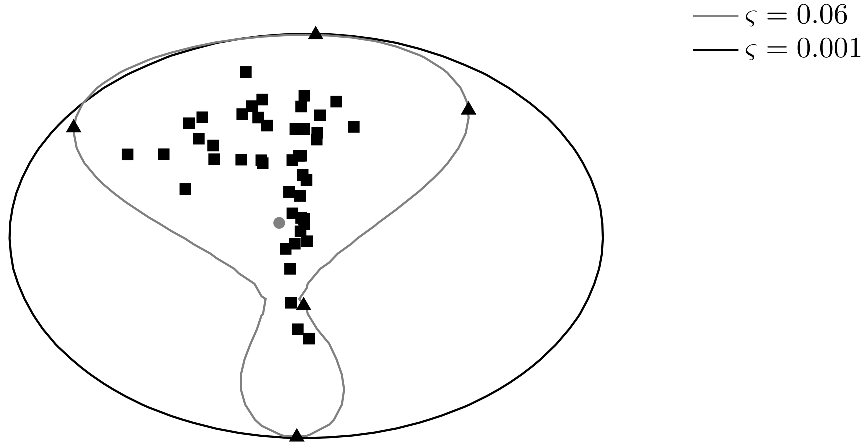
$$f(\mathbf{z}) = \mathbf{1}\left(\|\overline{\Phi(\mathbf{x})} - \Phi(\mathbf{z})\|_2^2 \leq \zeta^2 \right) = \begin{cases} 1 & \mathbf{z} \text{ is a target object} \\ 0 & \mathbf{z} \text{ is an outlier or a novelty} \end{cases} \tag{41}$$

where $\overline{\Phi(\mathbf{x})} = \sum_{i=1}^{N} \nu_i \Phi(\mathbf{x}_i)$. The square of the radius ζ^2 is given by $\zeta^2 = \|\overline{\Phi(\mathbf{x})} - \Phi(\mathbf{x}_i)\|_2^2$, where $\Phi(\mathbf{x}_i)$ is such that $0 < \nu_i < C$ i.e. a support vector. See Figure 3 for an illustration of the kernel SVDD.

The challenge in the implementation of kernel SVDD is the choice of the kernel function. Several kernel functions have been proposed for the support vector classifier. For other examples, readers are refereed to Scholkopf (1993), Smola et al. (1998), Kamath et al. (2010) and Bavaud (2011).

The two most kernel functions used in the SVDD are the polynomial kernel with $\theta \neq 0$ (see

Figure 3: Kernel SVDD



Equation (34)) and the Gaussian kernel defined by

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\varsigma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right), \ \varsigma \geq 0, \tag{42}$$

where ς is the width parameter playing the same role as m in polynomial kernel. Small value of ς correspond to large values of m. Tax and Duin (2004) and Tax (2001) proved that the Gaussian kernel deliver better performances than the polynomial kernel. However, these better performances depend profoundly both on the width parameter ς and the differences of scale in the input space.

For the calibration of ς, we refer the readers to Nader et al. (2014); Tax and Duin (2004); Tax and Juszczak (2003); Tax (2001). For the scaling problem, several rescaling approaches are suggested. We refer the readers to Forghani et al. (2011); Juszczak (2006); Juszczak et al. (2002); Lee et al. (2015); Nader et al. (2014); Tax and Juszczak (2003); Wang and Xiao (2017).

4 RQE Meets SVDD

This section contains the main result of our paper. It makes the link between the RQE portfolio diversification optimization problem and that of the SVDD. Note that the link between RQE and smallest enclosing ball problems (which is the core of the SVDD) was first mentioned by Pavoine (2005) in the context of the application of RQE to ecology. Here, this link is studied through the SVDD problem in the context of risk management, particularly in portfolio diversification.

Before, let us introduction some useful definitions. These definitions are taken from Pekalska and Duin (2005).

Definition 4.1. *An $N \times N$ dissimilarity matrix $\mathbf{D}$ is said to be Euclidean if and only if $\mathbf{D}$ can be embedded in a Euclidean space $(\mathbb{R}^M, \|.\|_2)$ i.e. there is N points $\mathbf{x}_1, ..., \mathbf{x}_N$ in space $(\mathbb{R}^M, \|.\|_2)$ such that $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2^2$.*

Definition 4.2 (Pseudo-Euclidean Spaces (Pekalska and Duin, 2005)). *The Pseudo-Euclidean Spaces $\mathcal{E} = \mathbb{R}^{(\underline{p},\underline{q})}$ is a real vector space equipped with a non-degenerate, indefinite inner product $\langle ., . \rangle_{\mathcal{E}}$. $\mathcal{E}$ admits a direct orthogonal decomposition $\mathcal{E} = \mathcal{E}_+ \oplus \mathcal{E}_-$, where $\mathcal{E}_+ = \mathbb{R}^{\underline{p}}$ and $\mathcal{E}_- = \mathbb{R}^{\underline{q}}$ and the inner product is positive definite on $\mathcal{E}_+$ and negative definite on $\mathcal{E}_-$. A vector $\mathbf{x} \in \mathcal{E}$ is represented as an ordered pair of two real vectors: $\mathbf{x} = (\mathbf{x}^+, \mathbf{x}^-)$. The inner product in $\mathcal{E}$ is defined as $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{E}} = \mathbf{x}^\top \mathbf{I}_{\underline{p}\,\underline{q}} \mathbf{y} = \sum_{i=1}^{\underline{p}} x_i^+ y_i^+ - \sum_{i=1}^{\underline{q}} x_i^- y_i^-$, where $\mathbf{I}_{\underline{p}\,\underline{q}} = [\mathbf{I}_{\underline{p} \times \underline{p}}\ \mathbf{0}; \mathbf{0}\ \mathbf{I}_{\underline{q} \times \underline{q}}]$. As a result, norm of $\mathbf{x} \in \mathcal{E}$ denoted by $\|.\|_{\mathcal{E}}^2$ is defined as $\|\mathbf{x}\|_{\mathcal{E}}^2 = \mathbf{x}^\top \mathbf{I}_{\underline{p}\,\underline{q}} \mathbf{x} = \sum_{i=1}^{p} x_i^+ x_i^+ - \sum_{i=1}^{p} x_i^- x_i^-$*

Definition 4.3 (Pseudo-Euclidean Embedding (Pekalska and Duin, 2005)). *The dissimilarity matrix $\mathbf{D}$ is Pseudo-Euclidean if it can be embedded in a Pseudo-Euclidean space $\left(\mathbb{R}^{(\underline{p},\underline{q})}, \|.\|_{\mathcal{E}}\right)$ i.e. there is N points $\mathbf{x}_1, ..., \mathbf{x}_N$ in space $\left(\mathbb{R}^{(\underline{p},\underline{q})}, \|.\|_{\mathcal{E}}\right)$ such that $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathcal{E}}^2$.*

We now present our main results. Consider first the case where the dissimilarity matrix $\mathbf{D}$ is conditionally definite negative (CDN). Recall that an $N \times N$ symmetric matrix $\mathbf{A}$ is said to be CDN if and only if $\mathbf{y}^\top \mathbf{A}\, \mathbf{y} \leq (\geq) 0$ for all $\mathbf{y} \in \mathbb{R}^N$ such as $\sum_{i=1}^{N} y_i = 0$.

Proposition 4.1 (RQE and SVDD: Case where D is CDN). *If $\mathbf{D}$ is CDN, then RQE portfolio diversification optimization problem is a hard-margin SVDD problem or a soft-margin SVDD problem with slack variables (or empirical errors) equal to zero i.e. $\xi_i = 0,\ \forall\, i = 1, ..., N$ or the parameter C is greater than one i.e. $C \geq 1$.*

Proof (of Proposition 4.1). Suppose that $\mathbf{D}$ is CDN. Then, from Lau (1985, Theorem 2.1), Pekalska and Duin (2005, Theorem 3.13) and Pekalska and Duin (2005, Definition 3.15), there is $p \leq N$ such that $\mathbf{D}$ can be embedded in a Euclidean space $(\mathbb{R}^p, \|.\|_2)$. More formally, there is $p \leq N$ and a set of points $\widetilde{\mathcal{R}} = \{\widetilde{\mathbf{r}}_1, ..., \widetilde{\mathbf{r}}_N\}$, $\widetilde{\mathbf{r}}_i \in \mathbb{R}^p$ such that

$$d_{ij} = \left\| \widetilde{\mathbf{r}}_i - \widetilde{\mathbf{r}}_j \right\|_2^2. \tag{43}$$

It follows that problem (3) can be rewritten as follows

$$\max_{\mathbf{w} \in \mathbb{W}} \frac{1}{2} \sum_{i,j=1}^{N} w_i\, w_j \left\| \widetilde{\mathbf{r}}_i - \widetilde{\mathbf{r}}_j \right\|_2^2 \tag{44}$$

Equation (44) can be rewritten as follows

$$\max_{\mathbf{w} \in \mathbb{W}} \sum_{i=1}^{N} w_i\, \widetilde{\mathbf{r}}_i^\top \widetilde{\mathbf{r}}_i - \mathbf{w}^\top \widetilde{\mathbf{r}}^\top \widetilde{\mathbf{r}}\, \mathbf{w}, \tag{45}$$

■

where $\widetilde{\mathbf{r}} = (\widetilde{\mathbf{r}}_1, ..., \widetilde{\mathbf{r}}_N)$ is an $p \times N$ matrix. First, it follows from problems (11)-(13) and (23)-(25) that problem (3) is equivalent to that of a hard-margin SVDD or a soft-margin SVDD with $\xi_i = 0,\ \forall\, i = 1, ..., N$ or with $C \geq 1$.

The RQE problem can also be defined as a SVDD problem in Euclidean space $\left(\mathbb{W}, \sqrt{D_{H_{\mathbf{D}}}(.,.)}\right)$, where $\sqrt{D_{H_{\mathbf{D}}}(.,.)}$ is an Euclidean distance (see Rao and Nayak, 1985) in $\mathbb{W}$ such that

$$D_{H_{\mathbf{D}}}(\mathbf{w}_1, \mathbf{w}_2) = 2H_{\mathbf{D}}\left(\mathbf{w}_1, \mathbf{w}_2\right) - H_{\mathbf{D}}(\mathbf{w}_1) - H_{\mathbf{D}}(\mathbf{w}_2). \tag{46}$$

Proposition 4.2 (RQE and SVDD in the Euclidean Space $\left(\mathbb{W}, \sqrt{D_{H_{\mathbf{D}}}(.,.)}\right)$). *If* $\mathbf{D}$ *is CDN, then RQE portfolio diversification optimization problem is a hard-margin SVDD problem or a soft-margin SVDD problem with slack variables (or empirical errors) equal to zero i.e.* $\xi_i = 0,\ \forall\, i = 1, ..., N$ *or the parameter* C *greater than one i.e.* $C \geq 1$ *in the space* $\left(\mathbb{W}, \sqrt{D_{H_{\mathbf{D}}}(.,.)}\right)$.

Proof (of Proposition 4.2). Let $\widetilde{\delta} = \{\boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_N\}$ be a set of N points or objects of $\mathbb{W}$. Consider the SVDD on $\widetilde{\delta}$

$$\min_{(\overline{\boldsymbol{\delta}}, \zeta^2) \in \mathbb{W} \times \mathbb{R}_+} \quad \zeta^2 \tag{47}$$

$$\text{s.t} \quad D_{H_{\mathbf{D}}}\left(\overline{\boldsymbol{\delta}}, \boldsymbol{\delta}_i\right) \leq \zeta^2,\ \ \forall\, i = 1, ..., N. \tag{48}$$

Let us prove that the SVDD problem (47)-(48) is equivalent to the RQE problem (). The Lagrangian of (47)-(48) is

$$L(\overline{\boldsymbol{\delta}}, \zeta^2, \boldsymbol{\nu}) = \zeta^2 + \sum_{i=1}^{N} \nu_i \left(D_{H_{\mathbf{D}}} \left(\overline{\boldsymbol{\delta}}, \boldsymbol{\delta}_i \right) - \zeta^2 \right). \tag{49}$$

where ν_i is the Lagrangian multiplier of constraint (48). The first derivatives of $L(\overline{\boldsymbol{\delta}}, \zeta^2, \boldsymbol{\nu})$ with respect of the variables ζ^2 and $\overline{\boldsymbol{\delta}}$ are respectively

$$\sum_{i=1}^{N} \nu_i = 1 \tag{50}$$

$$\overline{\boldsymbol{\delta}} = \boldsymbol{\nu} = \sum_{i=1}^{N} \nu_i \boldsymbol{\delta}_i \tag{51}$$

Reporting (50) and (51) into $L(\overline{\boldsymbol{\delta}}, \zeta^2, \boldsymbol{\nu})$, we obtain the dual problem

$$\max_{\boldsymbol{\nu} \in \mathbb{W}} \sum_{i=1}^{N} \nu_i D_{H_{\mathbf{D}}} \left(\overline{\boldsymbol{\delta}}, \boldsymbol{\delta}_i \right). \tag{52}$$

It is straightforward to verify that (52) is equivalent to

$$\max_{\boldsymbol{\nu} \in \mathbb{W}} H_{\mathbf{D}}(\boldsymbol{\nu}). \tag{53}$$

■

Now, consider the case where $\mathbf{D}$ is not CDN. The RQE problem can also be viewed as a SVDD problem, but in the pseudo-Euclidean space $\left(\mathbb{R}^{(p,q)}, \|.\|_{\mathcal{E}}\right)$.

Proposition 4.3 (RQE and SEB Problems in the Pseudo-Euclidean Space). *If* $\mathbf{D}$ *is not CDN, then RQE portfolio selection optimization problem is a hard-margin SVDD problem or a soft-margin SVDD problem with slack variables (or empirical errors) equal to zero i.e.* $\xi_i = 0$, $\forall\, i = 1, ..., N$ *or the parameter* C *greater than one i.e.* $C \geq 1$ *in a pseudo-Euclidean space* $\left(\mathcal{E} = \mathbb{R}^{(p,q)}, \|.\|_{\mathcal{E}}\right)$.

Proof (of Proposition 4.3). Suppose that $\mathbf{D}$ is not CDN. In that case, form Pekalska et al. (2001) $\mathbf{D}$ can be embedding in a pseudo Euclidean space. More formally, from (see Pekalska and Duin, 2005), there is $p, q \in \mathbb{N}$ and a set of points $\widetilde{\mathcal{R}} = \{\widetilde{\mathbf{r}}_1, ..., \widetilde{\mathbf{r}}_N\}$, $\widetilde{\mathbf{r}}_i \in \mathbb{R}^{p+q}$

such that

$$d_{ij} = \left\|\widetilde{\mathbf{r}}_i^+ - \widetilde{\mathbf{r}}_j^+\right\|_2^2 - \left\|\widetilde{\mathbf{r}}_i^- - \widetilde{\mathbf{r}}_j^-\right\|_2^2, \tag{54}$$

where $\widetilde{\mathbf{r}}_i = (\widetilde{\mathbf{r}}^+, \widetilde{\mathbf{r}}^-)$ with $\widetilde{\mathbf{r}}_i^+ \in \mathbb{R}^p$ and $\widetilde{\mathbf{r}}_i^- \in \mathbb{R}^q$. The RQE problem can then be rewritten as follows

$$\max_{\mathbf{w}\in\mathbb{W}} \frac{1}{2}\sum_{i,j=1}^{N} w_i\, w_j \left\|\widetilde{\mathbf{r}}_i^+ - \widetilde{\mathbf{r}}_j^+\right\|_2^2 - \frac{1}{2}\sum_{i,j=1}^{N} w_i\, w_j \left\|\widetilde{\mathbf{r}}_i^- - \widetilde{\mathbf{r}}_j^-\right\|_2^2. \tag{55}$$

Now, let us show that the problem (55) is a SVDD problem in the pseudo-Euclidean space $\left(\mathbb{R}^{(p,q)}, \|.\|_{\mathcal{E}}\right)$. We follow Huang et al. (2017); Ong et al. (2004). Consider the SVDD problem in $\left(\mathbb{R}^{(p,q)}, \|.\|_{\mathcal{E}}\right)$

$$\min_{\left(\overline{\widetilde{\mathbf{r}}}^+, \overline{\widetilde{\mathbf{r}}}^-, \zeta^2\right)\in\mathbb{R}^p\times\mathbb{R}^q\times\mathbb{R}_+} \zeta^2 \tag{56}$$

$$\text{s.t} \quad \left\|\overline{\widetilde{\mathbf{r}}}^+ - \widetilde{\mathbf{r}}_i^+\right\|_2^2 - \left\|\overline{\widetilde{\mathbf{r}}}^- - \widetilde{\mathbf{r}}_i^-\right\|_2^2 \leq \zeta^2, \ \ \forall\, i = 1, ..., N. \tag{57}$$

The Lagrangian of (56)-(57) is

$$L\left(\overline{\widetilde{\mathbf{r}}}^+, \overline{\widetilde{\mathbf{r}}}^-, \zeta^2, \boldsymbol{\nu}\right) = \zeta^2 - \sum_{i=1}^{N}\nu_i\left(\zeta^2 - \left\|\overline{\widetilde{\mathbf{r}}}^+ - \widetilde{\mathbf{r}}_i^+\right\|_2^2 + \left\|\overline{\widetilde{\mathbf{r}}}^- - \widetilde{\mathbf{r}}_i^-\right\|_2^2\right). \tag{58}$$

The first derivatives of $L\left(\overline{\widetilde{\mathbf{r}}}^+, \overline{\widetilde{\mathbf{r}}}^-, \zeta^2, \boldsymbol{\nu}\right)$ with respect of the variables ζ^2 and $\overline{\widetilde{\mathbf{r}}}^+, \overline{\widetilde{\mathbf{r}}}^-$ are

$$\sum_{i=1}^{N}\nu_i = 1 \tag{59}$$

$$\overline{\widetilde{\mathbf{r}}}^+ = \sum_{i=1}^{N}\nu_i\widetilde{\mathbf{r}}_i^+ \tag{60}$$

$$\overline{\widetilde{\mathbf{r}}}^- = \sum_{i=1}^{N}\nu_i\widetilde{\mathbf{r}}_i^- \tag{61}$$

Reporting (59) and (60) into $L\left(\overline{\widetilde{\mathbf{r}}}^+, \overline{\widetilde{\mathbf{r}}}^-, \zeta^2, \boldsymbol{\nu}\right)$, we obtain the dual problem

$$\max_{\boldsymbol{\nu}\in\mathbb{W}} \sum_{i=1}^{N}\nu_i\left(\left\|\overline{\widetilde{\mathbf{r}}}^+ - \widetilde{\mathbf{r}}_i^+\right\|_2^2 - \left\|\overline{\widetilde{\mathbf{r}}}^- - \widetilde{\mathbf{r}}_i^-\right\|_2^2\right), \tag{62}$$

which can be proved straightforward equivalent to RQE problem (4). ∎

Propositions 4.1 to 4.3 show that RQE portfolio optimization problem is the dual representation of a hard-margin SVDD problem or a soft-margin SVDD problem with slack variables (or empirical errors) equal to zero i.e. $\xi_i = 0, \ \forall\, i = 1, ..., N$ or the parameter C greater than one i.e. $C \geq 1$ in a Euclidean space or a pseudo-Euclidean space.

5 Discussion

Our findings (Propositions 4.1 to 4.3) have important implications for both the machine learning (SVDD problem) and the RQE portfolio diversification. On the one hand, the SVDD problem and its rich set of extensions are relevant for portfolio selection. This strengthens the believe that machine learning can play an important role in finance, in particular in risk management. On the other hand, our findings provide new insights on RQE portfolio diversification approach. First, they show that RQE portfolio diversification problem can be interpreted as a one-class classification in which the training set of object is the universe of assets. The support vectors represent assets with high diversification potential and the non support vectors represent assets with low diversification potential. As a result, only the assets held by the optimal diversified portfolio of RQE (hereafter Rao's Quadratic Entropy Portfolio and shortly RQEP) are required to describe the complete assets universe. It follows that leaving some assets out of RQEP does not entail information costs, but offers the advantage of transaction costs reduction. However, all is true only if the dissimilarity matrix $\mathbf{D}$ is adequately specified. This confirms how important or crucial the choice of $\mathbf{D}$ is for RQE.

Second, our findings show that RQE portfolio diversification optimization problem can be solved using or adapting the vast algorithms of the SVDD and smallest enclosing ball problem (see Fischer et al., 2003; Hansen et al., 2010; Kumar et al., 2003; Liu et al., 2016; Loosli et al., 2016; Nielsen and Nock, 2009; Wang et al., 2011; Yildirim, 2008). This helps reduce significantly the computational cost for a large universe of assets.

Third, our findings generalize and better express the result of Carmichael et al. (2015, Proposition 5.1 and Definition 5.1) on the diversification criteria behind RQE portfolio diversification. Indeed, Carmichael et al. (2015) show that diversification in RQEP is expressed in terms of distance between the single-asset portfolios and RQEP in space $\left(\mathbb{W}, \sqrt{D_{H_{\mathbf{D}}}(.,.)}\right)$: RQEP is equidistant from all assets that belong to it. Assets that do not belong to RQEP are those closer than the equidistant assets. They call this diversification criterion the *equally portfolio-asset distance*. Propositions 4.1 and 4.3 extend this result by showing that the same diversification criteria is obtained when we consider the Euclidean space $(\mathbb{R}^p, \|.\|_2)$ and pseudo-Euclidean space $\left(\mathcal{E} = \mathbb{R}^{(p,q)}, \|.\|_{\mathcal{E}}\right)$.

Fourth, in the presence of empirical errors in the matrix $\mathbf{D}$ or outliers, our findings suggest that RQE portfolio diversification optimization problem can be replaced by the following robust one

$$\max_{\mathbf{w}} H_{\mathbf{D}}(\mathbf{w}) \tag{63}$$

$$\mathbf{w}^\top \mathbf{1} = 1 \tag{64}$$

$$0 \leq w_i \leq C, \ \forall\, i = 1, ..., N \tag{65}$$

with $\frac{1}{N} \leq C < 1$. The problem (63)-(65) represents the weight constraint version of (4). It is well-known in the portfolio selection literature that upper and/or lower weight constraint improves portfolio performance reducing estimation errors (see Behr et al., 2013; Li, 2015; Ma and Jagannathan, 2003). Therefore the problem (63)-(65) can be considered as a natural approach to handle estimation errors in RQE problem. Following the SVDD literature (see Section 3.1), the parameter C can be calibrated such that

$$C = \frac{1}{N\varphi} \tag{66}$$

where $\varphi \in (0, 1]$ represents the upper-bound of the proportion of outliers or asset affected by estimation errors. The fraction φ can be identified via the distribution of $D_i = \sum_{j=1}^{N} d_{ij}$, the total dissimilarity of asset i. Note that C can also be calibrated in the case where errors are expected following the portfolio selection literature (see Ma and Jagannathan, 2003).

Remark 1. *In the problem (63)-(65), we have only upper weight constraint. The lower weight constraint of the RQE diversification problem can also be interpreted in the light of the one-class classification following the negative SVDD of Tax (2001) and Tax and Duin (2004).*

Remark 2. *In the problem (63)-(65), the weight constraint parameter C is constant for all* *assets. The weight constraint parameter C can also be defined for each asset as*

$$\max_{\mathbf{w}} H_{\mathbf{D}}(\mathbf{w}) \tag{67}$$

$$\mathbf{w}^\top \mathbf{1} = 1 \tag{68}$$

$$0 \leq w_i \leq C_i, \ \forall\, i = 1, ..., N, \tag{69}$$

where C_i is the weight constraint parameter of asset i. The problem (67)-(69) can also be interpreted in the light of the one-class classification. We refer the readers to Cha et al. (2014); Chen et al. (2015); Hamidzadeh et al. (2017); Liu et al. (2013) for more details.

Fifth, our findings also show that RQE diversification based on the dissimilarity matrix

$$d_{ij} = \|\widehat{\mathbf{r}}_i - \widehat{\mathbf{r}}_j\|_2^2 \tag{70}$$

will not robust, where $\widehat{\mathbf{r}}_i = \mathbf{r}_i - \overline{r}_i$, where $\overline{r}_i = \frac{\sum_{t=1}^{T} r_{it}}{T}$. It is sensitive to scaling and will provide poor diversification, when data is not spherical distributed, which is the case for asset returns.

To overcome these drawbacks, one can follow the kernel SVDD (see Section 3.2) solving the kernel RQE optimization problem

$$\max_{\mathbf{w}} \ \sum_{i=1}^{N} w_i\, k\left(\widehat{\mathbf{r}}_i, \widehat{\mathbf{r}}_i\right) - \sum_{i,j=1}^{N} w_i\, w_j\, k\left(\widehat{\mathbf{r}}_i, \widehat{\mathbf{r}}_j\right) \tag{71}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} w_i = 1 \tag{72}$$

$$0 \leq w_i \leq C \ \ \forall\, i = 1, ..., N \tag{73}$$

with $\mathbf{r}_i$ adequately scaling. The problem (71)-(73) can be rewritten as follow

$$\max_{\mathbf{w}} \ H_{\mathbf{D}_k}(\mathbf{w}) \tag{74}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} w_i = 1 \tag{75}$$

$$0 \leq w_i \leq C \ \ \forall\, i = 1, ..., N \tag{76}$$

where $H_{\mathbf{D}_k}(\mathbf{w})$ is the kernel RQE and $\mathbf{D}_k = (d_{k,ij})_{i,j=1}^{N}$ is the kernel dissimilarity matrix defined as

$$d_{k,ij} = k\left(\widehat{\mathbf{r}}_i, \widehat{\mathbf{r}}_i\right) + k\left(\widehat{\mathbf{r}}_i, \widehat{\mathbf{r}}_j\right) - 2k\left(\widehat{\mathbf{r}}_i, \widehat{\mathbf{r}}_j\right). \tag{77}$$

The problem (74)-(76) shows therefore that kernel SVDD provides new directions for the choice of the dissimilarity matrix in the use of RQE as portfolio diversification measure.

As mentioned in Section 3.2, the challenge in the implementation of (74)-(76) is the choice of the kernel function and the calibration of its hyper parameters. In the literature of the OCC and the economics (see Exterkate, 2013; Exterkate et al., 2016), the Gaussian and the polynomial kernel with $\theta \neq 0$ are the two most used kernel function. Thus, we limit our analysis to this two kernel functions. For other examples, readers are refereed to Scholkopf (1993), Smola et al. (1998), Kamath et al. (2010) and Bavaud (2011).

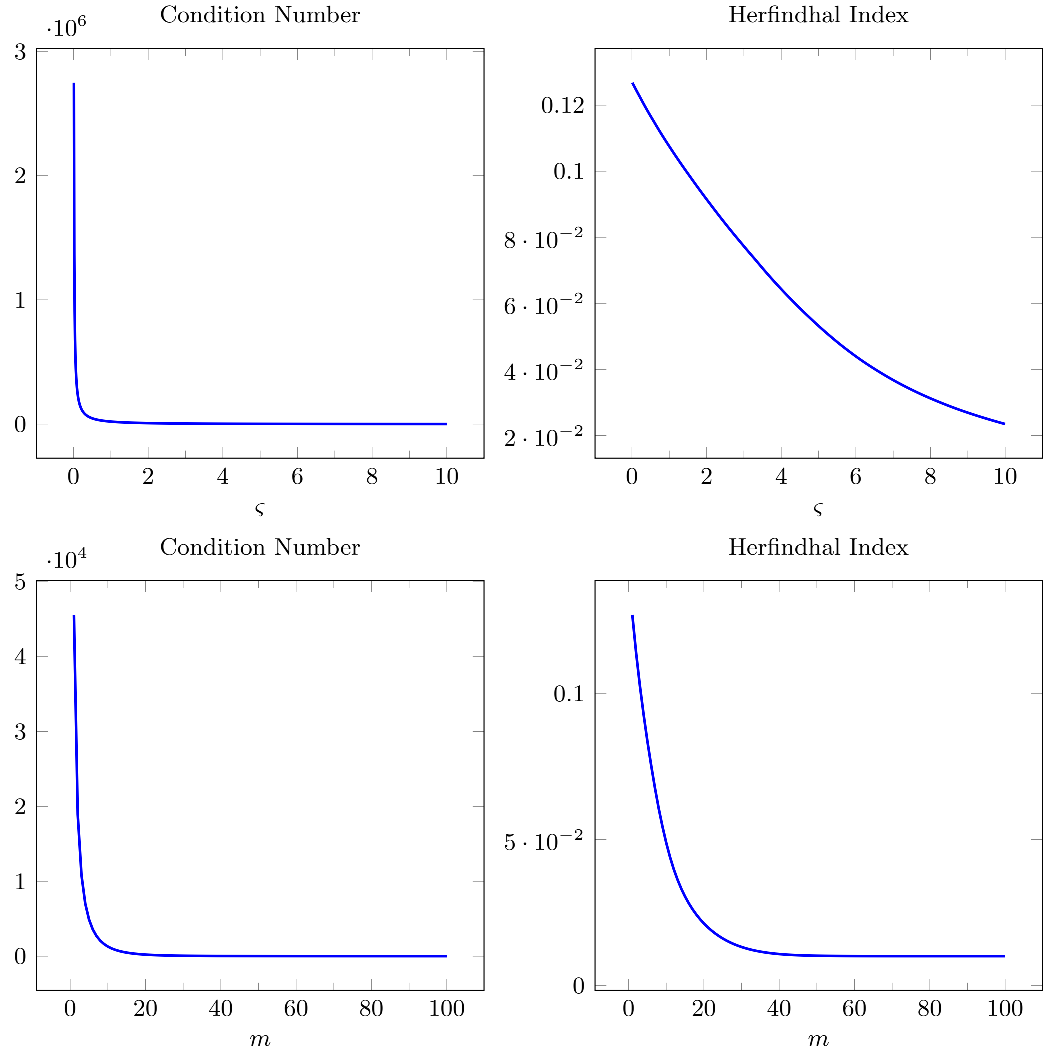
In the polynomial kernel there are two parameters to calibrated: θ and m. While in the Gaussian kernel there is one parameter to calibrate ς. In the literature θ is generally calibrate to equal to one i.e. $\theta = 1$. This calibration is adequate only if the elements of the matrix of covariances and 1 have the same scale, which is usually not the case. Thus, we suggest to calibrate θ such that $\theta = \max_{i,j} |\sigma_{ij}| = \sigma_{\max}$, where $|.|$ is the absolute value operator. In the OCC literature, the hyper parameters m and ς are calibrated optimizing a specific performance metric (see Tax and Duin, 2004; Tax and Juszczak, 2003; Tax, 2001) using the cross-validation approach, or to achieve good or tight description (see Chaudhuri et al., 2017; Deng and Xu, 2007; Evangelista et al., 2007; Kakde et al., 2017; Nader et al., 2014; Peredriy et al., 2016; Wang et al., 2012; Xiao et al., 2015, 2014) . In the case where one considers that a good or tight description is equivalent to a good diversification, all these approaches remain valid in the case of portfolio selection. However this not generally the case. A tight description generally implies moderate or high support vectors size which is not generally equivalent to good diversification. Thus, we now suggest alternative approaches.

A natural approach, similarly in Exterkate (2013); Exterkate et al. (2016) and Tax (2001), is to select m and ς optimizing a performance metric related to portfolio diversification. In the case of the out-of-sample exercise the k-fold cross validation approach is recommended.

The parameters m and ς can also be tuning exploiting their regularization or shrinkage properties. To see this, we depict the condition number (CN) of $\mathbf{K}$ and the Herfindahl index

(HI), $HI(\mathbf{w}) = \sum_{i=1}^{N} w_i^2$, of the optimal portfolio of the kernel RQE problem (74)-(76), for values of m range from 1 to 100 with 1 step and ς ranges from 0.01 to 10 with 0.01 step (see Figure 4). In the two cases we normalize the data by asset volatility such that in the case of Gaussian kernel $\mathbf{K} = \exp(-\varsigma\,(2 - \boldsymbol{\rho}))$, and in the case of polynomial kernel $\mathbf{K} = (1 + \boldsymbol{\rho})^m$. The data used to generate the correlation matrix $\boldsymbol{\rho}$ is the Fama-French 100 portfolios formed on size and book-to-market. The data is a daily data starting in 1 July 1969 and ending in 31 May 2018. As we can observe from Figure 4, the condition number and the Herfindahl index are decreasing functions of ς and m. This suggests that the kernel RQE problem (74)-(76) becomes more robust and well-conditioned when ς and m become large. As a consequence, the two parameters can be interpreted as regularization parameters. The solution of the problem (74)-(76), denoted $\mathbf{w}^{KRQE}$, is therefore the regularization of the solution of the problem (4), $\mathbf{w}^{RQE}$. They can also be interpreted as shrinkage intensity. For example, when ς converges to 0, $\frac{1-\exp(-\varsigma\,\mathbf{D})}{\varsigma}$ converges to $\mathbf{D}$. On the other hand, when ς converges to ∞, $\frac{1-\exp(-\varsigma\,\mathbf{D})}{\varsigma}$ converges to the equidistant dissimilarity matrix $\frac{\mathbf{1}\mathbf{1}^\top - \mathbf{I}}{\varsigma}$. The parameter ς shrinks $\mathbf{D}$ towards $\frac{\mathbf{1}\mathbf{1}^\top - \mathbf{I}}{\varsigma}$.

Figure 4: Condition numbers and Herfindahl index depending on kernel parameters ς and m



6 Conclusion

The focus of this paper is to establish the link between Rao's Quadratic Entropy (RQE) portfolio optimization and the support vector data description (SVDD) problems. We show that the RQE portfolio optimization problem is a dual representation of a SVDD problem in a Euclidean space or in a pseudo-Euclidean space (Propositions 4.1 to 4.3). Our findings demonstrate that machine learning can play an important role in finance, in particular in risk management. Moreover, it provides new insights on RQE portfolio diversification approach in terms of interpretation, understanding, implementation, specification and optimization.

In conclusion, machine learning can play an important role in risk management and RQE is an adequate framework to quantify and to manage portfolio diversification.

References

(2018). "Call for Papers: Special Issue on 'AI and machine learning in finance'," *Quantitative Finance*, 18, 533.

Abu-Mostafa, Y. S. and A. F. Atiya (1996). "Introduction to financial forecasting," *Applied Intelligence*, 6, 205–213.

Antunes, F., B. Ribeiro, and F. Pereira (2017). "Probabilistic modeling and visualization for bankruptcy prediction," *Applied Soft Computing*.

Atkins, A., M. Niranjan, and E. Gerding (2018). "Financial news predicts stock market volatility better than close price," *The Journal of Finance and Data Science*.

Ban, G.-Y., N. El Karoui, and A. E. Lim (2016). "Machine learning and portfolio optimization," *Management Science*.

Bartkowiak, A. M. (2011). "Anomaly, novelty, one-class classification: a comprehensive introduction," *International Journal of Computer Information Systems and Industrial Management Applications*, 3, 61–71.

Bavaud, F. (2011). "On the Schoenberg Transformations in Data Analysis: Theory and Illustrations," *Journal of Classification*, 28, 297–314.

Behr, P., A. Guettler, and F. Miebs (2013). "On portfolio optimization: Imposing the right constraints," *Journal of Banking & Finance*, 37, 1232–1242.

Booth, D. G. and E. F. Fama (1992). "Diversification Returns and Asset Contributions," *Financial Analysts Journal*, 48, 26–32.

Burnaev, E. and D. Smolyakov (2016). "One-class SVM with privileged information and its application to malware detection," in *Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference on*, IEEE, 273–280.

Carmichael, B., G. Koumou, and K. Moran (2015). "Unifying Portfolio Diversification Measures Using Rao's Quadratic Entropy," .

Carmichael, B., G. B. Koumou, and K. Moran (2018). "Rao's quadratic entropy and maximum diversification indexation," *Quantitative Finance*, 18, 1017–1031.

Cha, M., J. S. Kim, and J.-G. Baek (2014). "Density weighted support vector data description," *Expert Systems with Applications*, 41, 3343 – 3350.

Champely, S. and D. Chessel (2002). "Measuring Biological Diversity Using Euclidean Metrics," *Environmental and Ecological Statistics*, 9, 167–177.

Chaudhuri, A., D. Kakde, C. Sadek, L. Gonzalez, and S. Kong (2017). "The Mean and Median Criterion for Kernel Bandwidth Selection for Support Vector Data Description," *arXiv preprint arXiv:1708.05106*.

Chen, G., X. Zhang, Z. J. Wang, and F. Li (2015). "Robust support vector data description for outlier detection with noise or uncertain data," *Knowledge-Based Systems*, 90, 129 – 137.

Chen, L., B. Fang, and Z. Shang (2016). "Software fault prediction based on one-class SVM," in *2016 International Conference on Machine Learning and Cybernetics (ICMLC)*, vol. 2, 1003–1008.

Choueifaty, Y. and Y. Coignard (2008). "Toward Maximum Diversification," *Journal of Portfolio Management*, 35, 40–51.

Deng, H. and R. Xu (2007). "Model selection for anomaly detection in wireless ad hoc networks," in *Computational Intelligence and Data Mining, 2007. CIDM 2007. IEEE Symposium on*, IEEE, 540–546.

Dunis, C. L., P. W. Middleton, A. Karathanasopolous, and K. Theofilatos (2016). *Artificial Intelligence in Financial Markets Cutting Edge Applications for Risk Management, Portfolio Optimization and Economics*, New Developments in Quantitative Trading and Investment, London: Palgrave Macmillan UK.

El Azami, M., C. Lartizien, S. Canu, et al. (2014). "Robust outlier detection with L0-SVDD." in *ESANN*.

Elzinga, D. J. and D. W. Hearn (1972a). "The Minimum Covering Sphere Problem," *Management Science*, 19, 96–104.

Elzinga, J. and D. W. Hearn (1972b). "Geometrical Solutions for Some Minimax Location Problems," *Transportation Science*, 6, 379–394.

Evangelista, P. F., M. J. Embrechts, and B. K. Szymanski (2007). "Some properties of the Gaussian kernel for one class learning," in *International Conference on Artificial Neural Networks*, Springer, 269–278.

Evans, J. L. and S. H. Archer (1968). "Diversification and the reduction of dispersion: an empirical analysis," *The Journal of Finance*, 23, 761–767.

Exterkate, P. (2013). "Model selection in kernel ridge regression," *Computational Statistics & Data Analysis*, 68, 1–16.

Exterkate, P., P. J. Groenen, C. Heij, and D. van Dijk (2016). "Nonlinear forecasting with many predictors using kernel ridge regression," *International Journal of Forecasting*, 32, 736–753.

Feng, Z., J. Fu, D. Du, F. Li, and S. Sun (2017). "A new approach of anomaly detection in wireless sensor networks using support vector data description," *International Journal of Distributed Sensor Networks*, 13, 1550147716686161.

Fernholz, R. and B. Shay (1982). "Stochastic Portfolio Theory and Stock Market Equilibrium," *The Journal of Finance*, 37, 615–624.

Fischer, K., B. Gärtner, and M. Kutz (2003). *Fast Smallest-Enclosing-Ball Computation in High Dimensions*, Berlin, Heidelberg: Springer Berlin Heidelberg, 630–641.

Fischer, T. and C. Krauss (2017). "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*.

Forghani, Y., H. S. Yazdi, S. Effati, and R. S. Tabrizi (2011). "Support vector data description by using hyper-ellipse instead of hyper-sphere," in *Computer and Knowledge Engineering (ICCKE), 2011 1st International eConference on*, IEEE, 22–27.

Gangolf, C., R. Dochow, G. Schmidt, and T. Tamisier (2014). "SVDD: A proposal for automated credit rating prediction," in *2014 International Conference on Control, Decision and Information Technologies (CoDIT)*, 048–053.

Gartner, B. (1999). "Fast and Robust Smallest Enclosing Balls," in *Proceedings of the 7th Annual European Symposium on Algorithms*, London, UK, UK: Springer-Verlag, ESA '99, 325–338.

Goetzmann, W. N., L. Li, and K. G. Rouwenhorst (2005). "Long-Term Global Market Correlations," *The Journal of Business*, 78, 1–38.

Gyorfi, L., G. Ottucsak, and H. Walk (2012). *Machine learning for financial engineering*, vol. 8, World Scientific.

Hamidzadeh, J., R. Sadeghi, and N. Namaei (2017). "Weighted Support Vector Data Description based on Chaotic Bat Algorithm," *Applied Soft Computing*.

Hansen, M. S., K. Sjöstrand, and R. Larsen (2010). "On the regularization path of the support vector domain description," *Pattern Recognition Letters*, 31, 1919–1923.

Huang, X., A. Maier, J. Hornegger, and J. A. Suykens (2017). "Indefinite kernels in least squares support vector machines and principal component analysis," *Applied and Computational Harmonic Analysis*, 43, 162–172.

Juszczak, P. (2006). "Learning to recognise: A study on one-class classification and active learning," .

Juszczak, P., D. Tax, and R. P. Duin (2002). "Feature scaling in support vector data description," in *Proc. ASCI*, Citeseer, 95–102.

Kakde, D., A. Chaudhuri, S. Kong, M. Jahja, H. Jiang, and J. Silva (2017). "Peak criterion for choosing Gaussian kernel bandwidth in support vector data description," in *Prognostics and Health Management (ICPHM), 2017 IEEE International Conference on*, IEEE, 32–39.

Kallberg, L. and T. Larsson (2013). "Faster Approximation of Minimum Enclosing Balls by Distance Filtering and GPU Parallelization," *Journal of Graphics Tools*, 17, 67–84.

Kamath, U., A. Shehu, and K. A. De Jong (2010). "Feature and kernel evolution for recognition of hypersensitive sites in DNA sequences," in *International Conference on Bio-Inspired Models of Network, Information, and Computing Systems*, Springer, 213–228.

KENNEDY, K., B. MAC NAMEE, AND S. J. DELANY (2009). “Credit scoring: solving the low default portfolio problem using one-class classification,” .

KHAN, S. S. AND M. G. MADDEN (2009). “A survey of recent trends in one class classification,” in *Irish Conference on Artificial Intelligence and Cognitive Science*, Springer, 188–197.

——— (2014). “One-class classification: taxonomy of study and review of techniques,” *The Knowledge Engineering Review*, 29, 345–374.

KIM, K.-J. (2003). “Financial time series forecasting using support vector machines,” *Neurocomputing*, 55, 307–319.

KOUMOU, N. B. G. (2017). “Rao’s Quadratic Entropy, Risk Management and Portfolio Theory,” .

KRAUSS, C., X. A. DO, AND N. HUCK (2017). “Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500,” *European Journal of Operational Research*, 259, 689–702.

KUMAR, P., J. S. B. MITCHELL, AND E. A. YILDIRIM (2003). “Approximate Minimum Enclosing Balls in High Dimensions Using Core-sets,” *J. Exp. Algorithmics*, 8.

LARSSON, T. AND L. KALLBERG (2013). “Fast and Robust Approximation of Smallest Enclosing Balls in Arbitrary Dimensions,” *Computer Graphics Forum*, 32, 93–101.

LAU, K.-S. (1985). “Characterization of Rao’s Quadratic Entropies,” *Sankhy?: The Indian Journal of Statistics, Series A (1961-2002)*, 47, 295–309.

LEE, H., D. MOON, I. KIM, H. JUNG, AND D. PARK (2015). “Anomaly Intrusion Detection Based on Hyper-ellipsoid in the Kernel Feature Space.” *KSII Transactions on Internet & Information Systems*, 9.

LEE, K., D.-W. KIM, K. H. LEE, AND D. LEE (2007). “Density-induced support vector data description,” *IEEE Transactions on Neural Networks*, 18, 284–289.

LEE, S.-W., J. PARK, AND S.-W. LEE (2006). “Low resolution face recognition based on support vector data description,” *Pattern Recognition*, 39, 1809–1812.

LEE, Y.-C. (2007). "Application of support vector machines to corporate credit rating prediction," *Expert Systems with Applications*, 33, 67–74.

LI, B. AND S. C. HOI (2014). "Online portfolio selection: A survey," *ACM Computing Surveys (CSUR)*, 46, 35.

LI, J. (2015). "Sparse and stable portfolio selection with parameter uncertainty," *Journal of Business & Economic Statistics*, 33, 381–392.

LI, Y., J. WU, AND H. BU (2016). "When quantitative trading meets machine learning: A pilot survey," in *Service Systems and Service Management (ICSSSM), 2016 13th International Conference on*, IEEE, 1–6.

LIEW, J. K.-S. AND B. MAYSTER (2017). "Forecasting ETFs with Machine Learning Algorithms," *The Journal of Alternative Investments*, 20, 58–78.

LIN, W. Y., Y. H. HU, AND C. F. TSAI (2012). "Machine Learning in Financial Crisis Prediction: A Survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42, 421–436.

LIU, B., Y. XIAO, L. CAO, Z. HAO, AND F. DENG (2013). "SVDD-based outlier detection on uncertain data," *Knowledge and Information Systems*, 34, 597–618.

LIU, B., Y. XIAO, S. Y. PHILIP, Z. HAO, AND L. CAO (2014). "An efficient approach for outlier detection with imperfect data labels," *IEEE Transactions on Knowledge and Data Engineering*, 26, 1602–1616.

LIU, Y. AND T. XIE (2018). "Machine learning versus econometrics: prediction of box office," *Applied Economics Letters*, 1–7.

LIU, Y.-F., R. DIAO, F. YE, AND H.-W. LIU (2016). "An Efficient Inexact Newton-CG Algorithm for the Smallest Enclosing Ball Problem of Large Dimensions," *Journal of the Operations Research Society of China*, 4, 167–191.

LOOSLI, G., S. CANU, AND C. S. ONG (2016). "Learning SVM in Krein Spaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38, 1204–1216.

Ma, J. and S. Perkins (2003). "Time-series novelty detection using one-class support vector machines," in *Neural Networks, 2003. Proceedings of the International Joint Conference on*, IEEE, vol. 3, 1741–1745.

Ma, T. and R. Jagannathan (2003). "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps," *Journal of Finance*, 58, 57–72.

Manevitz, L. M. and M. Yousef (2001). "One-class SVMs for document classification," *Journal of Machine Learning Research*, 2, 139–154.

Markowitz, H. (1952). "Portfolio Selection," *The Journal of Finance*, 7, 77–91.

Meucci, A. (2009). "Managing Diversification," *Risk*, 22, 74–79.

Meucci, A., A. Santangelo, and R. Deguest (2014). "Measuring Portfolio Diversification Based on Optimized Uncorrelated Factors," SSRN Working Paper.

Mourao-Miranda, J., D. R. Hardoon, T. Hahn, A. F. Marquand, S. C. Williams, J. Shawe-Taylor, and M. Brammer (2011). "Patient classification as an outlier detection problem: an application of the one-class support vector machine," *Neuroimage*, 58, 793–804.

Mullainathan, S. and J. Spiess (2017). "Machine learning: an applied econometric approach," *Journal of Economic Perspectives*, 31, 87–106.

Mygdalis, V., A. Iosifidis, A. Tefas, and I. Pitas (2017). "Semi-supervised subclass support vector data description for image and video classification," *Neurocomputing*.

Nader, P., P. Honeine, and P. Beauseroy (2014). "l_p-norms in One-Class Classification for Intrusion Detection in SCADA Systems," *IEEE Transactions on Industrial Informatics*, 10, 2308–2317.

——— (2016). "Detection of cyberattacks in a water distribution system using machine learning techniques," in *2016 Sixth International Conference on Digital Information Processing and Communications (ICDIPC)*, 25–30.

Nanni, L. (2006). "Experimental comparison of one-class classifiers for online signature verification," *Neurocomputing*, 69, 869–873.

Nava, N., T. D. Matteo, and T. Aste (2018). "Financial Time Series Forecasting Using Empirical Mode Decomposition and Support Vector Regression," *Risks*, 6, 7.

Nayak, T. K. (1986a). "An analysis of Diversity Using Rao's Quadratic Entropy," *Indian Journal of Statistics*, 48, 315–330.

——— (1986b). "Sampling Distributions in Analysis of Diversity," *Indian Journal of Statistics*, 48, 1–9.

Nayak, T. K. and J. L. Gastwirth (1989). "The Use of Diversity Analysis to Assess the Relative Influence of Factors Affecting the Income Distribution," *Journal of Business & Economic Statistics*, 7, 453–460.

Nielsen, F. and R. Nock (2009). "Approximating Smallest Enclosing Balls with Application to Machine Learning," *International Journal of Computational Geometry & Applications*, 19, 389–414.

Ong, C. S., X. Mary, S. Canu, and A. J. Smola (2004). "Learning with non-positive kernels," in *Proceedings of the twenty-first international conference on Machine learning*, ACM, 81.

Pang, S., S. Li, and J. Xiao (2014). "Application of the algorithm based on the PSO and improved SVDD for the personal credit rating," *Journal of Financial Engineering*, 01, 1450037.

Park, J., D. Kang, J. Kim, J. T. Kwok, and I. W. Tsang (2007). "SVDD-based pattern denoising," *Neural computation*, 19, 1919–1938.

Pavoine, S. (2005). "Methodes Statistiques pour la Mesure de la Biodiversite," Ph.D. thesis, Universite Claude Bernard-Lyon I.

——— (2012). "Clarifying and developing analyses of biodiversity: towards a generalisation of current approaches," *Methods in Ecology and Evolution*, 3, 509–518.

Pavoine, S. and M. B. Bonsall (2009). "Biological diversity: Distinct distributions can lead to the maximization of Rao's quadratic entropy," *Theoretical Population Biology*, 75, 153–163.

Pavoine, S., S. Ollier, and D. Pontier (2005). "Measuring diversity from dissimilarities with Rao's quadratic entropy : Are any dissimilarities suitable?" *Theoretical Population Biology*, 67, 231–239.

Pekalska, E. and R. P. W. Duin (2005). *The Dissimilarity Representation for Pattern Recognition: Foundations And Applications (Machine Perception and Artificial Intelligence)*, World Scientific Publishing Company.

Pekalska, E., P. Paclik, and R. P. Duin (2001). "A Generalized Kernel Approach to Dissimilarity Based Classification," *Journal of Machine Learning Research*, 2, 175–211.

Peredriy, S., D. Kakde, and A. Chaudhuri (2016). "Kernel Bandwidth Selection for SVDD: Peak Criterion Approach for Large Data," *arXiv preprint arXiv:1611.00058.*

Rao, R. (1982a). "Diversity : Its Measurement, Decomposition, Apportionment and Analysis," *Indian Journal of Statistics*, 44, 1–22.

Rao, R. and T. K. Nayak (1985). "Cross Entropy, Dissimilarity Measures, and Characterizations of Quadratic Entropy," *Information Theory, IEEE Transactions*, 31, 589–593.

Rao, R. C. (1982b). "Diversity and Dissimilarity Coefficients: A Unified Approach," *Theoretical Population Biology*, 21, 24–43.

Ricotta, C. and L. Szeidl (2006). "Towards a unifying approch to diversity measures : Bridging the gap between the Shannon entropy and Rao's quadratic index," *Theoretical Population Biology*, 70, 237–243.

Rudin, A. M. and J. Morgan (2006). "A Portfolio Diversification Index," *The Journal of Portfolio Management*, 32, 81–89.

Sadeghi, R. and J. Hamidzadeh (2016). "Automatic support vector data description," *Soft Computing*, 1–12.

Schiilkop, P., C. Burgest, and V. Vapnik (1995). "Extracting support data for a given task," in *Proceedings, First International Conference on Knowledge Discovery & Data Mining. AAAI Press, Menlo Park, CA*, 252–257.

Scholkopf, B. (1993). "The kernel trick for distances," in *TR MSR 2000-51, Microsoft Research*, 5–3.

Scholkopf, B., S. Mika, C. J. Burges, P. Knirsch, K.-R. Muller, G. Ratsch, and A. J. Smola (1999a). "Input space versus feature space in kernel-based methods," *IEEE transactions on neural networks*, 10, 1000–1017.

Schölkopf, B., J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson (2001). "Estimating the Support of a High-Dimensional Distribution," *Neural Comput.*, 13, 1443–1471.

Scholkopf, B., R. C. Williamson, A. Smola, and J. Shawe-Taylor (1999b). "SV Estimation of a Distribution's Support," .

Sharpe, W. F. (1972). "Risk, Market Sensitivity and Diversification," *Financial Analysts Journal*, 28, 74–79.

Shawe-Taylor, J. and N. Cristianini (2004). *Kernel methods for pattern analysis*, Cambridge university press.

Smola, A. J., B. Scholkopf, and K.-R. Muller (1998). "The connection between regularization operators and support vector kernels," *Neural Networks*, 11, 637 – 649.

Statman, M. and J. Scheid (2005). "Global Diversification," *Journal of Investment Management*, 3, 1–11.

Stirling, A. (2010). "Multicriteria diversity analysis: A novel heuristic framework for appraising energy portfolios," *Energy Policy*, 38, 1622–1634.

Sylvester, J. J. (1857). "A Question in the Geometry of Situation," *Quarterly Journal of Pure and Applied Mathematics*, 1.

Tao, H., L. Yun, W. Ke, X. Jian, and L. Fu (2016). "A new weighted SVDD algorithm for outlier detection," in *Control and Decision Conference (CCDC), 2016 Chinese*, IEEE, 5456–5461.

Tax, D. M. and R. P. Duin (2004). "Support Vector Data Description," *Machine Learning*, 54, 45–66.

Tax, D. M. and P. Juszczak (2003). "Kernel whitening for one-class classification," *International Journal of Pattern Recognition and Artificial Intelligence*, 17, 333–347.

TAX, D. M. J. (2001). "One-class classification: Concept-learning in the absence of counter-examples," Ph.D. thesis, Technische Universiteit Delft.

TAY, F. E. AND L. CAO (2001). "Application of support vector machines in financial time series forecasting," *omega*, 29, 309–317.

VAPNIK, V. (1998). *Statistical learning theory. 1998*, Wiley, New York.

VERMORKEN, M. A., F. R. MEDDA, AND T. SCHRODER (2012). "The Diversification Delta : A Heigher-Moment Measure for Portfolio Diversification," *Journal Of Portfolio Management*, 39, 67–74.

WANG, H., L. ZHANG, Y. XIAO, AND W. XU (2012). "An approach to choosing gaussian kernel parameter for one-class svms via tightness detecting," in *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2012 4th International Conference on*, IEEE, vol. 2, 318–323.

WANG, K. AND H. XIAO (2017). "Ellipsoidal data description," *Neurocomputing*, 238, 328–339.

WANG, Q., L. S. LOPES, AND D. M. TAX (2004). "Visual object recognition through one-class learning," in *International Conference Image Analysis and Recognition*, Springer, 463–470.

WANG, X., F.-L. CHUNG, AND S. WANG (2011). "Theoretical analysis for solution of support vector data description," *Neural Networks*, 24, 360–369.

WESTON, J., B. SCHÖLKOPF, AND G. H. BAKIR (2004). "Learning to find pre-images," in *Advances in neural information processing systems*, 449–456.

WHITE, H. (1988). "Economic prediction using neural networks: the case of IBM daily stock returns," in *IEEE 1988 International Conference on Neural Networks*, 451–458 vol.2.

WOERHEIDE, W. AND D. PERSSON (1993). "An Index of Portfolio Diversification," *Financial Services Review*, 2, 73–85.

XIAO, Y., H. WANG, AND W. XU (2015). "Parameter selection of Gaussian kernel for one-class SVM," *IEEE transactions on cybernetics*, 45, 941–953.

Xiao, Y., H. Wang, L. Zhang, and W. Xu (2014). "Two methods of selecting Gaussian kernel parameters for one-class SVM and their application to fault detection," *Knowledge-Based Systems*, 59, 75–84.

Xuanthanh, V., T. Bach, H. A. Le Thi, and T. P. Dinh (2017). "Ramp Loss Support Vector Data Description," in *Asian Conference on Intelligent Information and Database Systems*, Springer, 421–431.

Yildirim, E. A. (2008). "Two Algorithms for the Minimum Enclosing Ball Problem," *SIAM Journal on Optimization*, 19, 1368–1391.

Yin, F. and G.-X. Huang (2011). "Improved density-induced support vector data description," in *Machine Learning and Cybernetics (ICMLC), 2011 International Conference on*, IEEE, vol. 2, 747–750.

Zhao, Y. and D. N. Naik (2012). "Hypothesis testing with Rao's quadratic entropy and its application to Dinosaur biodiversity," *Journal of Applied Statistics*, 39, 1667–1680.

Zhou, G., K.-C. Tohemail, and J. Sun (2005). "Efficient Algorithms for the Smallest Enclosing Ball Problem," *Computational Optimization and Applications*, 30, 147–160.

Zhou, R., R. Cai, and G. Tong (2013). "Applications of Entropy in Finance: A Review," *Entropy*, 15, 4909–4931.