Classification using proximity catch digraphs

Artür Manukyan[1] · Elvan Ceyhan[2]

Abstract

We employ random geometric digraphs to construct semi-parametric classifiers. These data-random digraphs belong to parameterized random digraph families called proximity catch digraphs (PCDs). A related geometric digraph family, class cover catch digraph (CCCD), has been used to solve the class cover problem by using its approximate minimum dominating set and showed relatively good performance in the classification of imbalanced data sets. Although CCCDs have a convenient construction in $\mathbb{R}^d$, finding their minimum dominating sets is NP-hard and their probabilistic behaviour is not mathematically tractable except for $d = 1$. On the other hand, a particular family of PCDs, called *proportional-edge* PCDs (PE-PCDs), has mathematically tractable minimum dominating sets in $\mathbb{R}^d$; however their construction in higher dimensions may be computationally demanding. More specifically, we show that the classifiers based on PE-PCDs are prototype-based classifiers such that the exact minimum number of prototypes (equivalent to minimum dominating sets) is found in polynomial time on the number of observations. We construct two types of classifiers based on PE-PCDs. One is a family of hybrid classifiers that depends on the location of the points of the training data set, and another type is a family of classifiers solely based on class covers. We assess the classification performance of our PE-PCD based classifiers by extensive Monte Carlo simulations, and compare them with that of other commonly used classifiers. We also show that, similar to CCCD classifiers, our classifiers tend to be robust to the class imbalance in classification as well.

Keywords Class cover problem · Delaunay tessellation · Digraph · Domination · Prototype selection · Support estimation

✉ Artür Manukyan
 artur-man@hotmail.com

 Elvan Ceyhan
 ceyhan@auburn.edu

[1] Department of Industrial and Systems Engineering, Yeditepe University, 34755 Istanbul, Turkey

[2] Department of Mathematics and Statistics, College of Sciences and Mathematics, Auburn University, Auburn, AL 36849, USA

1 Introduction

Classification methods based on set covering algorithms received considerable attention recently because of their use in prototype selection (Bien and Tibshirani 2011; Cannon and Cowen 2004; Angiulli 2012). Prototypes are selected members of a data set so as to attain various tasks including reducing, condensing or summarizing a data set. Many learning methods aim to carry out more than one of these tasks, thereby building efficient learning algorithms (Pękalska et al. 2006; Bien and Tibshirani 2011). A desirable prototype set reduces the data set in order to decrease running time, condenses the data set to preserve information, and summarizes the data set for better exploration and understanding. The methods we discuss in this work are considered as decision boundary generators where decisions are made based on class conditional regions, or *class covers*, that are composed of a collection of convex sets, each associated with a prototype (Toussaint 2002). The union of such convex sets constitutes a region for the class of interest, estimating the support of this class (Schölkopf et al. 2001). Support estimates have uses in both supervised and unsupervised learning schemes offering solutions to many problems in machine learning (Marchette 2004). We propose supervised learning methods, or classifiers, based on these estimates of the supports constructed with a random geometric digraph family called *proximity catch digraphs*.

Proximity Catch Digraphs (PCDs) are closely related to Class Cover Catch Digraphs (CCCDs) introduced by Priebe et al. (2001), and are vertex-random digraphs defined by the relationship between class-labeled observations. Priebe et al. (2001) introduced CCCDs to find graph theoretic solutions to the Class Cover Problem (CCP), and provided some results on the minimum dominating sets and the distribution of the domination number of such digraphs for one dimensional data. The goal of CCP is to find a set of hyperspheres (usually Euclidean balls) such that their union encapsulates, or *covers*, (a subset of) the training data set associated with a particular class, called the target class (Cannon and Cowen 2004). In addition, Priebe et al. (2003a) showed that approximate dominating sets of CCCDs, which were obtained by a greedy algorithm, can be used to establish efficient semi-parametric classifiers. Moreover, DeVinney et al. (2002) defined random walk CCCDs (RW-CCCDs) where balls of class covers are defined in a relaxed manner compared to the previously introduced CCCDs so as to avoid the overfitting problem. These digraphs have been used, e.g. in face detection (Eveland et al. 2005) and in latent class discovery for gene expression data (Priebe et al. 2003b). CCCDs also show robustness to the class imbalance in data sets (Manukyan and Ceyhan 2016). Class imbalance often occurs in real data sets; that is, some classes of the data sets have a large number of members whereas the remaining classes only have fewer, resulting in a bias towards the majority class (the class with abundant number of members) which drastically decreases the classification performance.

Class covers with Euclidean balls have been extended to allow the use of different type of regions in order to cover a class of interest. Serafini (2014) uses sets of boxes to find a cover of classes, and also defines the maximum redundancy problem. This is an optimization problem of covering as many points as possible by each box where the total number of boxes is kept to a (approximately) minimum. Hammer et al. (2004) investigates CCP using boxes with applications to the logical data analysis. Moreover, Bereg et al. (2012) extend covering boxes to rectilinear polygons to cover classes, and they report on the complexity of the CCP algorithms using such polygonal covering regions. Takigawa et al. (2009) incorporate balls and establish classifiers similar to the ones based on CCCDs, and they

also use sets of convex hulls. Ceyhan (2005) uses sets of triangles relative to the tessellation of the opposite class to analytically compute the minimum number of triangles required to establish a class cover. In this work, we study class covers with particular triangular regions (simplicial regions in higher dimensions).

CCCDs can be generalized using *proximity maps* (Jaromczyk and Toussaint 1992). Ceyhan (2005) defined PCDs and introduced three families of PCDs and investigated the distribution of the domination number of such digraphs in a two class setting. Domination number and, another graph invariant called the arc density (the ratio of number of arcs in a digraph to the total number of arcs possible) of these PCDs has been used for testing spatial patterns of segregation and association (Ceyhan and Priebe 2005; Ceyhan et al. 2006, 2007). In this article, we employ PCDs in statistical classification and investigate their performance. The PCDs of concern in this work are based on a particular family of proximity maps called *proportional-edge* (PE) proximity maps. The corresponding PCDs are called PE-PCDs, and are defined for target class (i.e. the class of interest) points inside the convex hull of non-target class points (Ceyhan 2005). However, this construction ignores the target class points outside the convex hull of the non-target class. We mitigate this shortcoming by partitioning the region outside of the convex hull into unbounded regions, called *outer simplices*, which may be viewed as extensions of outer intervals in $\mathbb{R}$ (e.g. intervals with infinite endpoints) to higher dimensions. We attain proximity regions in these outer simplices by extending PE proximity maps to outer simplices. We establish two types of classifiers based on PE-PCDs, namely, *hybrid* and *cover* classifiers. The first type incorporates the PE-PCD covers of only points in the convex hull and use other classifiers for points outside the convex hull of the non-target class, hence we have some kind of a hybrid classifier; the second type is further based on two class cover models where the first is a mixture of PE-PCDs and CCCDs (composite covers) whereas the second is purely based on PE-PCDs (standard covers).

One common property of most class covering (or set covering) methods is that none of the algorithms find the exact minimum number of covering sets in polynomial time, and solutions are mostly provided by approximation algorithms (Vazirani 2001). However, for PE-PCDs, the exact minimum number of covering sets (equivalent to prototype sets) can be found much faster; that is, the exact minimum solution is found in a running time polynomial in size of the data set but exponential in dimensionality. PE-PCDs have computationally tractable (exact) minimum dominating sets in $\mathbb{R}^d$ (Ceyhan 2010). Since the complexity of class covers based on this family of proximity maps exponentially increases with dimensionality, we apply dimension reduction methods (such as principal components analysis) to substantially reduce the number of features and thus reduce the dimensionality. Hence, based on the transformed data sets in the reduced dimensions, the PE-PCD based hybrid classifiers and, in particular, cover classifiers become more appealing in terms of both prototype selection and classification performance (in the reduced dimension). We use simulated and real data sets to show that these two types of classifiers based on PE-PCDs have either comparable or slightly better classification performance than other classifiers when the data sets exhibit the class imbalance problem.

The article is organized as follows: in Sect. 2, we introduce some auxiliary tools for defining PCDs, and in particular in Sect. 3, we describe PE proximity regions and PE-PCDs. In Sect. 4, we introduce two types of class cover models that are called composite and standard covers. In Sect. 5, we establish two types of statistical classifiers based on PE-PCDs which are called hybrid and cover PE-PCD classifiers. The latter type is defined for both class cover models described in Sect. 4. In Sect. 6, we assess the performance of PE-PCD classifiers and compare them with existing methods (such as kNN and support vector

machine classifiers) on simulated data sets. Finally, in Sect. 7, we assess our classifiers on real data sets, and in Sect. 8, we present discussion and conclusions as well as some future research directions.

2 Tessellations in $\mathbb{R}^d$ and auxiliary tools

In this section, we introduce the tools required for constructing PE-PCD classifiers. Let $(\Omega, \mathcal{M})$ be a measurable space, and let the training data set be composed of two non-empty sets, $\mathcal{X}_0$ and $\mathcal{X}_1$ with sample sizes $n_0 := |\mathcal{X}_0|$ and $n_1 := |\mathcal{X}_1|$ from classes 0 and 1, respectively. Also let $\mathcal{X}_0$ and $\mathcal{X}_1$ be sets of Ω-valued random variables with class conditional distributions F_0 and F_1, with supports $s(F_0)$ and $s(F_1)$, respectively. We develop rules to define proximity maps and regions for the points from the class of interest, i.e. *target class*, which is class j with respect to the *Delaunay tessellation* of the points from the class of non-interest, i.e. *non-target class* which is class $1 - j$ for $j = 0, 1$.

A tessellation in $\mathbb{R}^d$ is a collection of non-intersecting (or intersecting only on boundaries) convex d-polytopes such that their union covers a region. We partition $\mathbb{R}^d$ into non-intersecting d-simplices and d-polytopes to construct PE-PCDs that tend to have multiple disconnected components. We show that such a partitioning of the domain provides digraphs with computationally tractable minimum dominating sets. In addition, we use the barycentric coordinate system to characterize the target class points with respect to the Delaunay tessellation of the non-target class. Such a coordinate system simplifies the definitions of many tools associated with PE-PCD classifiers in $\mathbb{R}^d$, including minimum dominating sets of PE-PCDs and convex distance functions which are defined in Sect. 5.1.

2.1 Delaunay tessellation of $\mathbb{R}^d$

The convex hull of the non-target class points $C_H(\mathcal{X}_{1-j})$ can be partitioned into *Delaunay cells* through the Delaunay tessellation of $\mathcal{X}_{1-j} \subset \mathbb{R}^d$. For $d = 1$, the Delaunay tessellation is an intervalization (i.e., integer partition) of the the convex hull of $\mathcal{X}_{1-j}$ which is $C_H(\mathcal{X}_{1-j}) = \left(\min(X_{1-j}), \max(X_{1-j}) \right)$, where the middle intervals (i.e. intervals in the convex hull) are based on the order statistics of $\mathcal{X}_{1-j}$, and end intervals are $\left(-\infty, \min(X_{1-j}) \right)$ and $\left(\max(X_{1-j}), \infty \right)$. For $d = 2$, the Delaunay tessellation becomes a triangulation which partitions $C_H(\mathcal{X}_{1-j})$ into non intersecting triangles. For the points in the general position, the triangles in the Delaunay triangulation satisfy the property that the circumcircle of a triangle contains no points from $\mathcal{X}_{1-j}$ except for the vertices of the triangle. In higher dimensions, Delaunay cells are d-simplices (for example, a tetrahedron in $\mathbb{R}^3$). Hence, the $C_H(\mathcal{X}_{1-j})$ is the union of a set of disjoint d-simplices $\{\mathcal{S}_k\}_{k=1}^K$ where K is the number of d-simplices, or Delaunay cells. Each d-simplex has $d + 1$ non-coplanar vertices where none of the remaining points of $\mathcal{X}_{1-j}$ are in the interior of the circumsphere of the simplex (and the vertices of the simplex are points from $\mathcal{X}_{1-j}$). Hence, simplices of the Delaunay tessellations are more likely to be acute (simplices with no substantially small inner angles). Note that Delaunay tessellation is the dual of the *Voronoi diagram* of the set $\mathcal{X}_{1-j}$. A Voronoi diagram is a partitioning of $\mathbb{R}^d$ into convex polytopes such that the points inside each polytope is closer to the point associated with the polytope than any other point in $\mathcal{X}_{1-j}$. Hence, a polytope $V(y)$ associated with a point $y \in \mathcal{X}_{1-j}$ is defined as

$$V(\mathsf{y}) = \{v \in \mathbb{R}^d \,:\, \|v - \mathsf{y}\| \leq \|v - z\| \ \text{ for all } z \in \mathcal{X}_{1-j} \backslash \{\mathsf{y}\}\}.$$

Here, $\| \cdot \|$ stands for the usual Euclidean norm. Observe that the Voronoi diagram is unique for a fixed set of points. A Delaunay graph is constructed by joining the pairs of points in $\mathcal{X}_{1-j}$ whose boundaries of Voronoi polytopes intersect. The edges of the Delaunay graph constitute a partitioning of $C_H(\mathcal{X}_{1-j})$, hence the Delaunay tessellation. By the uniqueness of the Voronoi diagram, the Delaunay tessellation is also unique (except for cases where $d + 1$ or more points lie on the same hypersphere).

A Delaunay tessellation partitions only $C_H(\mathcal{X}_{1-j})$ and does not offer a partitioning of the complement $\mathbb{R}^d \backslash C_H(\mathcal{X}_{1-j})$ unlike the Voronoi diagrams. As we will see in the following sections, this drawback makes the definition of our semi-parametric classifiers more difficult. Let *facets* of $C_H(\mathcal{X}_{1-j})$ be the simplices on the boundary of $C_H(\mathcal{X}_{1-j})$. To partition $\mathbb{R}^d \backslash C_H(\mathcal{X}_{1-j})$, we define unbounded regions associated with each facet of $C_H(\mathcal{X}_{1-j})$, namely *outer simplices* in $\mathbb{R}^d$ or *outer triangles* in $\mathbb{R}^2$. Each outer simplex is constructed by a single facet of $C_H(\mathcal{X}_{1-j})$, denoted by $\mathcal{F}_l$ for $l = 1, \dots, L$ where, L is the number of boundary facets and, note that, each facet is a $(d-1)$-simplex. Let $\{P_1, P_2, \dots, P_N\} \subseteq \mathcal{X}_{1-j}$ be the set of points on the boundary of $C_H(\mathcal{X}_{1-j})$, and let $C_M := \sum_{i=1}^{N} p_i / N$ be the center of mass of $C_H(\mathcal{X}_{1-j})$. We use the bisector rays of Deng and Zhu (1999) as a framework for constructing outer simplices, however, such rays are not well defined for convex hulls in $\mathbb{R}^d$ for $d > 2$. Let the ray emanating from C_M through P_i be denoted as $\overrightarrow{C_M P_i}$. Hence, we define the outer simplices by rays emanating from each boundary vertex P_i to outside of $C_H(\mathcal{X}_{1-j})$ in the direction of $\overrightarrow{C_M P_i}$. Each facet $\mathcal{F}_l$ has d boundary points adjacent to it, and the rays associated with these boundary points establish an unbounded region together with the facet $\mathcal{F}_l$. Such a region can be viewed as an infinite "drinking glass" with $\mathcal{F}_l$ being the bottom while top of the glass reaching infinity, similar to the end intervals in $\mathbb{R}$ with one end being infinity. Let $\mathscr{F}_l$ denote the outer simplex associated with the facet $\mathcal{F}_l$. An illustration of a Delaunay triangulation and the corresponding outer triangles in $\mathbb{R}^2$ is given in Fig. 1 where $C_H(\mathcal{X}_{1-j})$ has six facets, hence $\mathbb{R}^2 \backslash C_H(\mathcal{X}_{1-j})$ is partitioned into six disjoint unbounded regions.
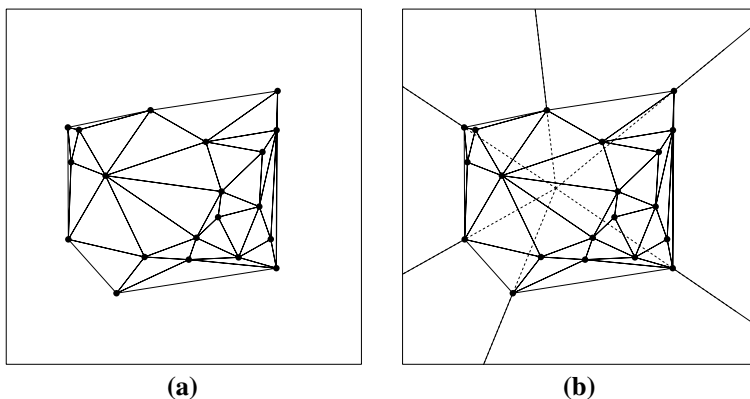

Fig. 1 **a** A Delaunay triangulation of 20 points $\mathcal{X}_{1-j} \subset \mathbb{R}^2$, partitioning $C_H(\mathcal{X}_{1-j})$. **b** The Delaunay tessellation of $\mathcal{X}_{1-j}$ with rays $\overrightarrow{C_M P_i}$ for $i = 1, \dots, 6$ that yield a partitioning of $\mathbb{R}^2 \backslash C_H(\mathcal{X}_{1-j})$. The dashed lines illustrate the direction of these rays where they meet at the point center of mass of $C_H(\mathcal{X}_{1-j})$

2.2 Barycentric coordinate system

The *barycentric coordinate system* was introduced by A.F. Möbius in his book "The Barycentric Calculus" in 1837. The idea is to assign weights w_1, w_2 and w_3 to points $\mathbf{y}_1$, $\mathbf{y}_2$ and $\mathbf{y}_3$ which constitute the vertices of a triangle T in $\mathbb{R}^2$, respectively (Ungar 2010). Hence, the center of mass, or the *barycenter* of $\mathbf{y}_1$, $\mathbf{y}_2$ and $\mathbf{y}_3$ is , for $w_1 + w_2 + w_3 \neq 0$, is given by

$$P = \frac{w_1\mathbf{y}_1 + w_2\mathbf{y}_2 + w_3\mathbf{y}_3}{w_1 + w_2 + w_3}. \tag{1}$$

Similarly, let $\mathcal{S} = \mathcal{S}(\mathcal{Y})$ be a d-simplex defined by the $d+1$ non-coplanar points $\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{d+1}\} \subset \mathbb{R}^d$ with weights $(w_1, w_2, \ldots, w_{d+1})$. Thus, the barycenter $W \in \mathbb{R}^d$ is given by

$$W = \frac{\sum_{i=1}^{d+1} w_i\mathbf{y}_i}{\sum_{i=1}^{d+1} w_i} \quad \text{with} \quad \sum_{i=1}^{d+1} w_i \neq 0. \tag{2}$$

The $(d+1)$-tuple $\mathbf{w} = (w_1, w_2, \ldots, w_{d+1})$ (also denoted as $(w_1 : w_2 : \ldots : w_{d+1})$) can also be viewed as a set of coordinates of W with respect to the (vertex) set $\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{d+1}\}$ for $d > 0$. Hence, the name *barycentric coordinates*. Observe that W in Eq. (2) is scale invariant (i.e. invariant under scaling of the weights of W). Therefore, the set of barycentric coordinates is homogeneous, i.e., for any $\lambda \in \mathbb{R}_+$,

$$(w_1, w_2, \ldots, w_{d+1}) = (\lambda w_1, \lambda w_2, \ldots, \lambda w_{d+1}). \tag{3}$$

This gives rise to *normalized barycentric coordinates* $\mathbf{w}' = (w'_1, w'_2, \ldots, w'_{d+1})$ of a point $x \in \mathbb{R}^d$ with respect to $\mathcal{Y}$ as follows:

$$\sum_{i=1}^{d+1} w'_i = \sum_{i=1}^{d+1} \frac{w_i}{w_{tot}} = 1, \tag{4}$$

where $w_{tot} := \sum_{j=1}^{d+1} w_j$. For simplicity, we use the normalized barycentric coordinates as "barycentric coordinates" throughout this work, and use $\mathbf{w}$ to denote the vector of the coordinates of x. That is, x is $(w_1, w_2, \ldots, w_{d+1})$ in barycentric coordinates so that $\sum_{i=1}^{d+1} w_i = 1$. The vector $\mathbf{w}$ has a unique solution given the linear systems of equations

$$\mathbf{Aw} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_{d+1} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{d+1} \end{bmatrix} = \begin{bmatrix} 1 \\ x \end{bmatrix}. \tag{5}$$

where the vectors $\mathbf{y}_k - \mathbf{y}_1 \in \mathbb{R}^d$ for $k = 2, \ldots, d+1$ are linearly independent (Lawson 1986). The vector $\mathbf{w}$ is unique but w_i are not necessarily in $(0, 1)$. Barycentric coordinates define whether the point x is in $\mathcal{S}(\mathcal{Y})$ or not, as follows:

- $x \in \mathcal{S}(\mathcal{Y})^o$ if $w_i \in (0, 1)$ for all $i = 1, \ldots, d+1$: the point x is inside of the d-simplex $\mathcal{S}(\mathcal{Y})$ where $\mathcal{S}(\mathcal{Y})^o$ denotes the interior of $\mathcal{S}(\mathcal{Y})$,
- $x \in \flat(\mathcal{S}(\mathcal{Y}))$, the point x is on the boundary of $\mathcal{S}(\mathcal{Y})$, if $w_i = 0$ and $w_j = (0, 1]$ for some i in I such that $I \subsetneq \{1, \ldots, d+1\}$ and $j \in \{1, \ldots, d+1\}\setminus I$,

- $x = y_i$ if $w_i = 1$ and $w_j = 0$ for all $i = 1, \ldots, d + 1$ and $j \neq i$: the point x is at a vertex of $S(\mathcal{Y})$,
- $x \notin S(\mathcal{Y})$ if $w_i \notin [0, 1]$ for some $i \in \{1, \ldots, d + 1\}$: the point x is outside of $S(\mathcal{Y})$.

Barycentric coordinates of a point $x \in S(\mathcal{Y})$ can also be viewed as the convex combination of the points of $\mathcal{Y}$, the vertices on the boundary of $S(\mathcal{Y})$.

2.3 *M*-vertex regions

A d-simplex is the smallest convex polytope in $\mathbb{R}^d$ constructed by a set of non-coplanar vertices $\mathcal{Y} = \{y_1, y_2, \ldots, y_{d+1}\}$. The boundary of a d-simplex consists of k-simplices called k-faces for $0 \leq k < d$. Each k-face is a simplex defined by a subset of $\mathcal{Y}$ with k elements, hence there are $\binom{d+1}{k+1}$ k-faces in a d-simplex. Let $S(\mathcal{Y})$ be the simplex defined by the set of points $\mathcal{Y}$. Given a *simplex center $M \in S(\mathcal{Y})^o$* (e.g. a triangle center in $\mathbb{R}^2$), there are $d + 1$ M-vertex regions constructed by the set $\mathcal{Y}$. The M-vertex region of the vertex y_i is denoted by $R_M(y_i)$ for $i = 1, 2, \ldots, d + 1$.

For $i = 1, \ldots, d + 1$, let f_i denote the $(d - 1)$-face opposite to the vertex y_i. Observe that the lines through the points y_i and M cross the face f_i, a $(d - 1)$-face, at the points M_i. Similarly, since the face f_i is a $(d - 1)$-simplex with a center M_i for any $i = 1, \ldots, d + 1$, we can find the centers of $(d - 2)$-faces of this $(d - 1)$-simplex. Note that both M_i and M are of the same type of centers of their respective simplices f_i and $S(\mathcal{Y})$. The vertex region $R_M(y_i)$ is the convex hull of the points y_i, $\{M_j\}_{j=1; j \neq i}^{d+1}$, and centers of all k-faces (which are also k-simplices) adjacent to y_i for $k = 1, \ldots, d - 2$. In Fig. 2, we illustrate the vertex regions of an acute triangle in $\mathbb{R}^2$ and the vertex regions $R_M(y_1)$ and $R_M(y_3)$ of a 3-simplex (tetrahedron). In $\mathbb{R}^2$, the 2-simplex is a triangle with vertices $\mathcal{Y} = \{y_1, y_2, y_3\}$ denoted as $T(\mathcal{Y}) = T(y_1, y_2, y_3)$ and the corresponding vertex regions are $R_M(y_1)$, $R_M(y_2)$, and $R_M(y_3)$ (see Fig. 2a, b). Notice that M_i lies on edge e_i which is opposite to vertex y_i for $i = 1, 2, 3$. Observe that, in Fig. 2c and d, each 2-face of this 3-simplex is a 2-simplex (a triangle). For example, in Fig. 2c, the points M_2, M_3 and M_4 are centers of f_2, f_3 and f_4, respectively. Moreover, these 2-simplices also have faces (1-faces or edges of the 3-simplex), and the centers of these faces are $\{M_{ij}\}_{i, j=1; i \neq j}^4$. Hence, the vertex region $R_M(y_1)$ is a convex polytope of points $\{y_1, M, M_2, M_3, M_4, M_{32}, M_{42}, M_{43}\}$ and $R_M(y_3)$ is a convex polytope of points $\{y_3, M, M_2, M_4, M_1, M_{42}, M_{41}, M_{21}\}$. Just as we can write the vertex region as the intersection of two triangles in $\mathbb{R}^2$, we can write the vertex region as intersections of three tetrahedrons in $\mathbb{R}^3$. For example, $R_M(y_1)$ is the intersection of tetrahedrons $T(y_1, M_{42}, y_2, y_4)$, $T(y_1, M_{43}, y_3, y_4)$ and $T(y_1, M_{32}, y_2, y_3)$.

Ceyhan and Priebe (2005) introduced the vertex regions as auxiliary tools to define proximity regions. They also gave the explicit functional forms of these regions as a function of the coordinates of vertices $\{y_1, y_2, y_3\}$. However, we characterize these regions based on barycentric coordinates as given in Proposition 1 and Theorem 1, as this coordinate system is more convenient for computation in higher dimensions.

Proposition 1 *Let $\mathcal{Y} = \{y_1, y_2, y_3\} \subset \mathbb{R}^2$ be a set of three non-collinear points, and let $\{R_M(y_i)\}_{i=1,2,3}$ be the vertex regions that partition $T(\mathcal{Y})$. Then for $x \in T(\mathcal{Y})$ and $M \in T(\mathcal{Y})^o$, we have $x \in R_M(y_i)$ if and only if*
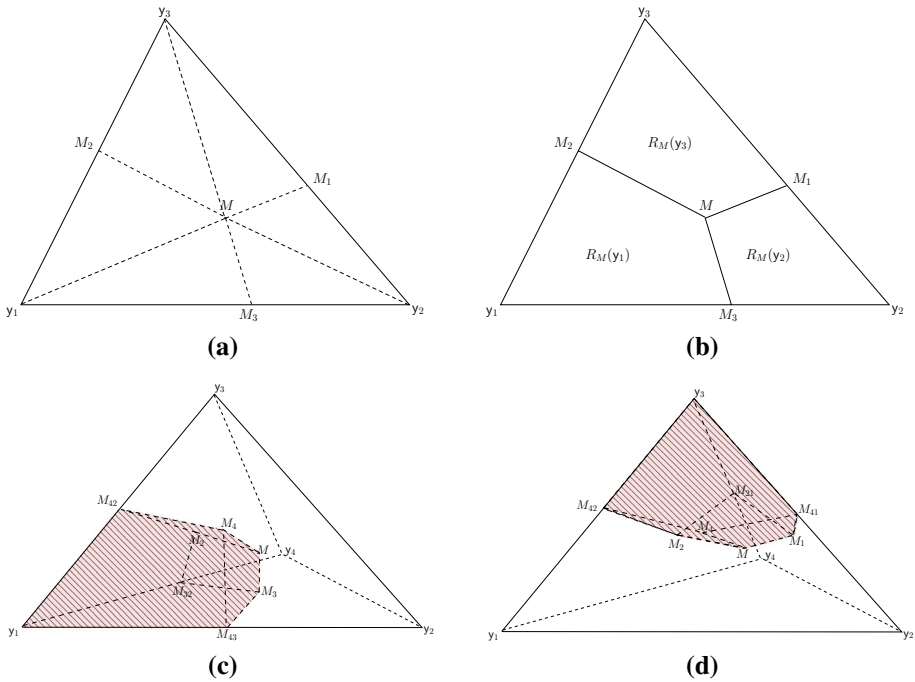
Fig. 2 M-vertex regions of an acute triangle $T(\mathcal{Y}) = T(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)$ with a center $M \in T(\mathcal{Y})^o$. **a** The dashed lines that constitute the vertex regions. **b** M-vertex regions associated with a vertex $\mathbf{y}_i$ for $i = 1, 2, 3$. **c** M-vertex region $R_M(\mathbf{y}_1)$ of vertex $\mathbf{y}_1$ and **d** $R_M(\mathbf{y}_3)$ of vertex $\mathbf{y}_3$ of a 3-simplex, or a tetrahedron. M-vertex regions are shaded in **c** and **d**

$$w_T^{(i)}(x) \geq \max_{\substack{j = 1, 2, 3 \\ j \neq i}} \frac{m_i w_T^{(j)}(x)}{m_j},$$

for $i = 1, 2, 3$ where $\mathbf{w}_T(x) = \left(w_T^{(1)}(x), w_T^{(2)}(x), w_T^{(3)}(x) \right)$ and $\mathbf{m} = (m_1, m_2, m_3)$ are barycentric coordinates of points x and M with respect to the triangle $T(\mathcal{Y})$, respectively.

Proof It is sufficient to show the result for $i = 1$, as $i = 2, 3$ cases would follow similarly. So we will show that, $x \in R_M(\mathbf{y}_1)$ iff

$$w_T^{(1)}(x) \geq \max \left\{ \frac{m_1 w_T^{(2)}(x)}{m_2}, \frac{m_1 w_T^{(3)}(x)}{m_3} \right\}.$$

Let $T_2(\mathcal{Y})$ and $T_3(\mathcal{Y})$ be two triangles formed by sets of points $\{\mathbf{y}_1, \mathbf{y}_2, M_2\}$ and $\{\mathbf{y}_1, \mathbf{y}_3, M_3\}$, respectively. First, we observe that $x \in R_M(\mathbf{y}_1)$ if and only if $x \in T_2(\mathcal{Y}) \cap T_3(\mathcal{Y})$. So, for the forward direction, assume $x \in T_2(\mathcal{Y}) \cap T_3(\mathcal{Y})$. Then $x \in T_2(\mathcal{Y})$ and $x \in T_3(\mathcal{Y})$. Since $x \in T_2(\mathcal{Y})$, we have $x = \alpha_1 \mathbf{y}_1 + \alpha_2 \mathbf{y}_2 + \alpha_3 M_2$, i.e., the barycentric coordinate vector of point x with respect to the triangle $T_2(\mathcal{Y})$ is $\mathbf{w}_{T_2}(x) = (\alpha_1, \alpha_2, \alpha_3)$. But since M_2 lies on edge e_2, we can write it as $M_2 = b\mathbf{y}_1 + (1 - b)\mathbf{y}_3$ for some $b \in (0, 1)$. Then $x = (\alpha_1 + \alpha_3 b)\mathbf{y}_1 + \alpha_2 \mathbf{y}_2 + \alpha_3(1 - b)\mathbf{y}_3$. Hence, by uniqueness of barycentric

coordinates for x with respect to $T(\mathcal{Y})$, it follows that $w_T^{(1)}(x) = \alpha_1 + \alpha_3 b$, $w_T^{(2)}(x) = \alpha_2$ and $w_T^{(3)}(x) = \alpha_3(1 - b)$. Also, since M_2 and M are on the same line which crosses edge e_2, we have $M = c\mathbf{y}_2 + (1 - c)M_2$ for some $c \in (0, 1)$. Then

$$M = c\mathbf{y}_2 + (1 - c)(b\mathbf{y}_1 + (1 - b)\mathbf{y}_3)$$
$$= b(1 - c)\mathbf{y}_1 + c\mathbf{y}_2 + (1 - b)(1 - c)\mathbf{y}_3.$$

Hence, by uniqueness of barycentric coordinates for M with respect to $T(\mathcal{Y})$, it follows that $m_1 = b(1 - c)$, $m_2 = c$, and $m_3 = (1 - b)(1 - c)$. Hence

$$\frac{w_T^{(1)}(x)}{w_T^{(3)}(x)} = \frac{\alpha_1 + \alpha_3 b}{\alpha_3(1 - b)} \geq \frac{b}{1 - b} = \frac{b(1 - c)}{(1 - b)(1 - c)} = \frac{m_1}{m_3}.$$

Then, $x \in T_2(\mathcal{Y})$ iff $w_T^{(1)}(x) \geq (m_1/m_3)w_T^{(3)}(x)$, and similarly, $x \in T_3(\mathcal{Y})$ iff $w_T^{(1)}(x) \geq (m_1/m_2)w_T^{(2)}(x)$. So, $x \in T_2(\mathcal{Y}) \cap T_3(\mathcal{Y}) = R_M(\mathbf{y}_1)$ implies $w_T^{(1)}(x) \geq \max \left\{ \dfrac{m_1 w_T^{(2)}(x)}{m_2}, \dfrac{m_1 w_T^{(3)}(x)}{m_3} \right\}$.

For the reverse direction, for a contradiction, assume that $x \in R_M(\mathbf{y}_1)$ and $w_T^{(1)}(x) < \max \left\{ \dfrac{m_1 w_T^{(2)}(x)}{m_2}, \dfrac{m_1 w_T^{(3)}(x)}{m_3} \right\}$. Without loss of generality, assume that $w_T^{(1)}(x) < \dfrac{m_1 w_T^{(3)}(x)}{m_3}$. Since $x \in T_2(\mathcal{Y})$, as before, we have $\dfrac{w_T^{(1)}(x)}{w_T^{(3)}(z)} = \dfrac{\alpha_1 + \alpha_3 b}{\alpha_3(1 - b)}$ which is less than $\dfrac{m_1}{m_3} = \dfrac{b}{1 - b}$. That is, $\dfrac{\alpha_1 + \alpha_3 b}{\alpha_3(1 - b)} < \dfrac{b}{1 - b}$ which implies $\alpha_1 < 0$ which implies $x \notin T(\mathcal{Y})$ contradicting the assumption that $x \in T_2(\mathcal{Y}) \subset T(\mathcal{Y})$. Thus,

$$R_M(\mathbf{y}_1) = \left\{ x \in T(\mathcal{Y}) : w_T^{(1)}(x) \geq \max \left\{ \frac{m_1 w_T^{(2)}(x)}{m_2}, \frac{m_1 w_T^{(3)}(x)}{m_3} \right\} \right\}.$$

$\square$

Note that, when $M := M_C$ (i.e., M is the centroid or the center of mass of the triangle $T(\mathcal{Y})$), we can further simplify the result of Proposition 1; that is, for any point $x \in T(\mathcal{Y})^o$, we have $x \in R_{M_C}(\mathbf{y}_i)$ if and only if $w_T^{(i)}(x) = \max_{j=1,2,3} w_T^{(j)}(x)$ since the vector of (special) barycentric coordinates of M_C is $\mathbf{m}_C = (1/3, 1/3, 1/3)$. The following theorem is an extension of Proposition 1 to higher dimensions.

Theorem 1 *Let* $\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{d+1}\} \subset \mathbb{R}^d$ *be a set of non-coplanar points for $d > 0$, and let* $\{R_M(\mathbf{y}_i)\}_{i=1}^{d+1}$ *be the M-vertex regions that partition* $\mathcal{S}(\mathcal{Y})$. *Then, for $x \in \mathcal{S}(\mathcal{Y})$ and $M \in \mathcal{S}(\mathcal{Y})^o$, we have $x \in R_M(\mathbf{y}_i)$ if and only if*

$$w_{\mathcal{S}}^{(i)}(x) \geq \max_{\substack{j = 1, \ldots, d+1 \\ j \neq i}} \frac{m_i w_{\mathcal{S}}^{(j)}(x)}{m_j}, \tag{6}$$

where $\mathbf{w}_{\mathcal{S}}(x) = \left(w_{\mathcal{S}}^{(1)}(x), \ldots, w_{\mathcal{S}}^{(d+1)}(x) \right)$ *and* $\mathbf{m} = (m_1, \ldots, m_{d+1})$ *are the barycentric coordinates of points x and M with respect to the simplex* $\mathcal{S}(\mathcal{Y})$, *respectively.*

See "Appendix" for the proof.

As in the triangle case above, for $M = M_C$ and for any point $x \in S(\mathcal{Y})^o$, we have $x \in R_{M_C}(\mathsf{y}_i)$ if and only if $w_T^{(i)}(x) = \max_j w_T^{(j)}(x)$ since the set of barycentric coordinates of M_C is $\mathbf{m}_C = (1/(d+1), 1/(d+1), \ldots, 1/(d+1))$. The M_C-vertex regions are particularly appealing for our proportional-edge proximity regions.

3 Proximity regions and proximity catch digraphs

We consider proximity regions for the (supervised) two-class classification problem, then perform complexity reduction via minimum dominating sets of the associated proximity catch digraphs. For $j = 0, 1$, the *proximity map* $N(\cdot) : \Omega \to 2^\Omega$ associates with each point $x \in \mathcal{X}_j$, a *proximity region* $N(x) \subset \Omega$. Consider the data-random (or vertex-random) proximity catch digraph $D_j = (\mathcal{V}_j, \mathcal{A}_j)$ with vertex set $\mathcal{V}_j = \mathcal{X}_j$ and arc set $\mathcal{A}_j$ defined by $(u, v) \in \mathcal{A}_j \iff \{u, v\} \subset \mathcal{X}_j$ and $v \in N(u)$, for $j = 0, 1$. The digraph D_j depends on the (joint) distribution of the sets of points $\mathcal{X}_0$ and $\mathcal{X}_1$, and on the map $N(\cdot)$. The adjective *proximity*—for the digraph D_j and for the map $N(\cdot)$ — comes from thinking of the region $N(x)$ as representing those points in Ω "closer" to x (Toussaint 1980; Jaromczyk and Toussaint 1992). Our proximity catch digraphs (PCDs) for $\mathcal{X}_j$ against $\mathcal{X}_{1-j}$ are defined by specifying $\mathcal{X}_j$ as the target class and $\mathcal{X}_{1-j}$ as the non-target class. Hence, in the definitions of our PCDs, the only difference is switching the roles of $\mathcal{X}_0$ and $\mathcal{X}_1$. For $j = 0$, 0 becomes the target class label and 1 becomes the non-target class label, and it is vice versa for $j = 1$.

The proximity regions associated with PCDs introduced by Ceyhan and Priebe (2005) are *simplicial* proximity regions (regions that constitute simplices in $\mathbb{R}^d$) defined for the target class points $\mathcal{X}_j$ in the convex hull of the non-target class points $C_H(\mathcal{X}_{1-j})$. However, by introducing the *outer simplices* associated with the facets of $C_H(\mathcal{X}_{1-j})$, we extend the definition of the simplicial proximity regions to $\mathbb{R}^d \backslash C_H(\mathcal{X}_{1-j})$. Such simplicial regions are d-simplices in $C_H(\mathcal{X}_{1-j})$ (intervals in $\mathbb{R}$, triangles in $\mathbb{R}^2$ and tetrahedrons in $\mathbb{R}^3$) and d-polytopes for $\mathbb{R}^d \backslash C_H(\mathcal{X}_{1-j})$. After partitioning $\mathbb{R}^d$ into disjoint regions, we further partition each simplex $\mathcal{S}_k$ (only the ones inside $C_H(\mathcal{X}_{1-j})$) into vertex regions, and define the simplicial proximity regions $N(x)$ for $x \in \mathcal{S}_k$. Here, we define the regions $N(x)$ as open sets in $\mathbb{R}^d$.

3.1 Class cover catch digraphs

Class Cover Catch Digraphs (CCCDs) are graph theoretic representations of the CCP (Priebe et al. 2001, 2003a). In a CCCD, for $x, y \in \mathcal{X}_j$; let $B = B(x, \varepsilon)$ be the ball centered at x with radius $\varepsilon = \varepsilon(x)$. A CCCD is a digraph $D_j = (\mathcal{V}_j, \mathcal{A}_j)$ with vertex set $\mathcal{V}_j = \mathcal{X}_j$ and the arc set $\mathcal{A}_j$ where $(x, y) \in \mathcal{A}_j$ iff $y \in B$. One particular family of CCCDs are called pure-CCCDs wherein, for all $x \in \mathcal{X}_j$, no non-target class point lies in B. Hence, for some $\theta \in (0, 1]$ and for all $x \in \mathcal{X}_j$, the open ball B is denoted by $B_\theta(x, \varepsilon_\theta(x))$ with the radius $\varepsilon_\theta(x)$ given by

$$\varepsilon_\theta(x) := (1 - \theta)d(x, \ell(x)) + \theta d(x, u(x)), \tag{7}$$

where $u(x) := \operatorname{argmin}_{y \in \mathcal{X}_{1-j}} d(x, y)$, and $\ell(x) := \operatorname{argmax}_{z \in \mathcal{X}_j} \{d(x, z) : d(x, z) < d(x, u(x))\}$. Here, $d(., .)$ can be any dissimilarity measure but we use the Euclidean distance henceforth. For all $x \in \mathcal{X}_j$, the definition of the radius $\varepsilon_\theta(x)$ keeps any non-target class point $v \in \mathcal{X}_{1-j}$ out of the ball B; that is, $\mathcal{X}_{1-j} \cap B = \emptyset$. We say the CCCD, D_j, is "pure" since the balls include only the target class points and none of the non-target class points. The CCCD, D_j,

is invariant to the choice of θ, but this parameter affects the classification performance. This parameter potentially establishes classifiers with increased performance (Priebe et al. 2003a). An illustration of the effect of parameter θ on the radius of $B_\theta(x, \varepsilon_\theta(x))$ is given in Fig. 3 (DeVinney 2003). In fact, CCCDs can be viewed as a family of PCDs using *spherical* proximity maps, letting $N(x) := B(x, \varepsilon(x))$. We denote the proximity regions associated with pure-CCCDs as $N_S(x, \theta) = B_\theta(x, \varepsilon_\theta(x))$. For simplicity, we refer to pure-CCCDs as CCCDs throughout this article.

3.2 Proportional-edge proximity maps

We use a type of proximity map with expansion parameter r, namely *proportional-edge* (PE) proximity map, denoted by $N_{PE}(\cdot, r)$. The PE proximity map and the associated digraphs, PE-PCDs, are defined in Ceyhan and Priebe (2005). Currently, PE-PCDs are only defined for the points in $\mathcal{X}_j \cap C_H(\mathcal{X}_{1-j})$. Hence, for the remaining target class points $\mathcal{X}_j$, i.e. $\mathcal{X}_j \setminus C_H(\mathcal{X}_{1-j})$, we extend the definition of PE proximity maps to the outer simplices. Hence, we will be able to show in the subsequent sections that the resulting PCDs have computationally tractable *minimum dominating sets* which are equivalent to the exact minimum prototype sets of PE-PCD classifiers for the entire data set.

3.2.1 PE proximity maps for the interior of convex hull of non-target points

For $r \in [1, \infty)$, we define $N_{PE}(\cdot, r)$ to be the PE proximity map associated with a triangle $T = T(\mathcal{Y})$ formed by the set of non-collinear points $\mathcal{Y} = \{y_1, y_2, y_3\} \subset \mathbb{R}^2$. Let $R_{M_C}(y_1)$, $R_{M_C}(y_2)$ and $R_{M_C}(y_3)$ be the vertex regions associated with vertices y_1, y_2 and y_3. Note that the barycentric coordinates of M_C are (1/3:1/3:1/3). For $x \in T^o$, let $v(x) \in \mathcal{Y}$ be the vertex whose region contains x; hence $x \in R_{M_C}(v(x))$. If x falls on the boundary of two vertex regions, or on M_C, we assign $v(x)$ arbitrarily. Let $e(x)$ be the edge of T opposite to $v(x)$. Let $\ell(v(x), x)$ be the line parallel to $e(x)$ through x. Let $d(v(x), \ell(v(x), x))$ be the Euclidean (perpendicular) distance from $v(x)$ to $\ell(v(x), x)$. For $r \in [1, \infty)$, let $\ell_r(v(x), x)$ be the line parallel to $e(x)$ such that $d(v(x), \ell_r(v(x), x)) = rd(v(x), \ell(v(x), x))$ and $d(\ell(v(x), x), \ell_r(v(x), x)) < d(v(x), \ell_r(v(x), x))$. Let $T_r(x)$ be the triangle similar to and with the same orientation as T where $T_r(x)$ has $v(x)$ as a vertex and and the edge opposite $v(x)$ lies on $\ell_r(v(x), x)$. Then the *proportional-edge* proximity region $N_{PE}(x, r)$ is defined to be

Fig. 3 The radius $\varepsilon_\theta(x)$ of a single target class point x in a two-class setting. Grey and black points represent the target class points $\mathcal{X}_j$ and the non-target class points $\mathcal{X}_{1-j}$, respectively. The solid circle is constructed with the radius $\varepsilon_\theta(x)$ given by $\theta = 1$, dashed one by $\theta = 0.5$ and the dotted one by $\theta = \epsilon$, where ϵ is the machine epsilon
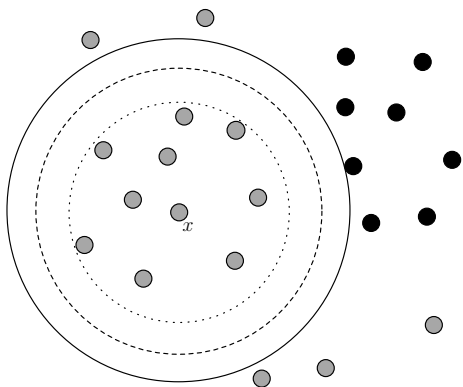
$(T_r(x) \cap T)^o$. Figure 4a illustrates a PE proximity region $N_{PE}(x,r)$ of a point x in an acute triangle.

The extension of $N_{PE}(\cdot,r)$ to $\mathbb{R}^d$ for $d > 2$ is straightforward. Now, let $\mathcal{Y} = \{y_1, y_2, \ldots, y_{d+1}\}$ be a set of $d+1$ non-coplanar points, and represent the simplex formed by the these points as $\mathcal{S} = \mathcal{S}(\mathcal{Y})$. We define the PE proximity map as follows. Given a point $x \in \mathcal{S}^o$, let $v(x)$ be the vertex in whose region x falls (if x falls on the boundary of two vertex regions or on M_C, we assign $v(x)$ arbitrarily.) Let $\varphi(x)$ be the face opposite to vertex $v(x)$, and $\eta(v(x),x)$ be the hyperplane parallel to $\varphi(x)$ which contains x. Let $d(v(x), \eta(v(x),x))$ be the (perpendicular) Euclidean distance from $v(x)$ to $\eta(v(x),x)$. For $r \in [1,\infty)$, let $\eta_r(v(x),x)$ be the hyperplane parallel to $\varphi(x)$ such that $d(v(x), \eta_r(v(x),x)) = r\, d(v(x), \eta(v(x),x))$ and $d(\eta(v(x),x), \eta_r(v(x),x)) < d(v(x), \eta_r(v(x),x))$. Let $\mathcal{S}_r(x)$ be the polytope similar to and with the same orientation as $\mathcal{S}$ having $v(x)$ as a vertex and $\eta_r(v(x),x)$ as the opposite face. Then the proportional-edge proximity region is given by $N_{PE}(x,r) := (\mathcal{S}_r(x) \cap \mathcal{S})^o$.

Notice that, so far, we assumed a single d-simplex for simplicity. For $n_{1-j} = d+1$, the convex hull of the non-target class $C_H(\mathcal{X}_{1-j})$ is a d-simplex. If $n_{1-j} > d+1$, then we consider the Delaunay tessellation (assumed to exist) of $\mathcal{X}_{1-j}$ where $\mathfrak{S}_{1-j}^{in} = \{\mathcal{S}_1, \ldots, \mathcal{S}_K\}$ denotes the set of all Delaunay cells (which are d-simplices). We construct the proximity region $N_{PE}(x,r)$ of a point $x \in \mathcal{X}_j$ depending on which d-simplex $\mathcal{S}_k$ this point resides in. Observe that, this construction pertains to points in $\mathcal{X}_j \cap C_H(\mathcal{X}_{1-j})$ only.

3.2.2 PE proximity maps for the exterior of convex hull of non-target points

For target class points $\mathcal{X}_j$ outside of the convex hull of the non-target class points $\mathcal{X}_{1-j}$, i.e. $\mathcal{X}_j \backslash C_H(\mathcal{X}_{1-j})$, we define the PE proximity maps similar to the ones defined for d-simplices. Let $\mathscr{F} \subset \mathbb{R}^2$ be an outer triangle defined by the adjacent boundary points which are without loss of generality assumed to be $\{y_1, y_2\} \subset \mathbb{R}^2$ of $C_H(\mathcal{X}_{1-j})$ and by rays $\overrightarrow{C_M y_1}$ and $\overrightarrow{C_M y_2}$ for C_M being the centroid of the boundary points of $C_H(\mathcal{X}_{1-j})$. Also, let $e = \mathcal{F}$ be the edge (or facet) of $C_H(\mathcal{X}_{1-j})$ adjacent to vertices $\{y_1, y_2\}$. Note that there is no center in an outer triangle, and hence no need for vertex regions. For $r \in [1,\infty)$, we define $N_{PE}(\cdot,r)$ to be the PE proximity map of the outer triangle as follows. For $x \in \mathscr{F}^o$, let $\ell(x,e)$ be the line
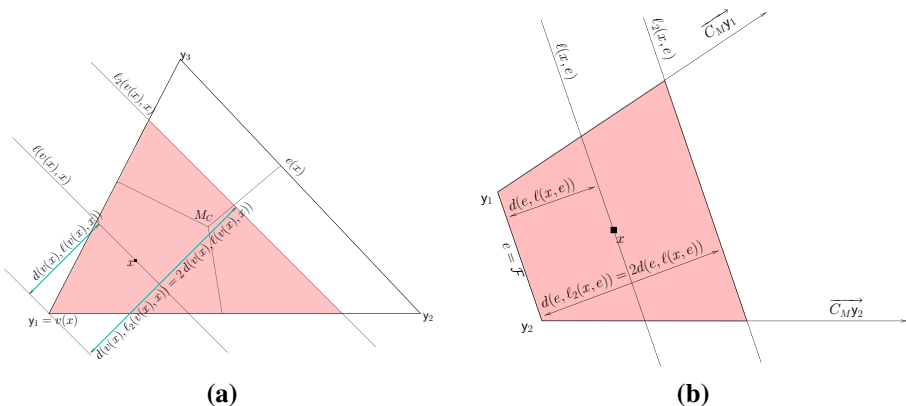

(a)   **(b)**

Fig. 4 The PE proximity region (shaded), $N_{PE}(x,r=2)$, **a** in a triangle $T \subseteq \mathbb{R}^2$ and **b** in an outer triangle $\mathscr{F} \subseteq \mathbb{R}^2$

parallel to e through x, and let $d(e, \ell(x, e))$ be the Euclidean distance from e to $\ell(x, e)$. For $r \in [1, \infty)$, let $\ell_r(x, e)$ be the line parallel to e such that $d(e, \ell_r(x, e)) = rd(e, \ell(x, e))$ and $d(\ell(x, e), \ell_r(x, e)) < d(e, \ell_r(x, e))$. Let $\mathscr{F}_r(x)$ be a polygon similar to the outer triangle $\mathscr{F}$ such that $\mathscr{F}_r(x)$ has e and $e_r(x) = \ell_r(x, e) \cap \mathscr{F}$ as its two edges, however $\mathscr{F}_r(x)$ is a bounded region whereas $\mathscr{F}$ is not. Then, the proximity region $N_{PE}(x, r)$ is defined to be $\mathscr{F}_r^o(x)$. Figure 4b illustrates a PE proximity region $N_{PE}(x, r)$ of a point x in an outer triangle.

The extension of $N_{PE}(\cdot, r)$ of outer triangles in $\mathbb{R}^2$ to $\mathbb{R}^d$ for $d > 2$ is also straightforward. Let $\mathscr{F} \subset \mathbb{R}^d$ be an outer simplex defined by the adjacent boundary points which are without loss of generality assumed to be $\{y_1, \ldots, y_d\} \subset \mathbb{R}^d$ of $C_H(\mathcal{X}_{1-j})$ and by rays $\{\overrightarrow{C_M y_1}, \ldots, \overrightarrow{C_M y_d}\}$. Also, let $\mathcal{F}$ be the facet of $C_H(\mathcal{X}_{1-j})$ adjacent to vertices $\{y_1, \ldots, y_d\}$. We define the PE proximity map as follows. Given a point $x \in \mathscr{F}^o$, let $\eta(x, \mathcal{F})$ be the hyperplane parallel to $\mathcal{F}$ through x and let $d(\mathcal{F}, \eta(x, \mathcal{F}))$ be the Euclidean distance from $\mathcal{F}$ to $\eta(x, \mathcal{F})$. For $r \in [1, \infty)$, let $\eta_r(x, \mathcal{F})$ be the hyperplane parallel to $\mathcal{F}$ such that $d(\mathcal{F}, \eta_r(x, \mathcal{F})) = rd(\mathcal{F}, \eta(x, \mathcal{F}))$ and $d(\eta(x, \mathcal{F}), \eta_r(x, \mathcal{F})) < d(\mathcal{F}, \eta_r(x, \mathcal{F}))$. Let $\mathscr{F}_r(x)$ be the polytope similar to the outer simplex $\mathscr{F}$ such that $\mathscr{F}_r(x)$ has $\mathcal{F}$ and $\mathcal{F}_r(x) = \eta_r(x) \cap \mathscr{F}$ as its two faces. Then, the proximity region $N_{PE}(x, r)$ is defined to be $\mathscr{F}_r^o(x)$.

The convex hull $C_H(\mathcal{X}_{1-j})$ has at least $d + 1$ facets (exactly $d + 1$ when $n_{1-j} = d + 1$), and since each outer simplex is associated with a facet, the number of outer simplices is at least $d + 1$. Let $\mathfrak{S}_{1-j}^{out} = \{\mathscr{F}_1, \ldots, \mathscr{F}_L\}$ denote the set of all outer simplices. This construction handles the points in $\mathcal{X}_j \backslash C_H(\mathcal{X}_{1-j})$ only. Together with the points inside $C_H(\mathcal{X}_{1-j})$, the PE-PCD, D_j, whose vertex set is $\mathcal{V}_j = \mathcal{X}_j$, has at least

$$\sum_{k=1}^{K} I(\mathcal{X}_j \cap \mathcal{S}_k \neq \emptyset) + \sum_{l=1}^{L} I(\mathcal{X}_j \cap \mathscr{F}_l \neq \emptyset)$$

many components where $I(\cdot)$ stands for the indicator function.

3.3 Minimum dominating sets of PCDs

We develop prototype-based classifiers with computationally tractable exact minimum prototype sets. We model the target class with a digraph D such that prototype sets of the target class are equivalent to dominating sets of D. Ceyhan (2010) determined the appealing properties of minimum dominating set of CCCDs in $\mathbb{R}$ as a guideline in defining new parametric digraphs relative to the Delaunay tessellation of points from the non-target class. In $\mathbb{R}$, finding the minimum dominating sets of CCCDs is computationally tractable, and the exact distribution of domination number is known for target class points which are uniformly distributed within each cell (Priebe et al. 2001). However, there is no polynomial time algorithm for finding the exact minimum dominating sets of CCCDs in $\mathbb{R}^d$ for $d > 1$. In this section, we provide a characterization of minimum dominating sets of PE-PCDs with the barycentric coordinate system and employ those coordinates to introduce algorithms for finding their minimum dominating sets in polynomial time.

We model the support of the class conditional distributions, i.e. $s(F_j)$, by a mixture of proximity regions. For a general proximity region $N(\cdot)$, we estimate the support of the class j as $Q_j := \cup_{x \in \mathcal{X}_j} N(x)$ such that $\mathcal{X}_j \subset Q_j$. Nevertheless, the support of the target class j can be estimated by a cover with lower complexity (i.e. with fewer proximity regions). For this purpose, we wish to reduce the model complexity by selecting an appropriate subset of proximity regions that still gives approximately the same estimate as Q_j. Let this (approximate) cover be defined as $C_j := \cup_{x \in S_j} N_{PE}(x, r)$, where S_j is a prototype set of points for $\mathcal{X}_j$

such that $\mathcal{X}_j \subset C_j$. A reasonable choice of the prototype set for our class covers is the minimum dominating set of PE-PCDs, whose elements are often more "central" than the arbitrary sets of the same size. Dominating sets of minimum size are desirable, since the size of the prototype sets determine the complexity of the model; that is, the smaller the set in cardinality (i.e. the model is lower in complexity), the higher the expected classification performance (Mehta et al. 1995; Rissanen 1989; Gao et al. 2013).

Algorithm 1 The greedy algorithm for finding an approximate minimum dominating set of a digraph D. Here, $D[H]$ is the digraph induced by the set of vertices $H \subseteq \mathcal{V}$ (see (West 2000)).

Input: A digraph $D = (\mathcal{V}, \mathcal{A})$
Output: An approximate minimum dominating set, S
    set $H = \mathcal{V}$ and $S = \emptyset$
    while $H \neq \emptyset$ **do**
        $v^* \leftarrow \mathrm{argmax}_{v \in \mathcal{V}(D)} |\{u \in \mathcal{V}(D) : (v, u) \in \mathcal{A}(D)\}|$
        $S \leftarrow S \cup \{v^*\}$
        $H \leftarrow \mathcal{V}(D) \setminus (\{u \in \mathcal{V}(D) : (v^*, u) \in \mathcal{A}(D)\} \cup \{v^*\})$
        $D \leftarrow D[H]$
    end while

In general, a digraph $D = (\mathcal{V}, \mathcal{A})$ of order $n = |\mathcal{V}|$, a vertex v *dominates* itself and all vertices of the form $\{u : (v, u) \in \mathcal{A}\}$. A *dominating set*, S_D, for the digraph D is a subset of $\mathcal{V}$ such that each vertex $v \in \mathcal{V}$ is dominated by a vertex in S_D. A *minimum dominating set* (MDS), S_{MD}, is a dominating set of minimum cardinality, and the *domination number*, $\gamma(D)$, is defined as $\gamma(D) := |S_{MD}|$. Finding a minimum dominating set is, in general, an NP-hard optimization problem (Karr 1992; Arora and Lund 1996). However, an approximately minimum dominating set can be obtained in $O(n^2)$ time using a well-known greedy algorithm as in Algorithm 1 (Chvatal 1979; Parekh 1991). PCDs using $N_S(\cdot, \theta)$ (or CCCDs with parameter θ) are examples of such digraphs. But, (exact) MDS of PE-PCDs are computationally tractable unlike PCDs based on $N_S(\cdot, \theta)$. Many attributes of these PE proximity maps and the proof of the existence of an algorithm to find a MDS are conveniently implemented through the barycentric coordinate system. Before proving the results on the MDS, we give the following proposition.

Proposition 2 *Let $\mathcal{Y} = \{y_1, y_2, \ldots, y_{d+1}\} \subset \mathbb{R}^d$ be a set of non-coplanar points for $d > 0$. For $x, x^* \in \mathcal{S} = \mathcal{S}(\mathcal{Y})$, we have $d(x, f_i) < d(x^*, f_i)$ if and only if $w_S^{(i)}(x) < w_S^{(i)}(x^*)$ for all $i = 1, \ldots, d + 1$, where $d(x, f_i)$ is the distance between point x and the face f_i.*

Proof For $i = 1, \ldots, d + 1$, note that f_i is the face of the simplex $\mathcal{S}$ opposite to the vertex y_i. Let $L(y_i, x)$ be the line through points x and y_i, and let $z \in f_i$ be the point that $L(y_i, x)$ crosses f_i. Also, recall that $\eta(y_i, x)$ denotes the hyperplane through the point x and parallel to f_i. Hence, for $\alpha \in (0, 1)$,

$$x = \alpha y_i + (1 - \alpha)z,$$

and since z is a convex combination of the set $\{y_k\}_{k \neq i}$,

$$x = \alpha \mathbf{y}_i + \left(\sum_{k=1; k \neq i}^{d+1} (1 - \alpha) \beta_k \mathbf{y}_k \right),$$

for $\beta_k \in (0, 1)$ for all k. Thus, $w_{\mathcal{S}}^{(i)}(x) = \alpha$ by the uniqueness of $\mathbf{w}_{\mathcal{S}}(x)$ for x with respect to $\mathcal{S}$. Observe that $\alpha = d(x, z)/d(\mathbf{y}_i, z) = d(x, f_i)/d(\mathbf{y}_i, f_i)$, since distances $d(x, z)$ and $d(x, f_i) = d(\eta(\mathbf{y}_i, x), f_i)$ are directly proportional (and so are $d(\mathbf{y}_i, z)$ and $d(\mathbf{y}_i, f_i)$). In fact, points that are on the same plane parallel to f_i have the same i^{th} barycentric coordinate as $w_{\mathcal{S}}^{(i)}(x) = \alpha$ corresponding to the vertex $\mathbf{y}_i$. Also, recall that, with decreasing α, the point x gets closer to f_i ($x \in f_i$ if $\alpha = 0$, and $x = \mathbf{y}_i$ if $\alpha = 1$). Then, for any two points $x, x^* \in \mathcal{S}$, we have $w_{\mathcal{S}}^{(i)}(x) = d(x, f_i)/d(\mathbf{y}_i, f_i)$ and $w_{\mathcal{S}}^{(i)}(x^*) = d(x^*, f_i)/d(\mathbf{y}_i, f_i)$. Thus $w_{\mathcal{S}}^{(i)}(x) < w_{\mathcal{S}}^{(i)}(x^*)$ if and only if $d(x, f_i) < d(x^*, f_i)$. □

Barycentric coordinates of a set of points in $\mathcal{S}(\mathcal{X}_{1-j})$ are useful in characterizing the set of *local extremum* points, which are extreme (having maximum or minimum distance) with respect to a subset of the class supports. A subset of local extremum points would constitute the minimum dominating set S_{MD}. We use Proposition 2 to prove the following theorem on MDS of a PE-PCD, D.

Theorem 2 *Let* $\mathcal{Z}_n = \{z_1, z_2, \ldots, z_n\} \subset \mathbb{R}^d$ *and* $\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{d+1}\} \subset \mathbb{R}^d$ *for* $d > 0$, *and let* $\mathcal{S} = \mathcal{S}(\mathcal{Y})$ *be the d-simplex given by the set* $\mathcal{Y}$ *such that* $\mathcal{Z}_n \subset \mathcal{S}$. *Let* D *be the PE-PCD associated with the proximity map* $N_{PE}(\cdot, r)$ *with vertex set* $\mathcal{V} = \mathcal{Z}_n$, *then we have* $\gamma(D) \leq d + 1$ *for all* $r > 1$.

Proof Let $x_{[i]} := \operatorname{argmin}_{x \in \mathcal{Z}_n \cap R_M(\mathbf{y}_i)} d(x, f_i)$, i.e., $x_{[i]}$ is the closest $\mathcal{Z}_n$ point in $R_M(\mathbf{y}_i)$ to face f_i (so $x_{[i]}$ is a local extremum in $\mathcal{Z}_n$ with respect to $R_M(\mathbf{y}_i)$), provided $\mathcal{Z}_n \cap R_M(\mathbf{y}_i) \neq \emptyset$. By Proposition 2, note that $d(x_{[i]}, f_i) \leq \min_{z \in \mathcal{Z}_n \cap R_M(\mathbf{y}_i)} d(z, f_i)$ if and only if $w_{\mathcal{S}}^{(i)}(x_{[i]}) \leq \min_{z \in \mathcal{Z}_n \cap R_M(\mathbf{y}_i)} w_{\mathcal{S}}^{(i)}(z)$. Hence, the *local extremum* point $x_{[i]}$ satisfies

$$x_{[i]} := \operatorname*{argmin}_{x \in \mathcal{Z}_n \cap R_M(\mathbf{y}_i)} w_{\mathcal{S}}^{(i)}(x).$$

Clearly, $\mathcal{Z}_n \cap R_M(\mathbf{y}_i) \subset N_{PE}(x_{[i]}, r)$ for all $r > 1$. Hence, $\mathcal{Z}_n \subset \cup_{i=1}^{d+1} N_{PE}(x_{[i]}, r)$. So, the set of all such local extremum points $E_L := \{x_{[1]}, \ldots, x_{[d+1]}\}$ (provided they all exist) is a dominating set for the PE-PCD with vertices $\mathcal{Z}_n$. If some of the $x_{[i]}$ do not exist, the set of such local extremum points will be a proper subset of E_L. Hence, we obtain $\gamma(D) \leq d + 1$. □

For $r = 1$, $x \notin N_{PE}(x, r)$, so $N_{PE}(x_{[i]}, r)$ does not cover the points on its boundary, in particular on its face coincident with $\eta_r(x_{[i]}, f_i)$ which is the same as $\eta(x_{[i]}, f_i)$ for $r = 1$. But $\eta(x_{[i]}, f_i)$ has Lebesgue measure zero in $\mathbb{R}^d$.

With $M = M_C$, MDSs of PE-PCDs are found by locating the closest point $x_{[i]}$ to face f_i in the vertex region $R_{M_C}(\mathbf{y}_i)$ for all $i = 1, \ldots, d + 1$. By Theorem 2, in $R_{M_C}(\mathbf{y}_i)$, the point $x_{[i]}$ is the closest point among $\mathcal{X}_j \cap R_{M_C}(\mathbf{y}_i)$ to the face f_i. For a set of d-simplices given by the Delaunay tessellation of $\mathcal{X}_{1-j}$, Algorithm 2 identifies all such local extremum points of each d-simplex in order to find the (exact) minimum dominating set $S_j = S_{MD}$.

Let $D_j = (\mathcal{V}_j, \mathcal{A}_j)$ be the PE-PCD with vertex set $\mathcal{V} = \mathcal{X}_j$. In Algorithm 2, we partition $\mathcal{X}_j$ into such subsets that each subset falls into a single d-simplex in the Delaunay tessellation of the set $\mathcal{X}_{1-j}$. Let $\mathfrak{S}_{1-j}$ be the set of all d-simplices associated with $\mathcal{X}_{1-j}$. Moreover, for each $\mathcal{S} \in \mathfrak{S}_{1-j}$, we further partition the subset $\mathcal{X}_j \cap \mathcal{S}$ into subsets that each subset falls into a single vertex region of $\mathcal{S}$. In each vertex region $R_{M_C}(\mathbf{y}_i)$, we find the closest point $x_{[i]}$ to

face f_i provided $\mathcal{X}_j \cap R_{M_C}(\mathbf{y}_i) \neq \emptyset$. Let $S(D)$ denote the minimum dominating set and $\gamma(D)$ denote the domination number of a digraph D. Also, let $D_j[\mathcal{S}]$ be the digraph induced by points of $\mathcal{X}_j$ inside the d-simplex $\mathcal{S}$, i.e. $\mathcal{X}_j \cap \mathcal{S}$. Recall that, as a result of Theorem 2, $\gamma(D_j[\mathcal{S}]) \leq d+1$ since $\mathcal{X}_j \cap \mathcal{S} \subset \cup_{i=1}^{d+1} N_{PE}(x_{[i]}, r)$. To find $S(D_j[\mathcal{S}])$, we sort all subsets of the set of such local extremum points, from smallest cardinality to highest, and check if $\mathcal{X}_j \cap \mathcal{S}$ is in the union of proximity regions of these subsets of local extremum points. For example, $S(D_j[\mathcal{S}]) = \{x_{[l]}\}$ and $\gamma(D_j[\mathcal{S}]) = 1$ if $\mathcal{X}_j \cap \mathcal{S} \subset N_{PE}(x_{[l]}, r)$ for some $l \in \{1, 2, \ldots, d+1\}$; else $S(D_j[\mathcal{S}]) = \{x_{[l_1]}, x_{[l_2]}\}$ and $\gamma(D_j[\mathcal{S}]) = 2$ if $\mathcal{X}_j \cap \mathcal{S} \subset N_{PE}(x_{[l_1]}, r) \cup N_{PE}(x_{[l_2]}, r)$ for some $\{l_1, l_2\} \in \binom{\{1, 2, \ldots, d+1\}}{2}$; or else $S(D_j[\mathcal{S}]) = \{x_{[1]}, x_{[2]}, x_{[3]}\}$ and $\gamma(D_j[\mathcal{S}]) = 3$ if $\mathcal{X}_j \cap \mathcal{S} \subset \cup_{l=1,2,3} N_{PE}(x_{[l]}, r)$, and so on. The resulting minimum dominating set of D_j for $\mathcal{X}_j \cap C_H(\mathcal{X}_{1-j})$ is the union of these sets, i.e., $S_j = \cup_{\mathcal{S} \in \mathfrak{S}_{1-j}} S(D_j[\mathcal{S}])$ and $\gamma(D_j) = |S_j|$. Observe that $S(D_j[\mathcal{S}]) = \emptyset$ if $\mathcal{X}_j \cap \mathcal{S} = \emptyset$. This algorithm is guaranteed to terminate, as long as n_0 and n_1 are both finite.

Algorithm 2 The algorithm for finding the (exact) minimum dominating set S_j of a PE-PCD, D_j, induced by $\mathcal{X}_j \cap C_H(\mathcal{X}_{1-j})$.

Input: The target class points $\mathcal{X}_j$, the Delaunay tessellation based on non-target class points $\mathcal{X}_{1-j}$ with Delaunay cells $\mathfrak{S}_{1-j}$, and the PE proximity map $N_{PE}(\cdot, r)$
Output: The minimum dominating set, S_j
1: $S_j = \emptyset$
2: **for all** $\mathcal{S} \in \mathfrak{S}_{1-j}$ where $\mathcal{X}_j \cap \mathcal{S} \neq \emptyset$ **do**
3:    $\mathcal{X}_j^* \leftarrow \mathcal{X}_j \cap \mathcal{S}$ and let $\{y_1, \ldots, y_{d+1}\}$ be the vertices of $\mathcal{S}$
4:    **for** $i = 1, \ldots, d+1$ **do**
5:       Let $x_{[i]} \leftarrow \underset{x \in \mathcal{X}_j^* \cap R_{M_C}(y_i)}{\operatorname{argmin}} w_{\mathcal{S}}^{(i)}(x)$
6:    **end for**
7:    **for** $t = 1, \ldots, d+1$ **do**
8:       **if** there exists a set $\{l_1, \ldots, l_t\} \in \binom{\{1, \ldots, d+1\}}{t}$ s.t. $\mathcal{X}_j^* \subset \cup_{a=1}^{t} N_{PE}(x_{[l_a]}, r)$
      then
9:          $S_j \leftarrow S_j \cup \{x_{[l_1]}, \ldots, x_{[l_t]}\}$
10:          **break**
11:       **end if**
12:    **end for**
13: **end for**

The level of reduction in the training data depends also on the magnitude of the expansion parameter r. In fact, the larger the magnitude of r, the more likely the $S(D_j[\mathcal{S}])$ have smaller cardinality, i.e. the more the reduction in the data set. Thus, we have a stochastic ordering as follows:

Theorem 3 *Let $\mathcal{S}$ be a d-simplex in $\mathbb{R}^d$ for $d > 0$ with $d+1$ non-coplanar vertices and $\mathcal{Z}_n = \{X_1, X_2, \ldots, X_n\}$ be a random sample from a continuous distribution F whose support is $\mathcal{S}$. Also let PE-PCD be defined with vertices $\mathcal{Z}_n$ (i.e., $\mathcal{Z}_n$ is from the target class j) and expansion parameter $r \geq 1$. Denote the domination number of this PE-PCD as $\gamma_d(r) = \gamma(\mathcal{Z}_n, D_j, r)$. Then for $r_1 < r_2$, we have $\gamma_d(r_2) \leq^{ST} \gamma_d(r_1)$ where $\leq^{ST}$ stands for "stochastically smaller than".*

Proof Suppose $r_1 < r_2$. Then for any $x \in \mathcal{S}$, we have $N_{PE}(x, r_1) \subseteq N_{PE}(x, r_2)$. Let $A_S := \{x \in \mathcal{S} : N_{PE}(x, r_1) \subsetneq N_{PE}(x, r_2)\}$. Since $r_1 < r_2$, we have $\lambda(A_S) > 0$ where λ is the

Lebesgue measure. For any $t \in \mathbb{Z}_+$, $\gamma_d(r) \le t$ iff there exist $x_1, x_2, \ldots, x_t \in \mathcal{Z}_n$ such that $\mathcal{Z}_n \subset \cup_{i=1}^{t} N_{PE}(x_i, r)$. But since $r_1 < r_2$, $\cup_{i=1}^{t} N_{PE}(x_i, r_1) \subseteq \cup_{i=1}^{t} N_{PE}(x_i, r_2)$, hence $\gamma_d(r_1) \le t$ implies $\gamma_d(r_2) \le t$, hence $P(\gamma_d(r_1) \le t) \le P(\gamma_d(r_2) \le t)$ for all $t > 0$. Furthermore, strict inequality holds for at least one $t \in \mathbb{Z}_+$, e.g., $t = 1$ since it is more likely that $N_{PE}(X, r_2)$ is more likely to cover all $\mathcal{Z}_n$ compared to $N_{PE}(X, r_1)$ for any $X \in \mathcal{Z}_n$ as $N_{PE}(x, r_1) \subsetneq N_{PE}(x, r_2)$ for $x \in A_S$. Hence, the desired result follows. □

Algorithm 2 ignores the target class points outside the convex hull of the non-target class. This is not the case with Algorithm 1, since the map $N_S(\cdot, \theta)$ is defined over all points in $\mathcal{X}_j$ whereas the original PE proximity map $N_{PE}(\cdot, r)$ is not. Hence, with Algorithm 2, the prototype set S_j only yields a reduction in the set $\mathcal{X}_j \cap C_H(\mathcal{X}_{1-j})$. We tackle this issue with various approaches. One approach is to define covering methods with two proximity maps that are the PE proximity map and another oone which does not require the target class points to be inside the convex hull of the non-target class points, e.g. spherical proximity regions (i.e. proximity maps $N_S(\cdot, \theta)$).

Algorithm 3 uses both maps $N_{PE}(\cdot, r)$ and $N_S(\cdot, \theta)$ to generate a prototype set S_j for the target class points $\mathcal{X}_j$. There are two separate MDSs, S_j^{in} which is exactly MDS, and S_j^{out} which is approximately MDS. The two maps are associated with two distinct digraphs such that $\mathcal{X}_j \cap C_H(\mathcal{X}_{1-j})$ constitutes the vertex set of one digraph and $\mathcal{X}_j \backslash C_H(\mathcal{X}_{1-j})$ constitute the vertex set of the other, where the non-target class is always $\mathcal{X}_{1-j}$. Algorithm 2 finds a prototype set S_j^{in} for $\mathcal{X}_j \cap C_H(\mathcal{X}_{1-j})$, and then the prototype set S_j^{out} for $\mathcal{X}_j \backslash C_H(\mathcal{X}_{1-j})$ is merged with the overall prototype set, i.e. $S_j = S_j^{\text{in}} \cup S_j^{\text{out}}$ as in Algorithm 3. Note that the set S_j is an approximately minimum dominating set, since S_j^{out} is an approximately minimum dominating set.

Algorithm 3 The algorithm for finding the minimum dominating set S_j of PCD, D_j, defined by using both the proximity maps $N_{PE}(\cdot, r)$ and $N_S(\cdot, \theta)$.

Input: The target class points $\mathcal{X}_j$, the Delaunay tessellation based on non-target class points $\mathcal{X}_{1-j}$ with Delaunay cells $\mathfrak{S}_{1-j}$, and the proximity maps $N_{PE}(\cdot, r)$ and $N_S(\cdot, \theta)$
Output: The approximate minimum dominating set, S_j
1: $S_j^{\text{in}} = \emptyset$ and $S_j^{\text{out}} = \emptyset$
2: Find the minimum dominating set of PE-PCD with the vertex set $\mathcal{X}_j \cap C_H(X_{1-j})$ as in Algorithm 2 and assign it to S_j^{in}
3: $\mathcal{X}_j' = \mathcal{X}_j \setminus C_H(\mathcal{X}_{1-j})$
4: Find the approximate minimum dominating set of the CCCD as in Algorithm 1 where the target class is $\mathcal{X}_j'$ and the non-target class is $\mathcal{X}_{1-j}$, and assign the MDS to S_j^{out}
5: $S_j = S_j^{\text{in}} \cup S_j^{\text{out}}$

Algorithm 4 uses only the PE proximity map $N_{PE}(\cdot, r)$ with the original version inside $C_H(\mathcal{X}_{1-j})$ and extended version outside $C_H(\mathcal{X}_{1-j})$. The cover is a mixture of d-simplices and d-polytopes. Given a set of d-simplices, $\mathfrak{S}_{1-j}^{\text{in}}$, and a set of outer simplices $\mathfrak{S}_{1-j}^{\text{out}}$, we find the respective local extremum points of each d-simplex and outer simplex where local extremum in a d-simplex is the closest point among the data points in the vertex region to the face opposite to the relevant vertex and the local extremum in an outer simplex is the furthest data point to the face which constitutes the bottom edge of the outer simplex. Local extremum points of d-simplices are found as in Algorithm 2, and then we find the local extremum points of the remaining points to get the prototype set for the entire target class points $\mathcal{X}_j$. The following theorem provides a result on the local extremum points in

an outer simplex $\mathscr{F}$. Note that, in Algorithm 4, the set S_j is the exact minimum dominating set, since both S_j^{in} and S_j^{out} are exact MDSs for the PE-PCDs induced by $\mathcal{X}_j \cap C_H(\mathcal{X}_{1-j})$ and $\mathcal{X}_j \backslash C_H(\mathcal{X}_{1-j})$, respectively.

Theorem 4 *Let $\mathcal{Z}_n = \{z_1, z_2, \ldots, z_n\} \subset \mathbb{R}^d$ and $\mathcal{F}$ be a facet of the $C_H(\mathcal{X}_{1-j})$ and $\mathscr{F}$ be the associated outer simplex such that $\mathcal{Z}_n \subset \mathscr{F}$. Then, the furthest point among $\mathcal{Z}_n$ points to the facet $\mathcal{F}$ is a minimum dominating set S_{MD} of the PE-PCD restricted to $\mathscr{F}$ and is found in linear time. Moreover, the domination number of this restricted PE-PCD equals to 1 (provided $n > 0$).*

Proof We show that there is a point $s \in \mathcal{Z}_n$ such that $\mathcal{Z}_n \subset N_{PE}(s, r)$ for all $r \in (1, \infty)$. Note that $\eta(x, \mathcal{F})$ denotes the hyperplane through x and parallel to $\mathcal{F}$. Thus, for $x, x^* \in \mathcal{F}$, observe that $d(x, \mathcal{F}) < d(x^*, \mathcal{F})$ if and only if $d(\eta(x, \mathcal{F}), \mathcal{F}) < d(\eta(x^*, \mathcal{F}), \mathcal{F})$. Then it follows that $N_{PE}(x, r) \subsetneq N_{PE}(x^*, r)$ which implies $\{x, x^*\} \subset N_{PE}(x^*, r)$. Then, for $s := \operatorname{argmax}_{x \in \mathcal{Z}_n} d(x, \mathcal{F})$, we have $\mathcal{Z}_n \subset N_{PE}(s, r)$. So, $S_{MD} = \{s\}$ and $\gamma = 1$. Also, since s is the furthest point among $\mathcal{Z}_n$ from the facet $\mathcal{F}$, finding the MDS is linear in n. □

For $r = 1$, some $\mathcal{Z}_n$ points may fall on $\eta(s, r)$ (i.e., on the boundary of $N_{PE}(s, r)$) so $\{s\}$ is not a dominating set in such a case, but $\eta(s, r)$ is of Lebesgue measure zero in $\mathbb{R}^d$.

Algorithm 4 The algorithm for finding the (exact) minimum dominating set S_j of PE-PCD, D_j, with vertex set $\mathcal{X}_j$.

Input: The target class points $\mathcal{X}_j$, the set of d-simplices $\mathfrak{S}_{1-j}^{in}$, the set of polytopes $\mathfrak{S}_{1-j}^{out}$
Output: The minimum dominating set, S_j
1: $S_j^{in} = \emptyset$ and $S_j^{out} = \emptyset$
2: Find the minimum dominating set of PE-PCD restricted to the interior points as in Algorithm 2 and assign it to S_j^{in}
3: $\mathcal{X}_j' = \mathcal{X}_j \setminus C_H(\mathcal{X}_{1-j})$
4: **for all** $\mathscr{F} \in \mathfrak{S}_{1-j}^{out}$ where $\mathcal{X}_j' \cap \mathscr{F} \neq \emptyset$ **do**
5: $\quad \mathcal{X}_j^* \leftarrow \mathcal{X}_j' \cap \mathscr{F}$
6: $\quad$ Let $s \in \mathcal{X}_j^*$ be the local extremum point (i.e. furthest point from the corresponding facet) in $\mathscr{F}$
7: $\quad S_j^{out} \leftarrow S_j^{out} \cup \{s\}$
8: **end for**
9: $S_j = S_j^{in} \cup S_j^{out}$

Given Theorems 2 and 4, Algorithm 4 may be the most appealing one, since it gives the exact minimum dominating set for the complete target class j. However, the following theorem shows that the cardinality of such sets increase exponentially with dimensionality of the data set, even though it is polynomial on the number of observations.

Theorem 5 *Algorithm 4 finds an exact minimum dominating set S_j of the target class points $\mathcal{X}_j$ in $\mathcal{O}\left(d^k n_{1-j}^2 + 2^d n_{1-j}^{\lceil d/2 \rceil}\right)$ time for $k > 1$ where $|S_j| = \mathcal{O}\left(d n_{1-j}^{\lceil d/2 \rceil}\right)$.*

Proof A Delaunay tessellation of the non-target class points $\mathcal{X}_{1-j} \subset \mathbb{R}^d$ is found in $\mathcal{O}(d^k n_{1-j}^2)$ time with the Bowyer-Watson algorithm for some $k > 1$, depending on the complexity of the algorithm that finds the circumcenter of a d-simplex (Watson 1981). The resulting tessellation with n_{1-j} vertices has at most $\mathcal{O}\left(n_{1-j}^{\lceil d/2 \rceil}\right)$ simplices and at most

$\mathcal{O}(n_{1-j}^{\lfloor d/2 \rfloor})$ facets (Seidel 1995). Hence, the union of sets of d-simplices $\mathfrak{S}_{1-j}^{\text{in}}$ and outer simplices $\mathfrak{S}_{1-j}^{\text{out}}$ is of cardinality at most $\mathcal{O}\left(n_{1-j}^{\lfloor d/2 \rfloor}\right)$. Now, for each simplex $\mathcal{S} \in \mathfrak{S}_{1-j}^{\text{in}}$ or each outer simplex $\mathcal{F} \in \mathfrak{S}_{1-j}^{\text{out}}$, the local extremum points are found in linear time. Each simplex is divided into $d + 1$ vertex regions with each having its own extremum point. Hence, a minimum cardinality subset of the set of local extremum points is of cardinality at most $d + 1$ and found in a brute force fashion. For outer simplices, however, the local extremum point is the furthest point to the associated facet of the Delaunay tessellation. Thus, it takes at most $\mathcal{O}(2^d)$ and $\mathcal{O}(n)$ time to find the exact minimum dominating sets which are subsets of local extremum points for each (inner) simplex and outer simplex, respectively. Hence, the desired result follows. □

Theorem 5 shows the exponential increase of the number of prototypes as dimensionality increases. So, the complexity of the class cover model also increases exponentially, which might lead to overfitting. We will investigate this issue further in Sects. 6 and 7.

4 PCD covers

We establish class covers with the PE proximity map $N_{PE}(\cdot, r)$ and spherical proximity map $N_S(\cdot, \theta)$. We define two types of class covers: one type is called *composite covers* which cover the points in $\mathcal{X}_j \cap C_H(\mathcal{X}_{1-j})$ with PE proximity maps and the points in $\mathcal{X}_j \backslash C_H(\mathcal{X}_{1-j})$ with spherical proximity maps, and the other is called *standard cover* incorporating the PE proximity maps for all points in $\mathcal{X}_j$. We use these two types of covers to establish a specific type of classifier that is more appealing in the sense of prototype selection.

Our composite covers are mixtures of simplicial and spherical proximity regions. Specifically, given a set of simplices and a set of spheres, the composite cover is the union of both these sets which constitute proximity regions of two separate PCD families, hence the name *composite cover*. Let $N_{\text{in}}(\cdot)$ and $N_{\text{out}}(\cdot)$ be the proximity maps associated with sets $\mathcal{X}_j \cap C_H(\mathcal{X}_{1-j})$ and $\mathcal{X}_j \backslash C_H(\mathcal{X}_{1-j})$, respectively. The set Q_j is partitioned into two: the cover Q_j^{in} of points inside the convex hull of non-target class points, i.e., $\mathcal{X}_j \cap C_H(\mathcal{X}_{1-j})$, and the cover Q_j^{out} of non-target class points outside, i.e., $\mathcal{X}_j \backslash C_H(\mathcal{X}_{1-j})$. Let $Q_j^{(1)} := \cup_{x \in \mathcal{X}_j \cap C_H(\mathcal{X}_{1-j})} N_{\text{in}}(x)$ and $Q_j^{\text{out}} := \cup_{x \in \mathcal{X}_j \backslash C_H(\mathcal{X}_{1-j})} N_{\text{out}}(x)$ such that $Q_j := Q_j^{\text{in}} \cup Q_j^{\text{out}}$. Hence, in composite covers, target class points inside $C_H(\mathcal{X}_{1-j})$ are covered with PE proximity map $N_{\text{in}}(\cdot) = N_{PE}(\cdot, r)$, and the remaining points are covered with spherical proximity map $N_{\text{out}}(\cdot) = N_S(\cdot, \theta)$. Given the covers Q_j^{in} and Q_j^{out}, let C_j^{in} and C_j^{out} be the class covers with lower complexity associated with the dominating sets S_j^{in} and S_j^{out}. Let $C_j^{\text{in}} := \cup_{s \in S_j^{\text{in}}} N_{\text{in}}(s)$ and $C_j^{\text{out}} := \cup_{s \in S_j^{\text{out}}} N_{\text{out}}(s)$. Then, the composite cover is given by

$$C_j := C_j^{\text{in}} \cup C_j^{\text{out}}.$$

An illustration of the class covers C_0 and C_1 with $N_{\text{in}}(\cdot) = N_{PE}(\cdot, r = 2)$ and $N_{\text{out}}(\cdot) = N_S(\cdot, \theta = 1)$ is given in Fig. 5b.

By definition, the spherical proximity map $N_S(\cdot, \theta)$ yields class covers for all points in $\mathcal{X}_j$. Figure 5a illustrates the class covers of the map $N_S(\cdot, \theta = 1)$. We call such covers which only constitute a single type of proximity map as *standard covers*. Hence, the standard cover of the PE-PCD, D_j, is a union of d-simplices and d-polytopes:
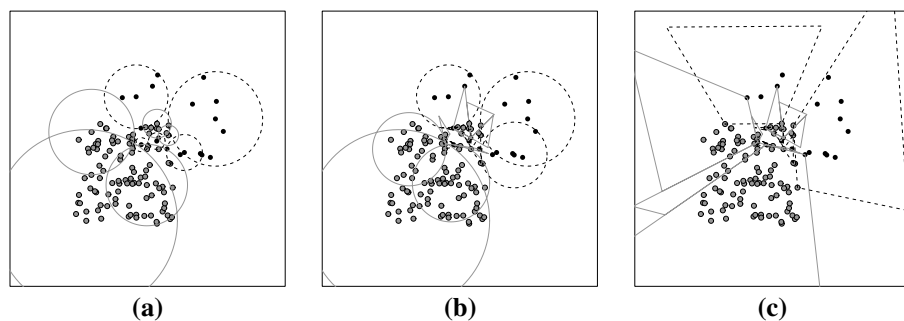
Fig. 5 Class covers of a data set in a two-class setting in $\mathbb{R}^2$ where grey and black points represent points of two distinct classes. The training data set is composed of two classes labeled as 0 and 1 wherein 100 and 20 data points are drawn from multivariate uniform distributions $U([0, 1]^2)$ and $U([0.5, 1.5]^2)$, respectively. Cover of one class is given by solid circle and solid line segments, and the cover of the other is given by dashed circles and dashed line segments. **a** Standard class covers with $N_{\text{in}}(\cdot) = N_{\text{out}}(\cdot) = N_S(\cdot, \theta = 1)$ **b** composite class cover with $N_{\text{in}}(\cdot) = N_{PE}(\cdot, r = 2)$ and $N_{\text{out}}(\cdot) = N_S(\cdot, \theta = 1)$ **c** standard class covers with $N_{\text{in}}(\cdot) = N_{\text{out}}(\cdot) = N_{PE}(\cdot, r = 2)$

$$C_j := \bigcup_{s \in S_j} N_{PE}(s, r).$$

Here, $N_{\text{in}}(\cdot) = N_{\text{out}}(\cdot) = N_{PE}(\cdot, r)$. An illustration is given in Fig. 5c.

PCD covers can easily be generalized to the multi-class case with J classes. To establish the set of covers $\mathcal{C} = \{C_1, C_2, \ldots, C_J\}$, the set of PCDs $\mathcal{D} = \{D_1, \ldots, D_J\}$, and the set of MDSs $\mathfrak{S} = \{S_1, S_2 \ldots, S_J\}$ associated with a set of classes with labels $\mathfrak{C} = \{1, 2, \ldots, J\}$, we gather the classes into two classes as $C_T = j$ and $C_{NT} = \cup_{t \neq j}\{t\}$ for $t, j = 1, \ldots, J$. We refer to classes C_T and C_{NT} as target and non-target classes, respectively. More specifically, target class is the class we want to find the cover of, and the non-target class is the union of the remaining classes. We transform the multi-class case into the two-class setting and find the cover of jth class, C_j for each $j = 1, 2, \ldots, J$.

5 Classification with PCDs

The elements of the minimum dominating set S_j are selected prototypes for the problem of modelling the class conditional discriminant regions via a collection of proximity regions (balls, simplices, polytopes, etc.). The sizes of these regions represent an estimate of the domain of influence, which is the region in which a given prototype should influence the class labelling. Our semi-parametric classifiers depend on the class covers given by these proximity regions. We define various classifiers based on the class covers (composite or standard) and some other classification methods. We approach classification of points in $\mathbb{R}^d$ in two ways:

Hybrid classifiers   Given the class covers C_0^{in} and C_1^{in} associated with classes with labels 0 and 1, we classify a given point $z \in \mathbb{R}^d$ with g_P if $z \in C_0^{\text{in}} \cup C_1^{\text{in}}$, and with g_A otherwise. Here, g_P is the pre-classifier and g_A is an alternative classifier.

Cover classifiers   These classifiers are constructed by class covers only; that is, a given point $z \in \mathbb{R}^d$ is classified as $g_C(z) = j$ if $z \in C_j \backslash C_{1-j}$ or if $\rho(z, C_j) < \rho(z, C_{1-j})$, hence class of the point z is estimated as j if z is only in cover C_j or closer to C_j than C_{1-j}. Here, $\rho(z, C_j)$ is a dissimilarity measure between point z and the cover C_j. Cover classifiers depend on the type of covers which are either *composite* or *standard*.

We incorporate PE-PCDs for establishing both of these types of classifiers. Hence, we will refer to them as hybrid PE-PCD and cover PE-PCD classifiers. Since the PE proximity maps were originally defined for points $\mathcal{X}_j \cap C_H(\mathcal{X}_{1-j})$, we develop hybrid PE-PCD classifiers to account for points outside of the convex hull of the non-target class in a convenient fashion. However, as we shall see later, cover PE-PCD classifiers have more appealing properties than hybrid PE-PCD classifiers in terms of both efficiency and classification performance. Nonetheless, we consider and compare both types of classifiers, but first we define the PE-PCD pre-classifier.

5.1 PE-PCD pre-classifier

Let $\rho(z, C)$ be a dissimilarity measure between z and the class cover C. The PE-PCD pre-classifier is given by

$$g_P(z) := \begin{cases} j & \text{if } z \in C_j^{\text{in}} \backslash C_{1-j}^{\text{in}} \text{ for } j = 0, 1 \\ I(\rho(z, C_1^{\text{in}}) < \rho(z, C_0^{\text{in}})) & \text{if } z \in C_0^{\text{in}} \cap C_1^{\text{in}} \\ -1 & \text{otherwise.} \end{cases} \tag{8}$$

Here $g_P(z) = -1$ denotes a "no decision" case. Given that class covers C_0^{in} and C_1^{in} are the unions of PE proximity regions $N_{PE}(x, r)$ of points in dominating sets S_0^{in} and S_1^{in}, the closest cover for a new point z is found by, first, finding the proximity region of a point in the cover closest to the point z:

$$\rho(z, C_j^{\text{in}}) = \min_{s \in S_j^{\text{in}}} \rho(z, N(s))$$

which is expressed based on a dissimilarity measure between the point z and the region $N_{PE}(s)$. For such measures, we employ *convex distance functions*. Let H be a convex set in $\mathbb{R}^d$ with $x \in H$ where the point x may be viewed as the center of the set H. Thus, the convex distance (or dissimilarity) between z and H be defined by

$$\rho(z, H) := \frac{d(z, x)}{d(t, x)},$$

where $d(\cdot, \cdot)$ is the Euclidean distance and t is a point of intersection for the half line $L(x, z) := \{x + \alpha(z - x) : \alpha \in [0, \infty)\}$ and $\partial(H)$. An illustration is given in Fig. 6 for several convex sets, including balls and simplices in $\mathbb{R}^2$.

For spherical proximity map $N_S(\cdot, \theta)$, the dissimilarity function is defined by putting the radius of that ball which is a spherical proximity region: $d(x, t) = \varepsilon_\theta(x)$ into the denominator (Priebe et al. 2003a). However, for d-simplices, we characterize the dissimilarity measure in terms of barycentric coordinates of z with respect to $\mathcal{S}(x) = N_{PE}(x, r)$.
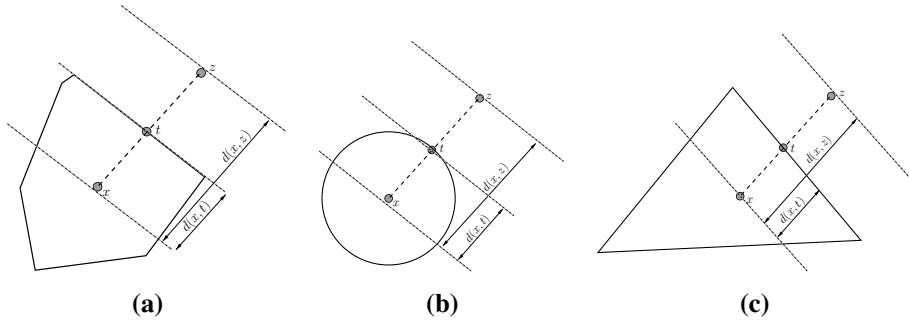
Fig. 6 Illustration of a convex distance between a point z and an arbitrary **a** convex set H, **b** ball and **c** 2-simplex in $\mathbb{R}^2$

Proposition 3 *Let $\{t_1, t_2, \ldots, t_{d+1}\} \subset \mathbb{R}^d$ be a set of non-coplanar points that are the vertices of simplex $\mathcal{S}(x) = N_{PE}(x, r)$ with the centroid $M_C(x) \in \mathcal{S}(x)^o$. Then, for $z \in \mathbb{R}^d$ and $t \in \partial(\mathcal{S}(x))$, the convex distance between z and $\mathcal{S}$ which is defined as $\rho(z, \mathcal{S}(x)) = d(M_C(x), z)/d(M_C(x), t)$ satisfies the following*

$$\rho(z, \mathcal{S}(x)) = 1 - (d+1)w_{\mathcal{S}(x)}^{(k)}(z),$$

where t is on the closest face f_k of $\mathcal{S}(x)$ to z and $w_{\mathcal{S}(x)}^{(k)}(z)$ is the kth barycentric coordinate of z with respect to $\mathcal{S}(x)$. Moreover, $\rho(z, \mathcal{S}(x)) \leq 1$ iff $z \in \mathcal{S}(x)$.

Proof Let the line segment $L(M_C(x), z)$ cross $\partial(\mathcal{S}(x))$ at the point $t \in f_k$ for f_k being the face of $\mathcal{S}(x)$ opposite to vertex t_k. Thus, for $\alpha_i \in (0, 1)$ and $\beta > 0$,

$$z = (1 - \beta)M_C(x) + \beta t = (1 - \beta)M_C(x) + \beta\left(\sum_{i=1; i \neq k}^{d+1} \alpha_i t_i\right).$$

Here, note that $\beta = d(M_C(x), z)/d(M_C(x), t) = \rho(z, \mathcal{S}(x))$ since z is a convex combination of t and $M_C(x)$. Also, since $M_C(x)$ is the centroid,

$$z = (1 - \beta)\frac{\sum_{i=1}^{d+1} t_i}{d+1} + \beta\left(\sum_{i=1; i \neq k}^{d+1} \alpha_i t_i\right) = \frac{1 - \beta}{d+1}t_k + \sum_{i=1; i \neq k}^{d+1}\left(\frac{1 - \beta}{d+1} + \beta\alpha_i\right)t_i.$$

Hence, $(1 - \beta)/(d+1) = w_{\mathcal{S}(x)}^{(k)}(z)$ which implies $\beta = 1 - (d+1)w_{\mathcal{S}(x)}^{(k)}(z)$.

For the second part, we first assume $\rho(z, \mathcal{S}(x)) \leq 1$. Then $d(M_C(x), z) \leq d(M_C(x), t)$ and $M_C(x)$ is in the interior of $\mathcal{S}(x)$ as $\mathcal{S}(x)$ is convex. Since t is on the face f_k of $\mathcal{S}(x)$ closest to z, z falls on the line segment joining $M_C(x)$ and t, denoted as $[M_C(x), t]$ which lies in $\mathcal{S}(x)$ as well. Hence, $z \in \mathcal{S}(x)$. For the reverse direction, assume that $z \in \mathcal{S}(x)$. Then there exists a face f_k of $\mathcal{S}(x)$ closest to z. Since f_k is the closest face to z and $M_C(x)$ is in the interior of $\mathcal{S}(x)$ as $\mathcal{S}(x)$ is convex, the line segment $[M_C(x), t]$ lies in $\mathcal{S}(x)$ as well, and z also lies on this line segment. Then it follows that $d(M_C(x), z) \leq d(M_C(x), t)$, so $\rho(z, \mathcal{S}(x)) = d(M_C(x), z)/d(M_C(x), t) \leq 1$. □

Observe that in Proposition 3, it also follows (from the proof) that $\rho(z, \mathcal{S}(x)) < 1$ iff $z \in \mathcal{S}(x)^o$ and $\rho(z, \mathcal{S}(x)) = 1$ iff $z \in \partial(\mathcal{S}(x))$.

For a (convex) proximity region $N_{PE}(x, r)$, the dissimilarity measure $\rho(z, \mathcal{S}(x)) = \rho(z, N_{PE}(x, r))$ indicates whether the point z is in the proximity region $N_{PE}(x, r)$ or not, since $\rho(z, \mathcal{S}(x)) < 1$ if $z \in N_{PE}(x, r)$ and ≥ 1 otherwise. Hence, the PE-PCD pre-classifier g_P may be simplified to

$$g_P(z) := \begin{cases} I(\rho(z, C_1^{in}) < \rho(z, C_0^{in})) & \text{if } z \in C_0^{in} \cup C_1^{in} \\ -1 & \text{otherwise.} \end{cases} \tag{9}$$

Here, without loss of generality, $z \in C_0^{in} \backslash C_1^{in}$ if and only if $\rho(z, C_0^{in}) < 1$. Let $\rho(z, x) := \rho(z, \mathcal{S}(x))$ be the dissimilarity between x and z, then the dissimilarity measure $\rho(\cdot, \cdot)$ violates the symmetry axiom of the metric, since $\rho(x, z) \neq \rho(z, x)$ unless $d(x, t(x)) = d(z, t(z))$ where proximity regions $N_{PE}(x, r)$ and $N_{PE}(z, r)$ intersect with the lines $L(M_C(x), z)$ and $L(M_C(z), x)$ at points $t(x)$ and $t(z)$, respectively.

5.2 Hybrid PE-PCD classifiers

Constructing hybrid classifiers has many purposes. Some classifiers are designed to solve harder classification problems by gathering many weak learning methods (often known as *ensemble* classifiers) while some others have advantages only when combined with another single classifier (Woźniak et al. 2014). Our hybrid classifiers are of the latter type. The PE-PCD pre-classifier, g_P, is able to classify points in the union of class covers, $C_0^{in} \cup C_1^{in}$, however classifying the remaining points in $\mathbb{R}^d$ requires incorporating an alternative classifier, often one that works for all points in $\mathbb{R}^d$. We use the PE-PCD pre-classifier, $g_P(\cdot)$, to classify all points of the test data, and if no decision are made for some of these points, we classify them with the alternative classifier g_A. Hence, let g_H be the hybrid PE-PCD classifier such that

$$g_H(z) := \begin{cases} g_P(z) & \text{if } z \in C_0^{in} \cup C_1^{in} \\ g_A(z) & \text{otherwise.} \end{cases} \tag{10}$$

That is, for "no decision" cases where $g_P(z) = -1$, we rely on the alternative classifier g_A; we will use the kNN, SVM and CCCD classifiers as alternative classifiers. The parameters are k, the number of closest neighbors to make a majority vote in the kNN classifier; γ, the scaling parameter of the radial basis function (RBF) kernel of the SVM classifier; and θ, the parameter of the CCCD classifier that regulates the size of each ball as described in Sect. 3.1.

5.3 Composite and standard cover PE-PCD classifiers

We propose PE-PCD classifiers g_C based on composite and standard covers. The classifier g_C is defined as

$$g_C(z) := I(\rho(z, C_1) < \rho(z, C_0)). \tag{11}$$

The cover is based on either composite covers or standard covers wherein $\mathcal{X}_j \subset C_j$ for $j = 1, 2$, hence a decision can be made without an alternative classifier. Note that composite cover PE-PCD classifiers are, in fact, different types of hybrid classifiers where the classifiers are only modelled by class covers but with multiple types of PCDs. Compared to hybrid PE-PCD classifiers, cover PE-PCD classifiers have many appealing properties.

Since a reduction is done over all target class points $\mathcal{X}_j$, depending on the percentage of reduction, classifying a new point $z \in \mathbb{R}^d$ is computationally faster and more efficient, whereas an alternative classifier might not provide such a reduction.

Note that, given the multi-class prototype sets, S_j, the two-class cover PE-PCD classifier, g_C, can be modified for the multi-class case as

$$g(z) = \operatorname*{argmin}_{j \in J} \left(\min_{s \in S_j} \rho(z, N(s)) \right) \tag{12}$$

for a general proximity map $N(\cdot)$.

5.4 Consistency analysis

In this section, we will prove consistency of cover and hybrid PCD classifiers when the two class conditional distributions are strictly δ-separable. For $\delta \in [0, \infty)$, the regions $A, B \subset \mathbb{R}^d$ are called *δ-separable* if

$$\inf_{x \in A, y \in B} d(x, y) \geq \delta$$

and *strictly δ- separable* if, moreover, $\delta > 0$. Notice that the definition of δ-separability allows overlap in the sets A and B with $\delta = 0$. Furthermore, if the continuous distributions F and G have δ-separable supports, then they are also called *δ-separable distributions*, and if $\delta > 0$, they are called *strictly δ- separable distributions* (Devroye et al. 1996).

Recall that cover classifiers are characterized by PCDs associated with proximity regions $N(x)$ for $x \in \mathbb{R}^d$, and thus, the consistency of such PCD classifiers depends on the proximity map $N(\cdot)$. We require that the proximity map $N(\cdot)$ satisfies the following properties:

P1   For all $x \in \mathbb{R}^d$, the proximity region $N(x)$ is either an open set or $N(x) = \{x\}$ and x is in the interior of $N(x)$ almost everywhere (a.e.) in Lebesgue measure.

P2   For two classes, the proximity map $N(x)$ is a function of x from target class and also depends on the non-target class points y in such a way that $N(x) \cap y = \emptyset$ a.e. in Lebesgue measure.

Notice that **P1** implies that $N(x)$ is an open set a.e. in $\mathbb{R}^d$-Lebesgue measure and **P2** implies that, the set $\{(x, y) : N(x) \cap y \neq \emptyset\}$ has zero $\mathbb{R}^{2d}$-Lebesgue measure. Both $N_S(\cdot, \theta)$ for $\theta \in (0, 1]$ and $N_{PE}(\cdot, r)$ for $r \in (1, \infty)$ satisfy the properties **P1** and **P2**. These will be useful in showing that the classifiers based on our class covers attain Bayes-optimal classification performance for classes with (strictly) δ-separable continuous distributions.

In the rest of this section, we assume that we have a random sample $\mathcal{X}_j$ of size n_j from class j with continuous distribution F_j whose support is $s(F_j)$ for $j = 0, 1$. Recall that the PCD class cover for class j based on $N(\cdot)$ is $C_j = \cup_{x \in S_j} N(x)$ with S_j being a prototype set of points for $\mathcal{X}_j$ (so $S_j \subseteq \mathcal{X}_j$). Note that all target class (say, class j) points reside inside the class cover C_j w.p. 1 by **P1**, i.e. $\mathcal{X}_j \subset C_j$ w.p. 1 for all $n_j > 0$. Hence, we have the following lemma.

Lemma 1 *Let $\mathcal{X}_j = \{X_1, X_2, \dots, X_{n_j}\}$ be a random sample of size n_j from a class j with a continuous distribution F_j whose support is $s(F_j) \subseteq \mathbb{R}^d$. Also, let the class cover for class j*

based on proximity map $N(\cdot)$ be denoted as $C_j := C(\mathcal{X}_j)$ such that $C_j = \cup_{X \in S_j} N(X)$ with prototype set $S_j \subseteq \mathcal{X}_j$. If $N(\cdot)$ satisfies property **P1**, then we have $\lambda(s(F_j) \backslash C_j) \to 0$ a.s. as $n_j \to \infty$.

Proof Assume that $N(\cdot)$ satisfies property **P1**. Then, $N(X)$ is open w.p. 1, and $P(X \in N(X)) = 1$ for $X \sim F_j$, which implies $P(\mathcal{X}_j \subset C_j) = 1$ for all $n_j > 0$. Assume, for a contradiction, that $\lambda(s(F_j) \backslash C_j) \to \varepsilon$ a.s. for some $\varepsilon > 0$ as $n_j \to \infty$. Then, as $n_j \to \infty$, there exists a region $s_\varepsilon(F_j)$ in $s(F_j)$ with positive Lebesgue measure so that $P(X \in s_\varepsilon(F_j)) > 0$ and such X's are not in C_j with positive probability. That is, $P(X_i \in s_\varepsilon(F_j)) > 0$ for any $X_i \in \mathcal{X}_j$, which implies $P(\mathcal{X}_j \cap s_\varepsilon(F_j) \neq \emptyset) > 0$ for all $n_j > 0$ and also in the limit. Therefore, it follows that $P(\mathcal{X}_j \subset C_j) < 1$ for all $n_j > 0$ and also in the limit, which is a contradiction. $\square$

Lemma 1 shows that the class cover, C_j, almost surely covers the support of its associated class (except perhaps on a region of Lebesgue measure zero) as $n_j \to \infty$. In particular, if the support $s(F_j)$ is bounded, then $P(\lambda(s(F_j) \backslash C_j) = 0) = 1$ for sufficiently large n_j and if the support $s(F_j)$ is unbounded, then $P(\lambda(s(F_j) \backslash C_j) > 0) > 0$ for all n_j but this probability converges to 0 as $n_j \to \infty$.

To show consistency of classifiers based on PCD class covers, we have to investigate the class covers under the assumption of (strict) δ-separability of class supports. Let the two classes be labeled as 0 and 1 with strictly δ-separable continuous distributions (i.e., $\delta > 0$), then a proximity map $N(\cdot)$ satisfying property **P2** establishes *pure* class covers that include none of the non-target class points w.p. 1, i.e. $C_j \cap \mathcal{X}_{1-j} = \emptyset$ w.p. 1. In this case, we have the following lemma showing that the intersection of the cover of the target class and the support of the non-target class is almost surely empty as $n_{1-j} \to \infty$ (except perhaps for a region of Lebesgue measure zero). Let P_j be the probability with respect to distribution F_j for $j = 0, 1$ and P_{01} be with respect to the joint distribution F_{01} of (X, Y) for $X \sim F_0$ and $Y \sim F_1$. Then, **P2** also implies that $P_1(s(F_0) \cap \mathcal{X}_1 = \emptyset) = 1$. Hence, it also follows that $P_{01}(\mathcal{X}_0 \cap \mathcal{X}_1 = \emptyset) = 1$, since $P_0(\mathcal{X}_0 \subset s(F_0)) = 1$.

Lemma 2 *Let the target and the non-target classes be labeled as 0 and 1 and $\mathcal{X}_0 = \{X_1, X_2, \ldots, X_{n_0}\}$ and $\mathcal{X}_1 = \{Y_1, Y_2, \ldots, Y_{n_1}\}$ be two random samples from classes 0 and 1 with class conditional continuous distributions F_0 and F_1 whose supports are strictly δ-separable (i.e., $\delta > 0$) in $\mathbb{R}^d$, respectively. If the proximity map $N(\cdot)$ satisfies properties* **P1** *and* **P2**, *then, for any fixed $n_0 > 0$, we have $\lambda(C_0 \cap s(F_1)) \to 0$ a.s. as $n_1 \to \infty$.*

Proof Let $n_0 > 0$. Notice that $s(F_1)$ is fixed and the randomness in $\lambda(C_0 \cap s(F_1))$ is due to $\mathcal{X}_0$ and $\mathcal{X}_1$, both of which are used in the construction of $N(X)$. Moreover, recall that $C_0 = \cup_{X \in S_0} N(X)$ for $S_0 \subset \mathcal{X}_0$ being a minimum prototype set of $\mathcal{X}_0$. Note that from **P2**, it follows that $P_1(N(x) \cap \mathcal{X}_1 = \emptyset) = 1$ for all $x \in s(F_0)$ and $P_{01}(N(X) \cap \mathcal{X}_1 = \emptyset) = 1$ for $X \sim F_0$. Then, as $n_1 \to \infty$, $P_{01}(C_0 \cap \mathcal{X}_1 = \emptyset) \to 1$ (or equivalently, $P_{01}(C_0 \cap \mathcal{X}_1 \neq \emptyset) \to 0$), since C_0 is the union of $N(X)$ for $X \in S_j \subseteq \mathcal{X}_j$. Now assume, for a contradiction, that $\lambda(C_0 \cap s(F_1)) \to \varepsilon$ w.p. 1 as $n_1 \to \infty$ for some $\varepsilon > 0$. Thus, as $n_1 \to \infty$, there exists a region $s_\varepsilon(F_1)$ in $s(F_1)$ with positive measure such that $P_{01}(C_0 \cap s_\varepsilon(F_1) \neq \emptyset)$ is positive in the limit and hence $P_{01}(C_0 \cap \mathcal{X}_1 \neq \emptyset)$ is positive in the limit, which is a contradiction. $\square$

Recall that PCD cover classifiers are defined with either standard covers which employ only one type of proximity map, or composite covers which employ two (or more) types of proximity maps. On the other hand, hybrid classifiers use cover classifiers for data points

from one class in the convex hull of points from the other class(es), and use an alternative classifier elsewhere.

We show the consistency of cover and hybrid PCD classifiers. That is, e.g., we show that the error rate of the cover classifier $L(g_C)$ converges to the Bayes optimal error rate, which is 0 for continuous class conditional distributions with strictly δ-separable supports as $n_0, n_1 \to \infty$ (Devroye et al. 1996). Then, we have the following theorem.

Theorem 6 *Let $\mathcal{X}_0$ and $\mathcal{X}_1$ be two random samples of size n_0 and n_1 from classes 0 and 1, respectively, such that the data set $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1$ is a random sample from the distribution $F = \pi_0 F_0 + \pi_1 F_1$ for some $\pi_0, \pi_1 \in [0, 1]$ and $\pi_0 + \pi_1 = 1$ where F_0 and F_1 are continuous class conditional distributions with finite dimensional strictly δ-separable supports $s(F_0)$ and $s(F_1)$, respectively. Then we have the following results.*

(1) *Let the cover classifier g_C be based on a standard cover with proximity map $N(\cdot)$ which satisfies **P1** and **P2** or based on a composite cover with proximity maps $N_i(\cdot)$ for $i = 1, \ldots, k$, each of which satisfies **P1** and **P2**. Then g_C is consistent; that is, $L(g_C) \to L^* = 0$ as $n_0, n_1 \to \infty$.*

(2) *Let the hybrid classifier g_H be based on g_C in $C^{\text{in}} := C_0^{\text{in}} \cup C_1^{\text{in}}$ where C_j^{in} is the cover of points $\mathcal{X}_j \cap C_H(\mathcal{X}_{1-j})$ for $j = 0, 1$ and based on an alternative classifier g_A which is different from g_C. If g_C is as in part (1) and g_A is consistent as $n_0, n_1 \to \infty$, then g_H is consistent as $n_0, n_1 \to \infty$.*

Proof

(1) It suffices to prove part (1) for standard cover classifiers, as the extension to the composite cover case is straightforward, since each $N_i(\cdot)$ also satisfies **P1** and **P2**. Let Z be a random variable from F. Then $Z = Z_j \sim F_j$ with probability π_j for $j = 0, 1$.

Then, by Lemma 1, we have $P(Z_j \in C_j) \to 1$ as $n_j \to \infty$. And by Lemma 2, $P(Z_j \in C_j \backslash s(F_{1-j})) \to 1$ as $n_{1-j} \to \infty$ for any fixed $n_j > 0$. Furthermore, $P(Z \in s(F)) = 1$ where $s(F) = s(F_0) \cup s(F_1)$. By Lemmas 1 and 2, as $n_0, n_1 \to \infty$, we have $\lambda(s(F) \backslash C_0 \cup C_1) \to 0$ w.p. 1 and $P(C_0 \cap C_1 \subset s(F)^c) \to 1$ and so $P(Z \in C_0 \triangle C_1) \to 1$ where $C_0 \triangle C_1$ is the symmetric difference between C_0 and C_1. Then, we have $\lambda(C_j \cap s(F_{1-j})) \to 0$ w.p. 1 for $n_j > 0$ as $n_{1-j} \to \infty$. Also, $P(Z_j \in C_j \backslash C_{1-j}) \to \pi_j$ as $n_0, n_1 \to \infty$, so $P(g_C(Z_j) = j) \to \pi_j$ (or equivalently $P(g_C(Z_j) \neq j) \to 0$) for $j = 0, 1$.

Therefore,

$$L(g_C) = \sum_{j=0,1} P(g_C(Z_j) \neq j)$$

$$= \sum_{j=0,1} P(g_C(Z) \neq j | Z \text{ is from class } j) P(Z \text{ is from class } j)$$

$$= P(g_C(Z_0) \neq 0)\pi_0 + P(g_C(Z_1) \neq 1)\pi_1$$

Hence, $L(g_C) \to 0$ as $n_0, n_1 \to \infty$.

(2) First observe that, for $j = 0, 1$, $C_H(\mathcal{X}_j)$ converges to $C_H(s(F_j))$ as $n_j \to \infty$ in the sense that $\lambda(C_H(s(F_j)) \backslash C_H(\mathcal{X}_j)) \to 0$ w.p. 1 as $n_j \to \infty$, which implies, for $X \sim F_j$, $P(X \in C_H(\mathcal{X}_j)) \to 1 = P(X \in s(F_j)) = P(X \in C_H(s(F_j)))$ as $n_j \to \infty$, since $s(F_j) \subseteq C_H(s(F_j))$. Let $s_j^{\text{in}} := s(F_j) \cap C_H(\mathcal{X}_{1-j})$. Without loss of generality, we assume

s_0^{in} or s_1^{in} has positive Lebesgue measure, since, otherwise, $P(g_H(Z) = g_A(Z)) \to 1$ and the result follows, since g_A is consistent. So, C_j^{in} is constructed with $S_j^{\text{in}} \subset s_j^{\text{in}}$ w.p. 1 for $j = 0, 1$. Let F_j^{in} be the distribution restricted to s_j^{in} (which can also be denoted as $F_j|_{s_j^{\text{in}}}$). Then F_0^{in} and F_1^{in} are continuous and strictly δ-separable as well. Furthermore, $g_H = g_C$ for points in C^{in} and g_C is consistent by part (1). For $j = 0, 1$, let $Z_j \sim F_j$ and let Y_j be the event that $Z_j \in C^{\text{in}}$ and $v_j := P(Y_j)$. Since s_0^{in} or s_1^{in} has positive Lebesgue measure, we can not have the case $v_1 = v_2 = 0$. Since C^{in} is not unique, there exist v_j^{sup} and v_j^{inf} such that $v_j^{\text{inf}} \leq \lim_{n_0, n_1 \to \infty} v_j \leq v_j^{\text{sup}}$ where v_j^{sup} corresponds to the supremum of the volume of C^{in} and v_j^{inf} corresponds to the infimum of the volume of C^{in} in the limit. Note also that

$$L(g_H) = P(g_H(Z_0) \neq 0)\pi_0 + P(g_H(Z_1) \neq 1)\pi_1.$$

And, for $j = 0, 1$,

$$
\begin{aligned}
P(g_H(Z_j) \neq j) &= P(g_H(Z_j) \neq j, Y_j) + P(g_H(Z_j) \neq j, Y_j^c) \\
&= P(g_H(Z_j) \neq j|Y_j)P(Y_j) + P(g_H(Z_j) \neq j|Y_j^c)(1 - P(Y_j)) \\
&= P(g_C(Z_j) \neq j|Y_j)P(Y_j) + P(g_A(Z_j) \neq j|Y_j^c)(1 - P(Y_j)).
\end{aligned}
$$

Hence, $L(g_H) \to L_H \leq L_C \max\left(v_0^{\text{sup}}, v_1^{\text{sup}}\right) + L_A\left(1 - \min\left(v_0^{\text{inf}}, v_1^{\text{inf}}\right)\right)$ as $n_0, n_1 \to \infty$. But, as $n_0, n_1 \to \infty$, g_C is consistent by part (1) (i.e., by part (1), $P(g_C(Z_j) \neq j|Y_j) \to 0$ for both $j = 0, 1$), hence $L_C = 0$. Moreover, $\sum_{j=0,1} P(g_A(Z_j) \neq j) \to L_A$ as $n_0, n_1 \to \infty$, since the classifier g_A is consistent with Bayes error being L_A in the limit. Notice also that $L_A = 0$, since the supports of the distributions restricted to the complement of C^{in} are also strictly δ-separable. Hence, $L_H = 0$, which is the desired result.

$\square$

As a corollary to Theorem 6 part (1), we have that classifier g_C of standard and composite covers with proximity maps $N_S(\cdot, \theta)$ for $\theta \in (0, 1]$ and $N_{PE}(\cdot, r)$ for $r > 1$ are consistent; and as a corollary to Theorem 6 part (2), we have that classifier g_H is consistent provided that g_C is based on standard and composite covers with proximity maps $N_S(\cdot, \theta)$ for $\theta \in (0, 1]$ and $N_{PE}(\cdot, r)$ for $r > 1$ and g_A is also consistent. A special case occurs when $r = 1$; that is, observe that $x \in \partial(N_{PE}(x, r = 1))$, and hence $N_{PE}(\cdot, r = 1)$ does not satisfy **P1**. Moreover, in part (1) we showed that a cover PE-PCD classifier is consistent since, as $n_0, n_1 \to \infty$, the PE-PCD cover excludes all non-target class points almost surely and if the support of the target class is bounded, it is a subset of the class cover, or if the support is unbounded, probability of observing a point in the support and outside of cover is zero. To show that the hybrid PE-PCD classifiers are consistent in part (2), we required alternative classifiers to be consistent as well.

In proving Theorem 6 using Lemmas 1 and 2, the assumption of strict δ-separability is crucial. If this assumption is dropped; that is, if $\lambda(s(F_0) \cap s(F_1)) > 0$, then no proximity map $N(\cdot)$ satisfies both **P1** and **P2** and consistency is not guaranteed to follow.

6 Monte Carlo simulations and experiments

In this section, we assess the classification performance of hybrid and cover PE-PCD classifiers. We perform simulation studies wherein observations of two classes are drawn from separate distributions where $\mathcal{X}_0$ is a random sample from a multivariate uniform

distribution $U([0, 1]^d)$ and $\mathcal{X}_1$ is a random sample from $U([v, 1 + v]^d)$ for $d = 2, 3, 5$ with the overlapping parameter $v \in [0, 1]$. Here, v determines the level of overlap between the two class supports. We regulate v in such a way that the overlapping ratio ζ is fixed for all dimensions, i.e. $\zeta = \text{Vol}(s(F_0) \cap s(F_1))/\text{Vol}(s(F_0) \cup s(F_1))$. When $\zeta = 0$, the supports are well separated, and when $\zeta = 1$, the supports are identical: i.e. $s(F_0) = s(F_1)$. Hence, the closer the ζ to 1, the more the supports overlap. Observe that $v \in [0, 1]$ can be expressed in terms of the overlapping ratio ζ and dimensionality d:

$$\zeta = \frac{\text{Vol}(s(F_0) \cap s(F_1))}{\text{Vol}(s(F_0) \cup s(F_1))} = \frac{(1 - v)^d}{2 - (1 - v)^d} \iff v = 1 - \left(\frac{2\zeta}{1 + \zeta}\right)^{1/d}. \quad (13)$$

In this simulation study, we train the classifiers with $n_0 = 400$ and $n_1 = qn_0$ with the imbalance level $q = |\mathcal{X}_1|/|\mathcal{X}_0| \in \{0.1, 0.5, 1.0\}$ and overlapping ratio $\zeta = 0.5$. For values of q closer to zero, classes of the data set are more imbalanced. On each replication, we form a test data with 100 random samples drawn from each of F_0 and F_1, resulting in a test data set of size 200. This setting is similar to a setting used by Manukyan and Ceyhan (2016), who showed that CCCD classifiers are robust to imbalance in data sets. We show that the same robustness extends to PE-PCD classifiers in this article. Using all classifiers, at each replication, we record F-measures for the test data, and also, we record the correct classification rates (CCRs) of each class of the test data separately. We perform these replications until the standard errors of F-measures of all classifiers are below 0.0005. We refer to the CCRs of two classes as "CCR0" and "CCR1", respectively. We consider the expansion parameters $r = 1, 1.1, 1.2, \ldots, 2.9, 3, 5, 7, 9$ for the PE-PCD classifiers. Our hybrid PE-PCD classifiers are referred as PE-SVM, PE-kNN and PE-CCCD classifiers with alternative classifiers SVM, kNN and CCCD, respectively.

Before the main Monte Carlo simulation, we perform a preliminary (pilot) Monte Carlo simulation study to determine the values of optimum parameters of SVM, CCCD and kNN classifiers. The same values will be used for alternative classifiers as well. We train the g_{svm}, g_{cccd} and g_{knn} classifiers, and classify the test data sets for each classifier to find the optimum parameters. We perform Monte Carlo replications until the standard errors of all F-measures are below 0.0005 and record which parameter produced the maximum F-measures among the set of all parameters in a trial. Specifically, on each replication, we (1) classify the test data set with each θ value (2) record the θ values with maximum F-measures and (3) update the count of the recorded θ values. Finally, given a set of counts associated with each θ value, we appoint the θ with the maximum count as the θ^*, the optimum θ (or the best performing θ). Later, we use θ^* as the parameter of alternative classifier g_{cccd} in our main simulations. Optimal parameter selection process is similar for classifiers g_{knn} and g_{svm} associated with the parameters k and γ.

The optimum parameters of each simulation setting are listed in Table 1. We consider parameters of SVM $\gamma = 0.1, 0.2, \ldots, 4.0$, of CCCD $\theta = 0, 0.1, \ldots, 1$ (here, $\theta = 0$ is actually equivalent to $\theta = \epsilon$, the machine epsilon), and of kNN $k = 1, 2, \ldots, 30$. In Table 1, as q and d increase, optimal parameters γ and θ decrease whereas k increases. Manukyan and Ceyhan (2016) showed that dimensionality d may affect the imbalance between classes when the supports overlap. Observe that in Table 1, with increasing d, optimal parameters are more sensitive to the changes in imbalance level q. For the CCCD classifier, $\theta = 1$ is usually preferred when the data set is imbalanced, i.e. $q = 0.1$ or $q = 0.5$. Bigger values of θ are better for the classification of imbalanced data sets, since with $\theta = 1$, the cover of the minority class is substantially bigger which increases the domain influence of the points of the minority class. For θ closer to 0, the class cover of the minority class is much smaller

Table 1 Optimum parameters for SVM, CCCD and kNN classifiers used in the hybrid PE-PCD classifiers

d	q	θ (CCCD)	k (kNN)	γ (SVM)
2	0.1	1	1	3.8
	0.5	1	1	4.0
	1.0	0	3	0.1
3	0.1	1	1	2.3
	0.5	1	1	0.4
	1.0	0	4	0.2
5	0.1	1	1	0.9
	0.5	1	4	0.3
	1.0	1	10	0.1

compared to the class cover of the majority class, and hence, the CCR1 is much smaller. Bigger values of parameter k are also detrimental for imbalanced data sets, the bigger the parameter k, the more likely a new point is classified as the majority class since the points tend to be labelled as the class of the majority of k nearest neighbors. As for the parameter γ, support vectors have more influence over the domain as γ decreases (Wang et al. 2003). Note that $\gamma = 1/(2\sigma^2)$ in the radial basis function (RBF) kernel. The smaller the γ, the bigger the σ. Hence, more points are classified as the majority class with decreasing γ since the majority class has more influence. Thus, bigger values of γ are better for the imbalanced data sets.

Average of F-measures and CCRs of three hybrid PE-PCD classifiers are presented in Fig. 7. For $q = 0.1$, the classifier PE-kNN, for $q = 0.5$, the classifier PE-CCCD and, for $q = 1.0$, the classifier PE-SVM performs better than others. Especially, when the data set is imbalanced, the CCR1 determines the performance of a classifier (thus the F-measure); that is, generally, the better a method classifies the minority class, the better the method performs overall. When the data is balanced (i.e. $q = 1$), PE-SVM is expected to perform
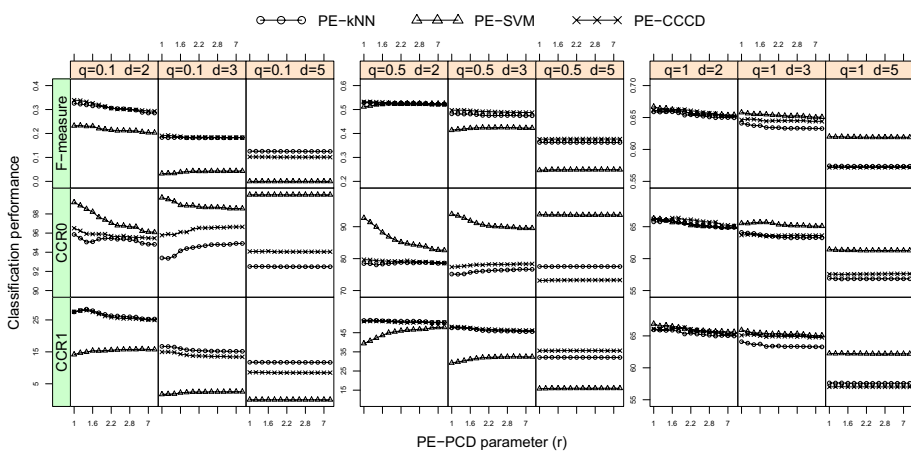

Fig. 7 F-measures and CCRs of the three hybrid PE-PCD classifiers versus expansion parameter $r = 1, 1.2, \ldots, 2.9, 3, 5, 7, 9$ and the alternative classifiers: CCCD, kNN and SVM. The data sets are random samples drawn as $\mathcal{X}_0 \sim U([0, 1]^d)$ and $\mathcal{X}_1 \sim U([\nu, 1 + \nu]^d)$ with several simulation settings based on $\zeta = 0.5$ given the Eq. 13, imbalance level $q = 0.1, 0.5, 1$, and dimensionality $d = 2, 3, 5$

well, however it is known that SVM classifiers are confounded by the imbalanced data sets (Akbani et al. 2004). Moreover, when $q = 0.1$, PE-kNN performs better than PE-CCD. This result contradicts the results of Manukyan and Ceyhan (2016). The reason for this is that hybrid PE-PCD classifiers incorporate alternative classifiers for points outside of the convex hull and kNN might perform better for these points. The kNN classifier is prone to missclassify points closer to the decision boundary when the data is imbalanced, and we expect points outside the convex hull to be far away from the decision boundary in our simulation settings.

In Fig. 7, CCR1 increases while CCR0 decreases for some settings of q and d, and vice versa for some other settings. Recall that Theorem 3 shows a stochastic ordering of the expansion parameter r; that is, with increasing r, there is an increase in the probability of exact MDS being less than or equal to some $\kappa = 1, \ldots, d + 1$. Hence, with increasing r, the proximity region $N_{PE}(x, r)$ gets bigger and the cardinality of the prototype set S_j gets lower. Therefore, we achieve a bigger cover of the minority class and more reduction in the majority class. The bigger the cover is, the higher the CCR1 is in the imbalanced data sets. However, the decrease in the performance, when r increases, may suggest that alternative classifiers perform better for these settings. For example, the CCR1 of PE-SVM increases as r increases for $q = 0.1, 0.5$ and $d = 2, 3$, but CCR1 of PE-CCD and PE-kNN decreases for $r \geq 1.6$. The higher the r, the more the reduction in data set. However, higher values of r may confound the classification performance. Hence, we choose an optimum value of r. Observe that for $d = 5$, the F-measures of all hybrid PE-PCD classifiers are equal for all r. With increasing dimensionality, the probability that a target class point falling in the convex hull of the non-target class points decreases, hence most target class points remain outside of the convex hull of non-target class points.

In Fig. 8, we compare the composite cover PE-PCD classifier and the standard cover PE-PCD classifier. The standard cover is slightly better in classifying the minority class, especially when there is imbalance between classes. In general, the standard cover PE-PCD classifier appear to have higher CCR1 than the composite cover PE-PCD classifiers.
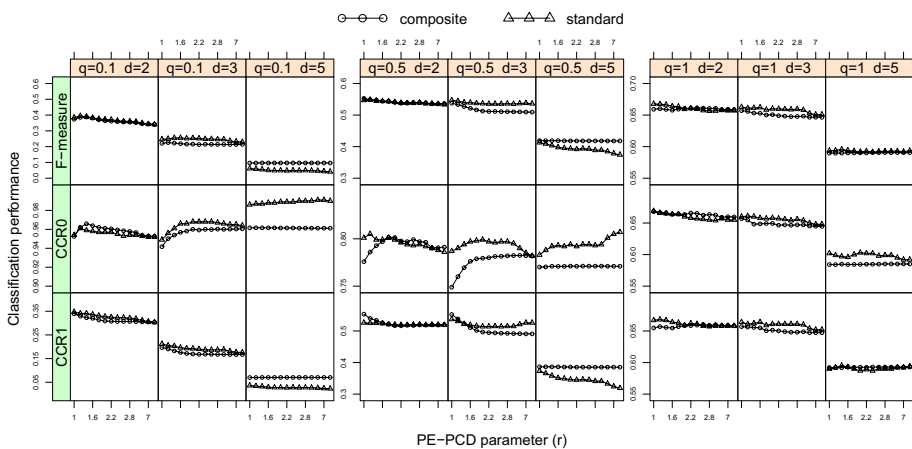

Fig. 8 F-measures and CCRs of the two cover PE-PCD classifiers versus expansion parameter $r = 1, 1.2, \ldots, 2.9, 3, 5, 7, 9$ with composite and standard covers. The data sets are random samples drawn as $\mathcal{X}_0 \sim U([0, 1]^d)$ and $\mathcal{X}_1 \sim U([\nu, 1 + \nu]^d)$ with several simulation settings based on $\zeta = 0.5$ given the Eq. 13, imbalance level $q = 0.1, 0.5, 1$, and dimensionality $d = 2, 3, 5$

However, the composite covers are better when $d = 5$. The PE-PCD class covers are surely influenced by the increasing dimensionality. Moreover, for $q = 0.1, 0.5$, we see that the CCR1 of standard cover PE-PCD classifier slightly decreases with r, even though the data set is more reduced with increasing r. Hence, we should choose an optimum value of r that can still be incorporated to both substantially reduce the data set and to achieve a good classification performance.

In Fig. 9, we compare all five classifiers, three hybrid and two cover PE-PCD classifiers. We consider the expansion parameter $r = 3$ since, in both Figs. 7 and 8, class covers with $r = 3$ perform well and, at the same time, substantially reduce the data set. For all $d = 2, 3, 5$, it appears that all classifiers show comparable performance when $q = 1$, but PE-SVM and SVM give slightly better results. However, when there is imbalance in the data sets, the performances of PE-SVM and SVM degrade, and hybrid and cover PE-PCD classifiers and CCCD classifiers have higher F-measures than others. Compared to all other classifiers, on the other hand, the standard cover PE-PCD classifier is clearly the best performing one for $d = 2, 3$ and $q = 0.1, 0.5$. Observe that the standard cover PE-PCD classifier achieves the highest CCR1 among all classifiers. Apparently, the standard cover constitutes the most robust (to class imbalance) classifier. The performance of standard cover PE-PCD classifier is usually comparable to the composite cover PE-PCD classifier, but slightly better. However, for $d = 5$, the performance of standard cover PE-PCD classifier degrades and composite cover PE-PCD classifiers usually perform better. These results show that cover PE-PCD classifiers are more appealing than hybrid PE-PCD classifiers. The reason for this is that the cover PE-PCD classifiers have both good classification performance and reduce the data considerably more since hybrid PE-PCD classifiers provide a data reduction for only $\mathcal{X}_j \cap C_H(\mathcal{X}_{1-j})$ whereas cover PE-PCD classifiers reduce the entire
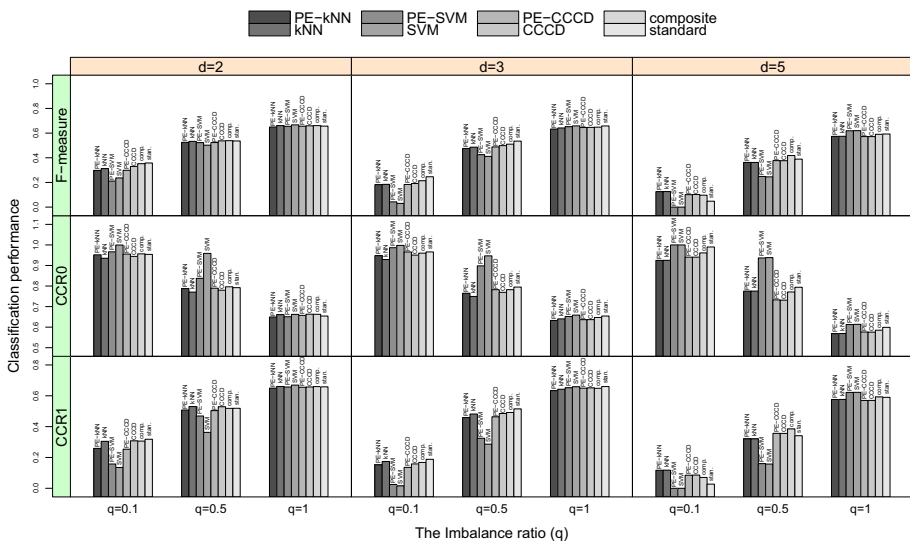

Fig. 9 F-measures and CCRs of the two cover, three hybrid PE-PCD classifiers with expansion parameter $r = 3$, and kNN, SVM and CCCD classifiers. The composite covers are indicated with "comp." and standard covers with "stan.". The data sets are random samples drawn as $\mathcal{X}_0 \sim U([0, 1]^d)$ and $\mathcal{X}_1 \sim U([v, 1 + v]^d)$ with several simulation settings based on $\zeta = 0.5$, imbalance level $q = 0.1, 0.5, 1$ and dimensionality $d = 2, 3, 5$

data set. The level of reduction, however, may decrease as the dimensionality of the data set increases.

In Fig. 10, we compare all five classifiers, three hybrid and two cover PE-PCD classifiers in a slightly different simulation setting where there exists an inherent class imbalance. We perform simulation studies wherein equal number of observations $n_0 = n_1 = n$ are drawn from separate distributions where $\mathcal{X}_0$ is a random sample from a multivariate uniform distribution $U([0, 1]^d)$ and $\mathcal{X}_1$ is a random sample from $U([0.3, 0.7]^d)$ for $d = 2, 3, 5$ and $n = 50, 100, 200, 500$. Observe that the support of one class is entirely inside of the other, i.e. $s(F_1) \subset s(F_0)$. The same simulation setting has been used to highlight the robustness of CCCD classifiers to imbalanced data sets (Manukyan and Ceyhan 2016). In Fig. 10, the performance of kNN and PE-kNN classifiers degrade as d increases and n decreases. With sufficiently high d and low n, the minority class points $\mathcal{X}_0$ is sparsely distributed around the overlapping region of class supports $s(F_1) \cap s(F_0)$ which is the support of $\mathcal{X}_1$. Hence, although the number of observations are equal in both classes, there exists a "local" imbalance between classes (Manukyan and Ceyhan 2016). However, CCCD and SVM classifiers, including the associated hybrid PE-PCD classifiers perform fairly well. Although the cover PE-PCD classifiers have considerably smaller CCR1, they perform relatively well compared to other classifiers and generally have higher CCR0 than other classifiers. Similar to other simulation settings, cover PE-PCD classifiers are also affected by the increasing dimensionality in this setting.

Although the PE-PCD based standard cover classifiers are competitive in classification performance, a case should be made on how much they reduce the data sets during the training phase. In Fig. 11, we illustrate the percentage of reduction in the training data set, and separately, in both minority and majority classes, using PE-PCD for $r = 1, 2, 3$. The overall reduction increases with r, which is also indicated by Theorem 3, and the reduction
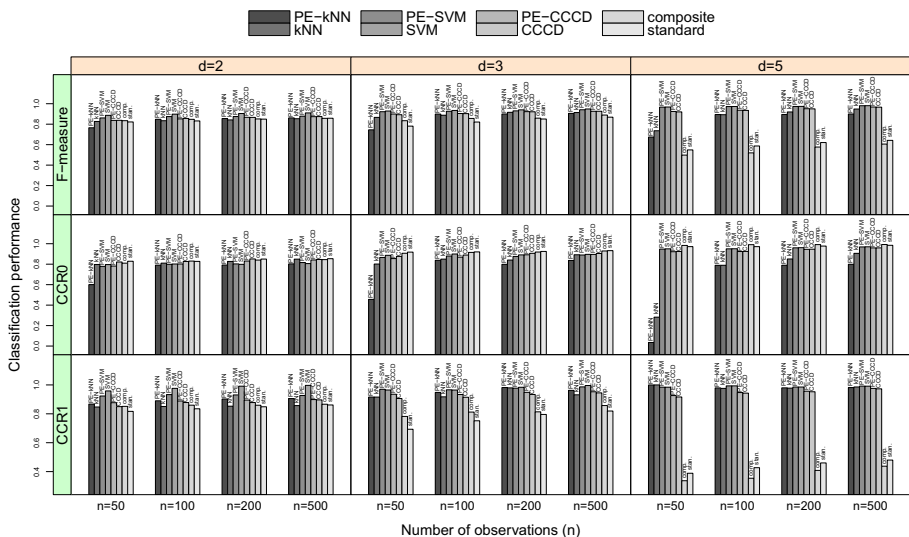

Fig. 10 F-measures and CCRs of the two cover, three hybrid PE-PCD classifiers with expansion parameter $r = 2.2$, and kNN, SVM and CCCD classifiers. The composite covers are indicated with "comp." and standard covers with "stan.". The data sets are random samples drawn as $\mathcal{X}_0 \sim U([0, 1]^d)$ and $\mathcal{X}_1 \sim U([0.3, 0.7]^d)$ with several simulation settings based on number of observations $n = 50, 100, 200, 500$ and dimensionality $d = 2, 3, 5$
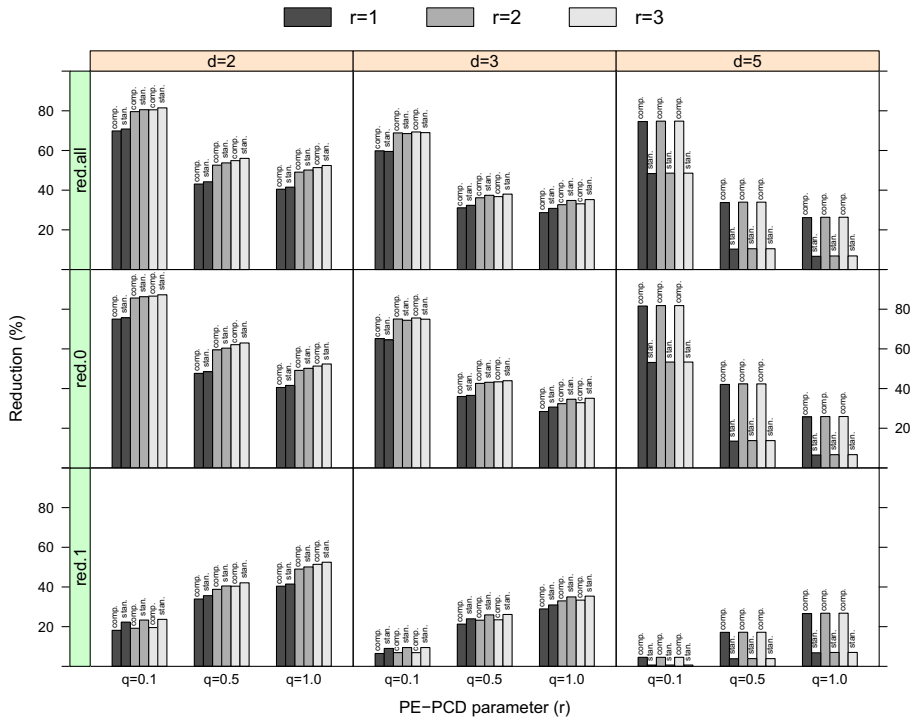
Fig. 11 The percentage of reduction of the composite (comp.) and standard (stan.) PE-PCD covers. The "red.all" indicates the overall reduction in the training data set, $1 - (|S_0 + S_1|/(n_0 + n_1))$, "red.0" the reduction in the $\mathcal{X}_0$ class, $1 - (|S_0|/n_0)$, and "red.1" the reduction in the $\mathcal{X}_1$ class, $1 - (|S_1|/n_1)$. The data sets are random samples drawn as $\mathcal{X}_0 \sim U([0, 1]^d)$ and $\mathcal{X}_1 \sim U([\nu, 1 + \nu]^d)$ with several simulation settings based on $\zeta = 0.5$, imbalance level $q = 0.1, 0.5, 1$ and dimensionality $d = 2, 3, 5$

in the majority class is much more than in minority class when $q = 0.1, 0.5$ since proximity regions of the majority class catch more points unlike the minority class. The majority class is reduced over nearly 60% when $q = 0.1$, and 40% when $q = 0.5$. Indeed, the higher the imbalance between classes, the higher the reduction in the abundantly populated classes. On the other hand, as the dimensionality increases, composite covers reduce the data set more than the standard covers. The number of the facets and simplices increases exponentially with d, and hence the cardinality of minimum dominating set (or the prototype set) also increases exponentially with d (see Theorem 5). As a result, composite PE-PCD covers achieve much higher reduction than standard PE-PCD covers.

7 Real data examples

In this section, we apply the hybrid and cover PE-PCD classifiers on UCI and KEEL data sets (Dua and Graff 2019; Alcalá et al. 2011). Most of these data sets were subjected to preprocessing before analysis such as log transformation, deletion of outliers and missing value imputation. We start with a trivial but a popular data set, `iris` with 150 flowers classified into three types based on their petal and sepal width and lengths. In Fig. 12, we
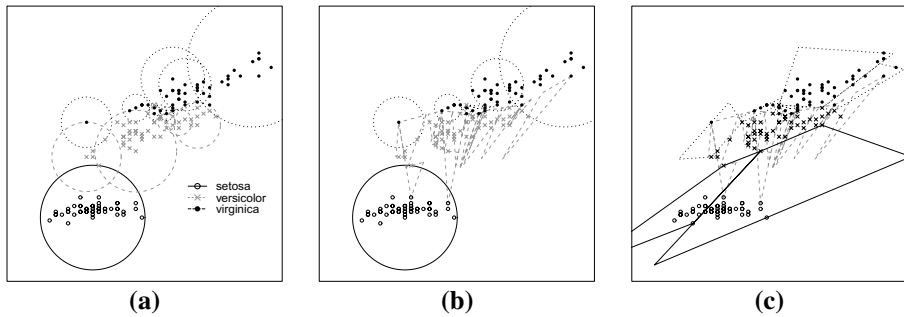
Fig. 12 Class covers of iris data set with variables sepal and petal length. **a** Standard covers with $N_S(\cdot, \theta = 1)$, **b** composite covers with $N_{in}(\cdot) = N_{PE}(\cdot, r = 1)$ and $N_{out}(\cdot) = N_S(\cdot, \theta = 1)$ and **c** standard covers with $N_{in}(\cdot) = N_{out}(\cdot) = N_{PE}(\cdot, r = 1)$

illustrate standard and composite PE-PCD covers, and CCCD covers of the first and the third variables of iris data set, sepal and petal lengths. Observe that in composite covers of Fig. 12b, only a few or no triangles are used to cover the setosa and virginica classes. Points from these classes are almost all outside of the convex hull of the versicolor class points, and hence covered mostly by spherical proximity regions. However, the standard cover of Fig. 12c covers setosa and virginica classes with polygons since these classes are in the outer triangles of the convex hull of the versicolor class.

We first assess the performance of PE-PCD classifiers and other classifiers (i.e. kNN, SVM and CCCD) on two real data sets. The first data set, High Time Resolution Universe Survey (HTRU), is composed of 91192 signals where only 1196 of these signals are labelled as pulsars (Jameson et al. 2010; Morello et al. 2014). A pulsar is a ratio emitting star that was formerly a massive star being on the verge of collapsing. Detecting whether a signal indicates the existence of candidate pulsar is of considerable interest in the field of radio astronomy (Lyon et al. 2016). This data set is preprocessed by Lyon et al. (2016) such that eight variables are generated to predict whether a signal is a pulsar or not. These features are the mean, standard deviations, kurtosis and skewness values of integrated pulse profiles and the DM-SNR (dispersion measure and signal-to-noise) curves. In HTRU data set, there are 1196 pulsar candidates and 89995 non-pulsar (stars that are not pulsars) candidates which makes the data set having an imbalance ratio of $89995/1196 = 75.24$. We randomly split the HTRU data set into training and test data sets which comprise of the 75% and 25% of the original HTRU data set, respectively, where both training and test sets have approximately the same level of imbalance.

Lyon et al. (2016) shows that all eight variables based on the pulse profiles and DM-SNR curves are fundamental for predicting a pulsar signal, but three variables, that are mean, kurtosis and the skewness of the pulse profiles, are more explanatory than the other variables. Also, recall that the number of prototypes in PE-PCD classifier increases exponentially with d as shown by Theorem 5. Simulation studies in Sect. 6 also indicated that the dimensionality of a data set affects the classification performance. Hence, we apply dimensionality reduction to the HTRU data set to mitigate the dimensionality effect. After preprocessing the HTRU data set, we used principal component analysis (PCA) to extract the three principal components with 96% of variation explained. In Fig. 13, we illustrate scatter diagrams of first two principal components. We observe that the classes are almost separated with a mild level of overlap.
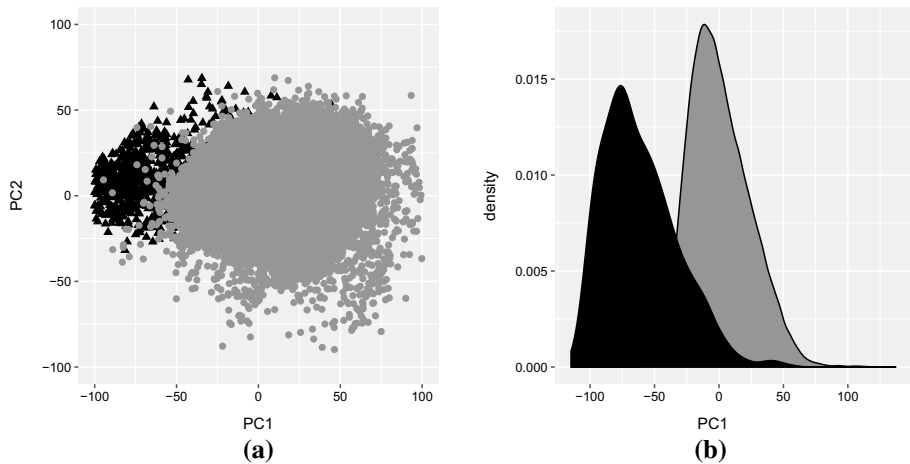
Fig. 13 The first two principal components (PCs) of the HTRU data set. **a** The scatter diagram of first and second PCs. **b** The density plot of pulsar and non-pulsar candidates with respect to the first PC. Here, grey points represent the non-pulsar candidates and black points represent the pulsar candidates

We establish the PE-PCD covers of the HTRU data sets for increasing values of $r = 1, 1.1, 1.2, \ldots, 2$. Also in Fig. 14, for all values of r, we give the levels of reduction and the imbalance ratio of these two classes after reduction, i.e. reduced imbalance ratio. In Fig. 14a, we ignore values $r > 2$ due to no substantial change in either the reduction percentage or reduced imbalance ratio. For increasing values of r, there exists a considerable reduction in the number of observations. With $r = 1$, the percentage of reduction in the number of non-pulsar candidates is almost 99%. Also, the reduction in the pulsar
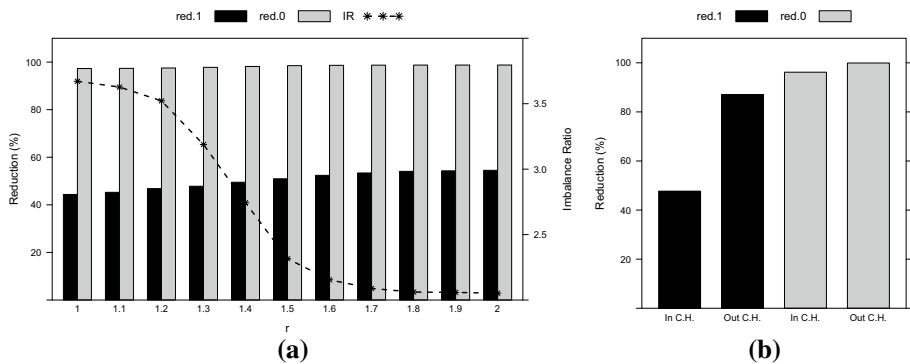

Fig. 14 **a** The percentage of reduction and the reduced imbalance ratios of the HTRU training data sets reduced by PE-PCDs (standard cover) for all values of $r = 1, 1.1, 1.2, \ldots, 2$. We ignore higher values of r due to no considerable change in either imbalance ratio or the reduction. **b** Total reduction in the two target classes that are inside and outside of the convex hull (denoted as "in C.H." and "out C.H.") of the non-target class (for example, if the set of pulsar candidates is the target class, then the non-pulsar candidates constitute the non-target class). Here, the "red.0" indicates the reduction percentage in the non-pulsar candidates (majority class), i.e. $|S_0|/|\mathcal{X}_0|$; and the "red.1" indicates the reduction percentage in pulsar candidates (minority class), i.e. $|S_1|/|\mathcal{X}_1|$. "IR" represents the imbalance ratio between the non-pulsar and pulsar classes, i.e. IR $= |S_0|/|S_1|$

candidates is 40%, reaching up to 50% with increasing values of r. PE-PCDs achieve a reduced imbalance ratio of approximately 3 where the global imbalance ratio of HTRU data set was originally 75.24. In Fig. 14b, we illustrate the reduction inside and outside of the convex hulls $C_H(\mathcal{X}_{1-j})$, for $j = 0, 1$, of both classes for $r = 2$. Both inside and outside of the convex hull of the non-target class, where the set of pulsar candidates is the non-target class, non-pulsar candidates achieve higher than 90% reduction. However, only the 50% of the pulsar candidates in the convex hull of the non-pulsar candidates are chosen as members of the prototype set. In both classes, the reduction is over 90% for those target class points that are outside of the convex hull of the non-target class. The reduction in the minority class (the pulsar candidates) is indeed lower than the reduction in the majority class (non-pulsar candidates), but it results in an undersampling of the majority class which successfully reduces the imbalance between the number of pulsar and non-pulsar signals.

In Fig. 15a, we illustrate the F-measure of standard cover classifier, SVM, kNN and CCCD with parameters $r = 1, 1.1, \dots, 2.9, 3, 5, 6, 7, 8, 9$, $\gamma = 0.1, 0.2, \dots, 4$, $k = 1, 2, \dots, 40$ and $\theta = 0, 0.1, \dots, 1$ measured on the HTRU test data set, respectively. The classification performance of SVM seems to be not affected by the increasing values of γ; that is, SVM achieves approximately 0.80 F-measure for all γ. The parameter $k = 1$ is not the best performing one for kNN; however, the best performance is achieved at $k = 5$ since, looking at Fig. 13, there exists some considerable separation between the pulsar and non-pulsar candidates. Although the HTRU data is highly imbalanced with IR = 75.24, due to the well separation of pulsar and non-pulsar candidates, moderate values of k may perform better. CCCD classifiers for all values of θ, however, achieves the least F-measure in classifying the test data set of HTRU. Moreover, standard cover classifiers with expansion parameter r have higher F-measures than CCCDs, where standard cover performs the best at $r = 1.5$. It was also observed in the simulation studies of Sect. 6 that an optimum value of r is preferred to achieve a considerable level of reduction while keeping the classification performance high. We observe a comparable level of F-measure in both kNN and standard cover classifiers while the cover reduces the number of observations and mitigate the effects of imbalance in classes.

In Fig. 15b, we illustrate the F-measure of hybrid classifiers and the composite cover classifier for PE-PCD parameter $r = 1, 1.1, \dots, 2.9, 3, 5, 6, 7, 8, 9$ measured on the HTRU test data set. The k, γ and θ parameters of the alternative classifiers and the parameter of the spherical proximity regions of the composite cover are fixed to those values that
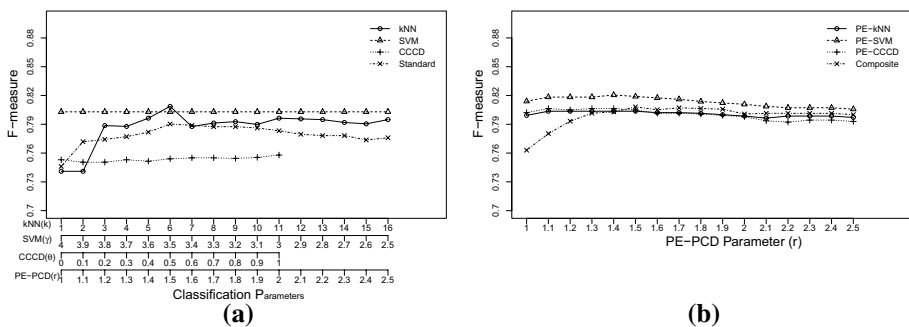

Fig. 15 **a** F-measures of the standard cover, kNN, SVM and CCCD classifiers measured on the HTRU test data set for increasing values of r, θ, k and γ, and (b) F-measure of the composite cover and hybrid classifiers for optimum θ, k and γ along with increasing values of r

performed the best in the experiments of Fig. 15a. We observe that the PE-SVM hybrid classifier slightly outperforms the SVM classifiers with the best performance achieved at $r = 1.4$. PE-CCCDs also have higher F-measure then CCCDs, considerably increasing the prediction accuracy of the CCCDs with the addition of the PE-PCDs. All hybrid classifiers seem to have similar classification performances, even though lower values of r may produce composite cover classifiers with slightly smaller F-measures. The increase in the classification performance of the hybrid classifiers may indicate that correctly classifying the pulsar candidates in the overlapping region of two classes is of higher importance. Both Fig. 15a and b indicate that, if the dimensionality of the data set is sufficiently reduced, standard classifiers and PE-PCD based hybrid classifiers may produce comparable classification performances.

Cover PE-PCD classifiers perform better if the data set has low dimensionality. Hence, we reduce the dimensionality of data sets by means of PCA and then classify the data set with the cover PE-PCD classifiers trained over this data set in the reduced dimension. Although PE-PCD classifiers have computationally tractable MDSs and potentially have comparable performance to those other classifiers, the moderately high dimensionality of the data sets are detrimental for these classifiers based on PE-PCD class covers. Now we apply all classifiers to the Letter data set which is composed of 20,000 black and white pixels. Each of these pixels represents one of the 26 alphabetic letters (Frey and Slate 1991; Dua and Graff 2019), and each pixel is converted into 16 integer features. For the sake of investigating the performance of standard cover classifiers under class imbalance, we restrict our attention to successfully recognizing the letter "M" where only 792 of examples represent this letter. Thus, the transformed data set "LetterM" has an imbalance ratio of $19208/792 = 24.25$. We apply dimensionality reduction to the data set after some preprocessing and extract 5 principal components with total explained variance of 72%. In Fig. 16, we illustrate two of these principal components that illustrate the separability of the LetterM data set. There exists a medium level of separability between classes that may help in achieving a high classification performance. We then randomly split the LetterM data set into training and
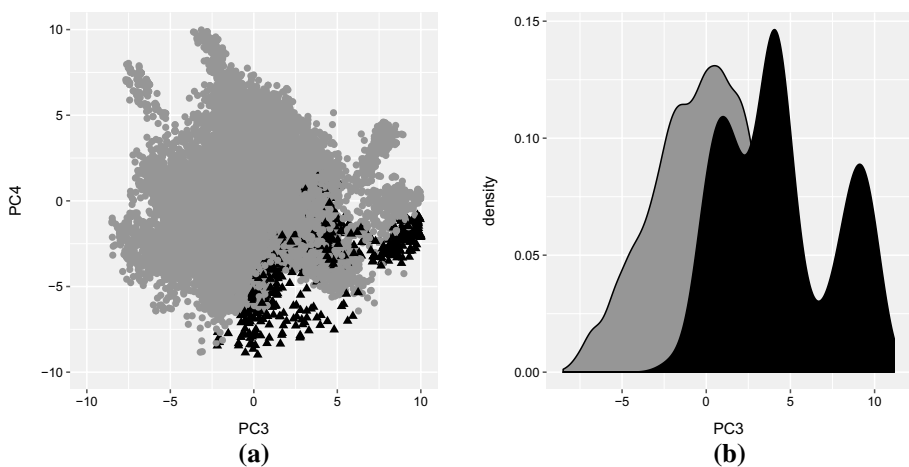

Fig. 16 The third and fourth principal components (PC) of the LetterM data set. **a** The scatter diagram of two PCs. **b** The density plots of the minority and majority class with respect to the first PC. Here, black points represent letters "M" and grey points represent the other letters

test data sets that both constitute equally 50% of all observations of the LetterM data set with the approximately same level of class imbalance.

In Fig. 17, we establish the PE-PCD cover of LetterM data set for all values of r, and then, observe the reduction in the number of observations and the reduced imbalance ratios by means of the prototype sets. Although the LetterM data set exhibits some level of separability between classes similar to HTRU data set, the reduction in the minority class is far less than what was from HTRU. Most reduction is achieved in the observations outside of convex hull of the non-target class, but the number of observations in the minority class is only reduced to 40% outside of the convex hull. Here, the convex hull becomes smaller compared to the entire domain with increasing dimensionality. Outside the convex hull of the minority class, we observe a 90% reduction in the majority class while the minority class achieves only 20% reduction, but the reduced imbalance level is approximately 2.4. The reduced imbalance ratio does not considerably change with increasing expansion parameter r since most points are outside of the convex hull and, by the Theorem 4, the number of prototypes (dominating points) outside of the convex hull is fixed for all r.

In Fig. 18a, we illustrate the F-measure of standard cover classifier, SVM, kNN and CCCD measured on the LetterM test data set. Contrary to the performance of standard cover classifier on HTRU dataset, the cover classifier based on the standard PE-PCD cover shows much worse performance compared to the other classifiers. There is almost no change in the performance of hybrid classifiers with increasing r since only few number of target class points(es) fall into the convex hull of the non-target class(es). Composite cover classifiers achieve nearly 0.70 F-measure with increasing r but both PE-CCCD hybrid and CCCD classifier outperform the composite cover classifier. CCCDs, kNN, SVM and all hybrid classifiers have nearly 0.80 F-measure on the test data while the cover classifiers achieve F-measures approximately between 0.60 and 0.70. The best performing values of r for composite and standard covers are $r = 1.7$ and $r = 1.9$, respectively. The F-measure seems to be stable for increasing values of θ and γ, but lower values of k perform better for kNN. Although it is possible to achieve approximately 0.80 F-measure with other classifiers, the standard cover classifiers suffer from model complexity of the PE-PCD cover which depends on d, dimensionality of the data set. This is again due to the results of Theorem 5 where a dominating set of PE-PCD $|S_j|$ is of complexity $\mathcal{O}\left(dn_{1-j}^{[d/2]}\right)$. Here, the car-
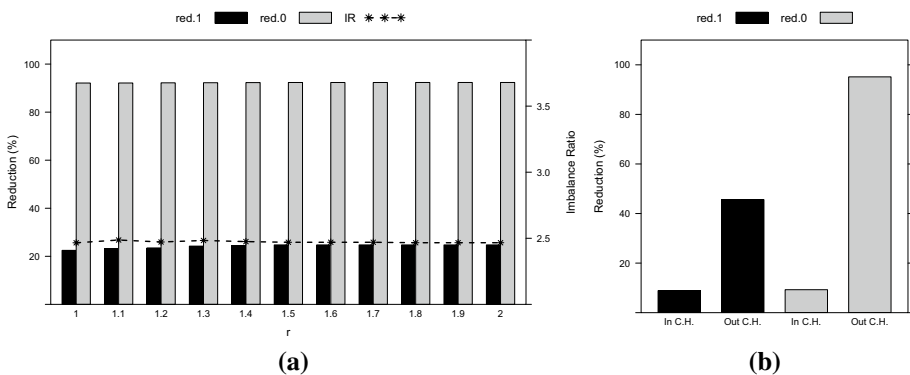

Fig. 17 The percentage of reduction and the reduced imbalance ratios (**a**) and the total reduction (**b**) of the LetterM data sets. Majority class is letters different from M, and minority class is the letter M. The description and labeling of the plots are as in Fig. 14
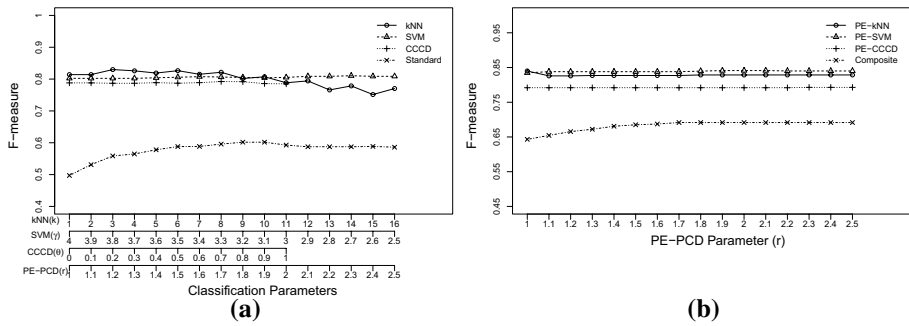
Fig. 18 **a** *F*-measures of the standard cover, *k*NN, SVM and CCCD classifiers measures on LetterM test data set for increasing values of r, θ, k and γ, and **b** *F*-measure of the composite cover and hybrid classifiers for optimum θ, k and γ along with increasing values of r

dinality of the dominating set increases exponentially on d which is the total exact minimum number of d-simplices and outer-simplices needed to cover the target class. Although the dimensionality reduction may be helpful in reducing the complexity of the class cover, one may not be able to reduce the dimensionality of the data in such a way that the reduced data set has considerably high fraction of explained variance and still be eligible to be trained by PE-PCD based classifiers.

In Table 2, we apply all classifiers on 17 UCI and KEEL data sets including iris, HTRU and LetterM data sets. For testing the statistical difference between the *F*-measures of classifiers, we employ the combined 5×2 CV *F*-test (Dietterich 1998; Alpaydın 1999). We also use micro *F*-measure for data sets with multiple number of classes since micro *F*-measure is more suitable for multiple imbalanced classes than macro *F*-measure (Narasimhan et al. 2016). The test works as an omnibus test for all ten possible 5×2 CV *t*-tests (for each five repetitions there are two folds, hence ten folds in total). Basically, if a majority of ten 5×2 CV *t*-tests suggest that two classifiers are significantly different in terms of performance, the *F*-test also suggests a significant difference. Hence, an *F*-test with high *p*-value suggests that some of the ten *t*-tests fail to reject the null-hypothesis (i.e. they have high *p*-value); that is, it is very likely that there exist no significant difference between the *F*-measures of two classifiers. We only report the *p*-values of the difference between the *F*-measures of standard cover classifier with all other classifiers (including composite cover and hybrid classifiers). In Table 2, we report on the *F*-measures of all classifiers along with the optimum values of associated tuning parameters. We either reduce the dimensionality of each data set, empirically select some subset of features or use the original set of features of each data set (hence called the unreduced data set). We report the best performing number of extracted or selected features for each data set. We mostly avoid applying PE-PCD based hybrid and cover classifiers to unreduced data sets due to the high model complexity of PE-PCDs with moderately high number of dimensions (or features), hence we only apply SVM, *k*NN and CCCD classifiers to these unreduced data sets.

Alongside with Table 2, we report on the reduced imbalance ratios (the ratio between the majority and minority class after PE-PCDs are applied and the prototype set is extracted), and reduction rates of standard and composite cover classifiers of all these 17 data sets in Table 3. Here, we employ a multi class imbalance ratio notation which indicates the reduced imbalance ratio of all classes with one class being the reference class; for example, IR $= n_3/n_1 \mid n_2/n_1 \mid n_1/n_1 = n_3/n_1 \mid n_2/n_1 \mid 1$. We report on the

Table 2 Average (± standard deviations) of ten folds of 5 × 2 CV F-test F-measures (or micro F-measure for data sets with multiple classes) values of all classifiers on 17 eleven KEEL and UCL data sets

Data set	d	PE-kNN	kNN	PE-SVM	SVM	PE-CCCD	CCCD	Composite	Standard
iris	4	$d=4$	$d=4$	$d=4$	$d=4$	$d=4$	PCd = 2 (85%)	PCd = 2 (85%)	PCd = 2 (85%)
		$r=1.8\,k=7$	$k=7$	$r=1.8\,\gamma=0.3$	$\gamma=0.3$	$r=1.8\,\theta=0.2$	$\theta=0.9$	$r=8\,\theta=0.9$	$r=8$
		0.962 ± 0.018	$\mathbf{0.967\pm0.014}$	0.954 ± 0.022	0.958 ± 0.020	0.946 ± 0.027	0.957 ± 0.022	$\mathbf{0.935\pm0.025}$	0.918 ± 0.041
Wine	13	PCd = 3 (44%)	PCd = 2 (33%)	PCd = 2 (33%)	$d=13$	PCd = 2 (33%)	PCd = 2 (33%)	PCd = 2 (33%)	PCd = 2 (33%)
		$r=1.3\,k=26$	$k=24$	$r=1.4\,\gamma=1$	$\gamma=0.1$	$r=1.4\,\theta=0.4$	$\theta=0.4$	$r=1.4\,\theta=0.4$	$r=1.4$
		0.975 ± 0.013	0.972 ± 0.019	0.983 ± 0.010	0.983 ± 0.010	$\mathbf{0.987\pm0.009}$	0.961 ± 0.022	$\mathbf{0.987\pm0.009}$	0.983 ± 0.014
Thyroid1	4	$d=4$	$d=4$	$d=4$	$d=4$	$d=4$	$d=4$	$d=4$	PCd = 2 (88%)
		$r=2.6\,k=1$	$k=1$	$r=2.6\,\gamma=0.2$	$\gamma=0.2$	$r=2.6\,\theta=1$	$\theta=1$	$r=2.6\,\theta=1$	$r=1.3$
		0.892 ± 0.043	0.892 ± 0.043	$0.951\pm0.036^{*}$	$0.951\pm0.036^{*}$	0.870 ± 0.073	0.870 ± 0.073	0.870 ± 0.073	$\mathbf{0.878\pm0.059}$
Thyroid2	4	$d=4$	$d=4$	$d=4$	$d=4$	$d=4$	$d=4$	$d=4$	PCd = 2 (88%)
		$r=1.9\,k=1$	$k=1$	$r=1.9\,\gamma=0.2$	$\gamma=0.2$	$r=1.9\,\theta=0.7$	$\theta=0.7$	$r=1.9\,\theta=0.7$	$r=9$
		0.885 ± 0.069	0.885 ± 0.069	$0.947\pm0.049^{**}$	$0.947\pm0.049^{***}$	0.861 ± 0.071	0.861 ± 0.071	0.861 ± 0.071	$\mathbf{0.898\pm0.036}$
Ionosphere	33	PCd = 7 (52%)	PCd = 7 (52%)	PCd = 7 (52%)	$d=22$	PCd = 7 (52%)	$d=22$	PCd = 7 (52%)	PCd = 4 (36%)
		$r=1.6\,k=1$	$k=1$	$r=1.6\,\gamma=0.4$	$\gamma=0.1$	$r=1.6\,\theta=0.6$	$\theta=1$	$r=1.6\,\theta=0.6$	$r=1.8$
		0.833 ± 0.032	0.833 ± 0.032	$0.921\pm0.015^{**}$	$0.930\pm0.018^{**}$	$0.895\pm0.021^{**}$	$0.919\pm0.028^{**}$	$\mathbf{0.837\pm0.043}$	0.793 ± 0.021
Banknote	4	$d=4$	$d=4$	$d=4$	$d=4$	$d=4$	$d=4$	$d=4$	$d=4$
		$r=1.6\,k=1$	$k=1$	$r=1.6\,\gamma=4$	$\gamma=4$	$r=1.6\,\theta=0.8$	$\theta=0.8$	$r=1.6\,\theta=0.8$	$r=1.6$
		1.000 ± 0.000	$\mathbf{1.000\pm0.000}$	$\mathbf{1.000\pm0.000}$	$\mathbf{1.000\pm0.000}$	0.996 ± 0.005	0.996 ± 0.005	0.996 ± 0.005	$\mathbf{0.998\pm0.004}$
Yeast4	8	PCd = 5 (76%)	$d=8$	PCd = 5 (76%)	PCd = 5 (76%)	PCd = 5 (76%)	$d=8$	PCd = 5 (76%)	PCd = 2 (40%)
		$r=9\,k=1$	$k=1$	$r=9\,\gamma=0.6$	$\gamma=0.6$	$r=9\,\theta=1$	$\theta=0.9$	$r=9\,\theta=1$	$r=2.8$
		0.248 ± 0.069	$\mathbf{0.315\pm0.075}$	0.169 ± 0.074	0.159 ± 0.079	0.225 ± 0.051	0.285 ± 0.091	$\mathbf{0.194\pm0.072}$	0.160 ± 0.084
Yeast1289vs7	8	PCd = 3 (44%)	$d=8$	PCd = 2 (31%)	$d=8$	PCd = 5 (68%)	PCd = 5 (68%)	PCd = 5 (68%)	PCd = 2 (31%)
		$r=4\,k=1$	$k=1$	$r=5\,\gamma=4$	$\gamma=0.7$	$r=7\,\theta=1$	$\theta=1$	$r=7\,\theta=1$	$r=5$
		0.215 ± 0.112	$\mathbf{0.233\pm0.092}$	0.145 ± 0.067	0.024 ± 0.050	0.176 ± 0.065	0.174 ± 0.064	0.135 ± 0.073	$\mathbf{0.147\pm0.069}$

Table 2 (continued)

Data set	d	PE-kNN	kNN	PE-SVM	SVM	PE-CCCD	CCCD	Composite	Standard
Yeast5	8	PCd = 4 (57%)	d = 8	PCd = 4 (57%)	d = 8	PCd = 4 (57%)	d = 8	PCd = 4 (57%)	PCd = 2 (31%)
		r = 6 k = 1	k = 4	r = 6γ = 0.5	γ = 0.2	r = 6θ = 1	θ = 1	r = 6θ = 1	r = 8
		0.578 ± 0.052*	**0.660 ± 0.081***	0.504 ± 0.110	0.543 ± 0.110	0.540 ± 0.086**	0.577 ± 0.079	**0.430 ± 0.080**	0.415 ± 0.098
Yeast6	8	PCd = 4 (57%)	d = 8	PCd = 4 (57%)	d = 8	PCd = 4 (57%)	d = 8	PCd = 4 (57%)	PCd = 2 (31%)
		r = 6 k = 4	k = 8	r = 6γ = 1.3	γ = 0.5	r = 6θ = 0.5	θ = 0.6	r = 7θ = 0.9	r = 7
		0.511 ± 0.063**	**0.588 ± 0.081***	0.358 ± 0.103	0.513 ± 0.091**	0.458 ± 0.059	0.563 ± 0.082*	**0.334 ± 0.095**	0.318 ± 0.127
Shuttle0vs4	9	PCd = 3 (59%)	PCd = 3 (59%)	PCd = 2 (43%)	PCd = 2 (43%)	PCd = 2 (43%)	PCd = 2 (43%)	PCd = 2 (43%)	PCd = 2 (43%)
		r = 1.4 k = 1	k = 1	r = 1.2 γ = 0.1	γ = 0.1	r = 1.2 θ = 0	θ = 0	r = 1.2 θ = 0	r = 1.2
		1.000 ± 0.000	**1.000 ± 0.000**	1.000 ± 0.000	**1.000 ± 0.000**	0.998 ± 0.004	0.998 ± 0.004	**0.998 ± 0.004**	0.997 ± 0.004
Ozone	72	PCd = 2 (24%)	d = 72	PCd = 3 (30%)	PCd = 3 (30%)	PCd = 3 (30%)	d = 72	PCd = 3 (30%)	PCd = 2 (24%)
		r = 9 k = 3	k = 3	r = 3 γ = 2.5	γ = 2.5	r = 3θ = 1	θ = 0.9	r = 3θ = 1	r = 9
		0.086 ± 0.036	**0.240 ± 0.060**	0.054 ± 0.005	0.006 ± 0.018	0.136 ± 0.048	0.178 ± 0.107	0.113 ± 0.042	**0.124 ± 0.065**
Segment0	18	PCd = 3 (88%)	d = 18	PCd = 5 (98%)	d = 16	PCd = 5 (98%)	d = 18	PCd = 5 (98%)	PCd = 5 (98%)
		r = 3 k = 1	k = 1	r = 2.4 γ = 0.4	γ = 0.4	r = 2.4 θ = 1	θ = 0.9	r = 2.4 θ = 1	r = 2.4
		0.469* ± 0.029	0.970 ± 0.009	0.987 ± 0.004*	**0.987 ± 0.004***	0.916 ± 0.014**	0.920 ± 0.014**	0.915 ± 0.014**	**0.961 ± 0.012**
Pageblocks0	10	PCd = 4 (99%)	PCd = 4 (99%)	PCd = 3 (98%)	PCd = 4 (99%)	PCd = 4 (99%)	PCd = 4 (99%)	PCd = 4 (99%)	PCd = 4 (99%)
		r = 1.9 k = 1	k = 1	r = 6γ = 0.9	γ = 3.2	r = 1.9θ = 0.4	θ = 0.4	r = 1.9θ = 0.4	r = 1.9
		0.611 ± 0.058	0.603 ± 0.072	0.347 ± 0.044	0.589 ± 0.067	0.555 ± 0.056	0.558 ± 0.067	0.420 ± 0.078	**0.436 ± 0.092**
AlivaD	10	d = 2	d = 2	d = 2	d = 10	d = 2	d = 2	d = 2	d = 2
		r = 1.7 k = 1	k = 1	r = 1.7γ = 0.1	γ = 0.6	r = 1.7θ = 0.6	θ = 0.6	r = 1.7θ = 0.6	r = 1.7
		0.957 ± 0.005	**1.000 ± 0.000***	0.9490 ± 0.011	0.727 ± 0.024***	0.957 ± 0.005	0.985 ± 0.005**	**1.000 ± 0.000***	0.972 ± 0.007
LetterM	16	PCd = 5 (72%)	d = 16	PCd = 5 (72%)	d = 16	PCd = 5 (72%)	d = 16	PCd = 4 (65%)	PCd = 4 (65%)
		r = 2.3 k = 4	k = 1	r = 2.3 γ = 0.3	γ = 4	r = 2.3 θ = 0.8	θ = 0.9	r = 2.2 θ = 0.7	r = 2.2
		0.843 ± 0.018**	**0.973 ± 0.004***	0.840 ± 0.013**	0.931 ± 0.015**	0.810 ± 0.015**	0.947 ± 0.006**	**0.707 ± 0.025***	0.653 ± 0.029

Table 2 (continued)

Data set	d	PE-kNN	kNN	PE-SVM	SVM	PE-CCCD	CCCD	Composite	Standard
HTRU	8	PCd = 3 (96%)	d = 8	PCd = 3 (96%)	d = 8	PCd = 3 (96%)	d = 8	PCd = 3 (96%)	PCd = 3 (96%)
		r = 1.9 k = 9	k = 12	r = 1.9 γ = 1.3	γ = 4	r = 1.9 θ = 0.1	θ = 0	r = 1.9 θ = 0.1	r = 1.9
		0.798 ± 0.008**	0.823 ± 0.010**	0.801 ± 0.008**	**0.863 ± 0.012****	0.794 ± 0.011**	0.821 ± 0.011**	**0.791 ± 0.013****	0.779 ± 0.012

Maximum F-measures (or micro F-measures) of hybrid or alternative classifiers, and cover classifiers are given in bold for each data set

The single asterisk symbol * indicates a difference with the F-measure of standard cover PE-PCD classifier at significant level of 0.1, and the double asterisk symbol ** at level 0.05. The letter "d" represents the number of selected features and "PCd" represents the number of extracted features by PCA along with percentage of explained variance. The optimum (best performing) parameters of r, k, γ and θ for all classifiers are given

Table 3 Characteristics of each data set in Table 2 given the number of observations, number of features, local (with optimum σ of one class SVM) and global imbalance levels. "O.R." represents the percentage of minority class members in the overlapping region, the "Red" represents the fraction of reduction (in decimal) and "Red.IR" represents the reduced imbalance ratio of the reduced data sets (the prototype sets) given by composite and standard covers, separately

Data	Size	d	q = m/n	q(E)	σ	O.R. (%)	Composite		Standard	
							Red	Red.IR.	Red	Red.IR.
Iris	150	4	1|1|1	0.50 (2vs3)	3	6	0.16 ± 0.03	2.09 | 1 | 2.15	0.18 ± 0.03	1 | 5.95 | 3.95
Wine	178	13	1 | 1.20 | 0.83	NA	–	0	0.12 ± 0.03	4.47 | 3.03 | 1	0.12 ± 0.03	4.6 | 2.7 | 1
New-Thyroid1	215	5	5.14	NA	–	0	0.07 ± 0.05	1.06 ± 0.53	0.07 ± 0.02	1.08 ± 0.37
New-Thyroid2	215	5	5.14	NA	–	0	0.08 ± 0.01	0.83 ± 0.12	0.07 ± 0.02	1.12 ± 0.33
Ionosphere	351	33	1.78	70.5	7	2	0.32 ± 0.04	1.76 ± 0.96	0.43 ± 0.03	1.42 ± 0.63
Banknote	1372	4	1.24	1.65	8	9	0.03 ± 0.00	1.08 ± 0.03	0.06 ± 0.00	0.26 ± 0.06
Yeast4	1484	8	28.10	30.19	5	72	0.09 ± 0.00	1.94 ± 0.18	0.08 ± 0.00	1.72 ± 0.09
Yeast1289vs7	1484	8	30.70	26.42	5	63	0.11 ± 0.01	2.77 ± 0.31	0.08 ± 0.01	1.69 ± 0.16
Yeast5	1484	8	32.70	1.35	6	68	0.04 ± 0.00	1.01 ± 0.19	0.05 ± 0.01	1.04 ± 0.18
Yeast6	1484	8	41.40	4.52	3	66	0.05 ± 0.01	1.53 ± 0.19	0.05 ± 0.01	1.41 ± 0.19
Shuttle0vs4	1829	9	13.87	NA	-	0	0.003 ± 0.00	0.75 ± 0.26	0.003 ± 0.00	1.05 ± 0.69
Ozone	1848	72	31.42	NA	5	12	0.10 ± 0.01	2.45 ± 0.22	0.08 ± 0.01	1.71 ± 0.10
Segment0	2308	18	6.02	NA	2	2	0.05 ± 0.00	1.30 ± 0.1	0.07 ± 0.01	1.34 ± 0.01
Pageblocks0	5473	10	46.59	6.16	3	33	0.04 ± 0.00	1.53 ± 0.10	0.05 ± 0.00	1.85 ± 0.15
AlivaD	20,867	10	28.59	13.65	5	79	0.01 ± 0.00	2.32 ± 0.00	0.01 ± 0.00	2.05 ± 0.00
LetterM	20,000	16	27.36	0.99	6	17	0.08 ± 0.00	1.99 ± 0.12	0.09 ± 0.00	2.13 ± 0.12
HTRU	91,192	8	75.24	122.893	6	26	0.02 ± 0.00	0.45 ± 0.03	0.02 ± 0.00	0.46 ± 0.03

global imbalance ratio $q = |\mathcal{X}_0|/|\mathcal{X}_1|$ for $|\mathcal{X}_0|$ being cardinality of the majority and $|\mathcal{X}_1|$ of minority class. We also report on the local imbalance ratio of two classes and the percentage of minority class members in the overlapping region of two classes. Local imbalance ratio and the overlapping ratio of imbalanced classes have also been investigated by Manukyan and Ceyhan (2016), who showed that, although two classes are balanced, there may be some region $E \subset \mathbb{R}^d$ that two classes show some level of imbalance, and this region E is usually where subsets of two classes are close in proximity, e.g., where two classes overlap. We employ one-class SVMs with radial basis function (RBF) kernels to estimate the support of each class with $\nu = 0.01$ and an optimum γ. We calculate the imbalance ratio of two classes within the overlapping region of their estimated supports, i.e. the local class imbalance. We also provide the percentage of minority class members falling into this region. The emphasis is that, the higher the (local) class imbalance or the higher the percentage of minority class members in the overlapping region, the smaller the F-measure since it would be harder to predict the true labels of the minority class (Manukyan and Ceyhan 2016).

In Tables 2 and 3, we observe that composite and standard cover classifiers usually perform the best in lower dimensions; that is, dimensionality reduction helps increasing the F-measure and mitigating the drawbacks caused by moderately high number of features. However, in order for the feature selection or extraction to succeed, the reduced set of features should be explanatory enough to help classifiers achieve a considerable performance. We select two features from the AlivaD data set where we aim to predict if a letter recognized as "D" or not. Hence, we reduce the dimensionality of the data set drastically, and as a result, cover classifiers nearly achieve the Bayes optimal performance. However, applying the classifiers to Pageblocks0 data set, we require four extracted principal components to achieve a 99% explained variance where cover classifiers perform poorly against the other classifiers because of the rapidly increasing complexity of the PE-PCD based covers. A similar argument can be made for LetterM data set since four principal components are clearly not enough. F-measure is naturally effected by the overlapping ratio of two classes since it would be harder to correctly predict the minority class members closer to the points of the majority class. As seen in Table 3, the overlapping ratios of data sets like Iris, Thyroid (1 and 2) and Banknote are small and these data sets are also quite balanced. Hence, there are a handful of minority class members in the support of the majority class which results in a high F-measure. There are some data sets however that, although they exhibit some global class imbalance, they do not have any local imbalance within the overlapping region. Shuttle0vs4 and Segment0 data sets are examples of such cases where all classifiers, including the cover classifiers, perform well even in reduced dimensions with PE-kNN classifier being an exception; that is, a hybrid of PE-PCD and kNN classifiers drastically confounds the performance. On the other hand, Yeast data sets have high overlapping ratios and both local and global imbalance ratios of these data sets are also notably high. Although PE-PCD based (hybrid and cover) classifiers perform the best in lower dimensions, this may result in a considerable loss of information since the percentage of explained variance is, for example in Yeast data sets, nearly 35%. Therefore, all other (nonhybrid or non-cover) classifiers enjoy high F-measure due to the employment of all the features (of unreduced data set). PE-PCD based cover classifiers may also perform well for classifying data sets like Ionosphere and Ozone data with moderately high number of features; that is, after dimensionality reduction, standard cover classifier achieves comparable performance to other classifiers but only underperforms against SVM and PE-SVM classifiers in classifying the Ionosphere data set. Hybrid classifiers often perform comparable to their non-hybrid counterparts, but in some cases, they slightly increase the classification

performance; see for example, the *F*-measures of PE-SVM classifiers in Yeast 4 and Ozone data sets.

In Table 3, we observe that all data sets with number of observations bigger than 5000 are reduced to at most 10% of the original number of observations; that is, cover classifiers prune almost 90% of all observations by only choosing nearly 10% of all observations as the members of the prototype (i.e. minimum dominating) set. Moreover, both types of covers reduce the imbalance between two classes; that is, in all data sets, the reduced imbalance ratio is nearly 2 or lower. The level of reduction of PE-PCD covers is highly dependant on the dimensionality. In Table 2, standard and composite cover classifiers perform better in Ozone with fewer dimensions than in Ionosphere data set, hence the cardinality of dominating sets in both covers are higher in Ionosphere data where approximately 35% of all observations constitute the dominating set. The exponentially increasing complexity of the Delaunay tessellation effects both the performance and the model complexity of the PE-PCD covers. Optimal values of k, γ, θ and r is also similarly affected by the global (or local) class imbalance levels and the dimensionality. Lower values of k perform better for locally imbalanced data sets like Yeast4 and Yeast1289vs7, but γ is mostly effected by the dimensionality. The higher the dimension d or more the extracted features, the lower the gamma, and hence the higher the bandwidth (σ) of the RBF kernel. But most importantly, there is a positive trend in γ as the class imbalance increases. CCCDs mostly perform best in unreduced data sets, therefore higher values of θ are preferred. However, Shuttle0vs4 is a well separated data set with CCCD achieving higher *F*-measure in reduced dimension, and hence, optimal θ is set to the lowest, i.e. $\theta = 0$. The cardinality of minimum dominating set of the PE-PCDs decrease with r, but as it is also demonstrated in Sect. 6, a high value of r is detrimental for the performance of the PE-PCD based cover classifiers. Hence, in most data sets, moderate values of r achieve the best *F*-measure. One apparent difference in optimum values of r among data sets is between locally imbalanced and balanced data sets. Yeast5, Yeast6, LetterM and Segment0 data sets have high global imbalance but low local imbalance despite the fact that majority and minority classes overlap. An optimum value of r for "global only" imbalanced data sets is high which undersamples majority classes as much as possible. The bias in globally imbalanced data sets is originated from the abundant number of majority class members closer to the decision boundary but not in the overlapping region, hence higher the r the better the performance.

8 Summary and discussion

We use *proximity catch digraphs* (PCDs) to construct semi-parametric classifiers that show potential in solving problems with substantial class imbalance. These families of random geometric digraphs constitute class covers of a class of interest (i.e. the target class) in order to generate decision-boundaries for classifiers. PCDs are generalized versions of Class Cover Catch Digraphs (CCCDs). For imbalanced data sets, CCCDs showed better performance than some other commonly used classifiers in previous studies (Manukyan and Ceyhan 2016; DeVinney et al. 2002). CCCDs are actually examples of PCDs with spherical proximity maps. Our PCDs, however, are based on simplicial proximity maps, e.g. proportional-edge (PE) proximity maps. Our PCD, or PE-PCD, class covers are extended to be unions of simplicial and polygonal regions whereas original PE-PCD class covers were composed of only simplicial regions. The most important advantage of these family of PE proximity maps is that their respective digraphs, or

namely PE-PCDs, have computationally tractable minimum dominating sets (MDSs). The class covers of such digraphs are minimum in complexity, offering maximum reduction of the entire data set with comparable and, potentially, better classification performance. PE-PCDs are one of many PCD families using simplicial proximity maps investigated in Ceyhan (2010). Their construction is also based on the Delaunay tessellations of the non-target class, and similar to PE-PCDs, they enjoy various properties that CCCDs do in $\mathbb{R}$, and they can also be used to establish PCD classifiers.

The PE-PCDs are defined in the Delaunay tessellation of the points from the non-target class (i.e. the class not of interest). PE-PCDs, and associated proximity maps, were only defined for the points inside of the convex hull of the non-target class points, $C_H(\mathcal{X}_{1-j})$, in previous studies. Here, we introduce the outer simplices associated with facets of $C_H(\mathcal{X}_{1-j})$ and thus extend the definition of the PE proximity maps to these outer simplices. Hence, the class covers of PE-PCDs apply for all target class points $\mathcal{X}_j$. PE-PCDs are based on the regions of simplices associated with the vertices of these simplices, called M-vertex regions. We characterize these vertex regions with barycentric coordinates of target class points with respect to the vertices of the d-simplices. However, the barycentric coordinates only apply for the target class points inside the convex hull of non-target class points $C_H(\mathcal{X}_{1-j})$. For those points outside the convex hull, we may incorporate the generalized barycentric coordinates of, for example, the coordinate system of Warren (1996). Such coordinate systems are convenient for locating points outside $C_H(\mathcal{X}_{1-j})$ since outer simplices are similar to convex d-polytopes even though they are unbounded. However, generalized barycentric coordinates of the points with respect to these convex polytopes are not unique. Hence, the associated properties of MDSs and convex distance measures are not well-defined.

We define two types of classifiers based on PE-PCDs, namely, hybrid and cover PE-PCD classifiers. We show that these classifiers are better in classifying the minority class in particular. This makes cover PE-PCD classifiers more appealing since they present slightly better performance than other classifiers (including hybrid PE-PCD classifiers) with a high reduction in the data set. In hybrid PE-PCD classifiers, alternative classifiers are used when PE-PCD pre-classifiers are unable to make a decision on a query point. These pre-classifiers are only defined by the simplices provided in the Delaunay tessellation of the set $\mathcal{X}_{1-j}$, hence only for target class points in $C_H(\mathcal{X}_{1-j})$. We considered alternative classifiers kNN, SVM and CCCD. In both our simulation studies and real data experiments, there are some cases where hybrid classifiers outperform their non-hybrid counterparts; for example, PE-SVMs outperform SVM classifiers in dimensionally reduced HTRU data set. This may be an indication that, if used alongside with proper alternative classifiers, PE-PCD classifiers could be better in modelling the decision boundary closer to the overlapping region of classes. The cover PE-PCD classifiers, on the other hand, are based on two types of covers: *composite* covers where the target class points inside and outside of the convex hull of the non-target class points are covered with separate proximity regions, and *standard* covers where all points are covered with regions based on the same family of proximity maps. For composite covers, we consider a composition of spherical proximity maps (used in CCCDs) and PE proximity maps. We observe that, in general, standard cover classifiers perform slightly better or comparable to composite cover classifiers in reduced and imbalanced data sets unless the standard cover suffers from the high dimensionality. In general, however, results on both hybrid and cover PE-PCD classifiers indicate that when the dimensionality is low and classes are imbalanced, standard cover PE-PCD classifiers achieve either comparable or slightly better classification performance than others.

PE-PCD class covers are low in complexity (with respect to the number of observations); that is, by finding the MDSs of these PE-PCDs, we can construct class covers with minimum number of proximity regions. The minimum dominating set, or the prototype set, is viewed as a reduced data set that potentially increases the testing speed of a classifier. CCCDs have the same properties, but only for data sets in $\mathbb{R}$. By extending end intervals, i.e. intervals with infinite end points, to outer simplices in $\mathbb{R}^d$ for $d > 1$, we established classifiers having the same appealing properties of CCCDs in $\mathbb{R}$. Experiments on both simulated and real data sets indicate that the expansion parameter r of the PE proximity maps substantially decreases the cardinality of the minimum dominating set, but the classification performance decreases if r is very large. Although PE-PCDs substantially reduce the number of observations of almost all of the real data sets that are considered in this work, higher values of r actually degrade the classification performance of both PE-PCD based classifiers. Hence, an optimal choice of r value is preferred. But a major drawback of PE-PCDs is the exponentially increasing complexity of the prototype set on the dimensionality of the data set d. This fact is due to the Delaunay tessellation of the non-target class since the number of simplices and facets increase exponentially in d (see Theorem 5). Therefore, these class covers become inconvenient for modelling the supports of the points from the classes in high dimensions. We employ methods of dimensionality reduction, e.g. principal components analysis, to mitigate the effects of high dimensionality. PE-PCD cover classifiers perform well in reduced dimensions only if the extracted set of principal components provide a high percentage of explained variance. For some real data sets, however, it is inefficient to rely on a few number of features, and hence, PE-PCD classifiers are outperformed by other classifiers which make use of a set of higher number of explanatory input variables.

PE-PCDs offer classifiers of (exact) minimum complexity based on estimation of the class supports. The MDSs of PE-PCDs are computationally tractable, and hence, the maximum reduction is achieved in polynomial time (on the size of the training data set). This property of PE-PCDs, however, achieved by partitioning of $\mathbb{R}^d$ by Delaunay tessellation, and as a result, the number of the simplices and facets of the convex hull of the non-target class determines the complexity of the model which increases exponentially fast with the dimensionality d of the data set, i.e. $\mathcal{O}\left(n_{1-j}^{\lceil d/2 \rceil}\right)$ for n_{1-j} being the number of non-target class points. Indeed, this leads to an overfitting of the data set. We employ PCA to extract the features with the most variation, and thus reduce the dimensions to mitigate the effects of dimensionality. PCA, however, is one of the oldest dimensionality reduction methods, and there are many dimension reduction methods in literature that may potentially increase the classification performance of PCD classifiers. One other case to be made on PE-PCD covers is that, with the assumption of strict δ-separability between two classes in a data set, PE-PCDs can be shown to be consistent and they achieve Bayes optimal performance of $L^* = 0$. However, Devroye et al. (1996) suggests that classifiers with homogeneous decision regions often lead to overfitting and they are best fit for data sets with separable class conditional distributions. The PCDs considered in this work, including CCCDs, are pure of non-target class points, hence none of the non-target class points reside in the class cover. DeVinney et al. (2002) introduced random walk CCCDs (RW-CCCDs) that are non-pure alternatives of CCCDs where some non-target class points are allowed inside of the class cover in order to mitigate the effects of overfitting. Although PE-PCDs offer class covers with computationally tractable minimum prototype set, PE-PCDs are also pure class covers, and hence, they also build homogeneous regions for decision making. We believe that, as a follow-up to this

work, it is worthwhile to define non-pure and relaxed PCDs that constitute both exact minimum dominating sets and heterogeneous class covers. Such directed graphs would enjoy appealing properties of PCDs and they would also be the building blocks of classifiers that are consistent for a general set of real life data sets.

Although, our work proves the idea that relatively good performing classifiers with minimum prototype sets can be provided with PCDs, a discussion raises if there exist PCDs with alternative partitioning methods whose exact minimum dominating sets are fixed-parameter tractable with respect to d, unlike PCDs based on Delaunay tessellations. A problem is said to be *fixed-parameter tractable* (FPT), if there exists an algorithm to solve the problem, running in $f(k)|x|^c$ time where $c \in \mathbb{R}^+$ is a constant, f is an arbitrary computable and non-decreasing function of the *parameter k*, and $|x|$ is the size of the input x (Downey and Fellows 2013). It is often appealing to try to find an FPT algorithm for a problem initially shown to be solvable in $O(n^{f(k)})$ time (which is not FPT). Note that, as in Theorem 5, it only takes at most $\mathcal{O}(2^d)$ or $\mathcal{O}(1)$ time to find the exact extremum points of each d-simplex $\mathcal{S}$. Hence, a possible line of research for PCDs could be to employ alternative partitioning methods such that $\mathbb{R}^d$ is partitioned in at most $\mathcal{O}(n_{1-j}^c)$ time and the extremum points are found in $\mathcal{O}(2^k)$ with parameter $k = d$. We believe such a partitioning method, say for example a rectangular partitioning scheme with polynomial running time on both n and d, that produces less partitioning than a Delaunay tessellation could be more appealing for the class cover. Classifiers based on such PCDs and their classification performance are topics of ongoing research.

Appendix

Proof of Theorem 1

We prove this theorem by induction on dimension d. The proof of the case $d = 1$ is trivial. For $\mathcal{S}(\mathcal{Y}) = (y_1, y_2) \subset \mathbb{R}$ and $y_1 < y_2$, the vertex regions $R_M(y_1)$ and $R_M(y_2)$ are the intervals (y_1, M) and (M, y_2), respectively ($\{x = M\}$ and $\{x = y_i\}$ for $i = 1, 2$ have zero $\mathbb{R}$-Lebesgue measure). For $\alpha_1 \in (0, 1)$ and $\alpha_2 = 1 - \alpha_1$, let $\alpha_1 y_1 + \alpha_2 y_2$ be the convex (or barycentric) combination of $x \in \mathcal{S}(\mathcal{Y})$. Hence, for $m_1 y_1 + m_2 y_2$ being the convex combination of M, we have $x \in (y_1, M) = R_M(y_1)$ if and only if $\alpha_1/\alpha_2 > m_1/m_2$. The case $d = 2$ is proved in Proposition 1. Thus, there only remains the case $d > 2$. Suppose the statement is true for all faces of the d-simplices which are $d - 1$ dimensional, and by that, we will show that the statement is also true for the d-simplex which is d dimensional.

It is sufficient to show the result for y_1, as the others follow by symmetry. Let $x \in R_M(y_1)$ and note that the elements of the set of $(d - 1)$-faces, $\{f_j\}_{j=2}^{d+1}$, are adjacent to y_1. Each of these faces is of $d - 1$ dimensions. Hence, they are $(d - 1)$-simplices and they also have their own vertex regions. Thus, let $R_{M_i}(y_j, f_i)$ be the vertex region of y_j with respect to $(d - 1)$-simplex f_i for $j \neq i$. Note that M_i is the center of f_i. Now, let $w_{f_i}(z, y_j) = w_{ij}$ be the barycentric coordinate of point z corresponding to y_j with respect to the f_i. Observe that w_{ii} is not defined since y_i is not a vertex of the face f_i.

Moreover, let $\mathbf{m}' = (m'_1, \ldots, m'_{i-1}, m'_{i+1}, \ldots, m'_{d+1})$ be the vector of barycentric coordinates of M_i with respect to f_i, and note that M_i is a linear combination of M and $\mathbf{y}_i$. Also, observe that m'_i is not defined since the vertex $\mathbf{y}_i$ is not a vertex of f_i. Hence, for $\beta \geq 1$,

$$M_i = \beta M + (1 - \beta)\mathbf{y}_i = \beta\left(\sum_{t=1; t \neq i}^{d+1} m_t \mathbf{y}_t\right) + (1 - \beta)\mathbf{y}_i. \tag{14}$$

Therefore, by the uniqueness of barycentric coordinates, we have $m'_t = \beta m_t$ for $t = 1, \ldots, d+1$ and $t \neq i$. Note that $(1 - \beta) = 0$ since $M_i \in f_i$ and also $f_i \subset \partial(\mathcal{S}(\mathcal{Y}))$. Hence, $\beta = 1$ which implies $m'_t = m_t$ for all $t \neq i$. Then, $m'_1/m'_j = m_1/m_j$ for $j = 2, 3, \ldots, d+1$ and $j \neq i$. We use this result on our induction hypothesis.

Now, for $i = 2, \ldots, d+1$, let the face f_i and line defined by x and $\mathbf{y}_i$ cross at the point z_i Observe that $z_i \in f_i$, and since f_i is a $(d-1)$-simplex and $x \in R_M(\mathbf{y}_1)$, see that $z_i \in R_{M_i}(\mathbf{y}_1, f_i)$. By induction hypothesis and Equation (14), we observe that $z_i \in R_{M_i}(\mathbf{y}_1, f_i)$ if and only if $w_{i1} > (m'_1/m'_j)w_{ij}$ if and only if $w_{i1} > (m_1/m_j)w_{ij}$ for $j = 2, 3, \ldots, d+1$ and $j \neq i$. Since the point x is the convex (and linear) combination of z_i and $\mathbf{y}_i$, for $\alpha \in (0, 1)$, we have

$$x = (1 - \alpha)\mathbf{y}_i + \alpha z_i = (1 - \alpha)\mathbf{y}_i + \alpha\left(\sum_{k=1; k \neq i}^{d+1} w_{ik}\right).$$

By the uniqueness property of barycentric coordinates, it follows that $w_{\mathcal{S}}^{(1)}(x) = \alpha w_{i1}$ and $w_{\mathcal{S}}^{(j)}(x) = \alpha w_{ij}$. Hence,

$$\frac{w_{\mathcal{S}}^{(1)}(x)}{w_{\mathcal{S}}^{(j)}(x)} = \frac{w_{i1}}{w_{ij}} > \frac{m_1}{m_j}. \tag{15}$$

Since Eq. (15) is true for all $i = 2, \ldots, d+1$, we see that $x \in R_M(\mathbf{y}_1)$ if and only if $w_{\mathcal{S}}^{(1)}(x) > (m_1/m_i)w_{\mathcal{S}}^{(i)}(x)$. Hence, the desired result follows. $\qquad\square$

References

Akbani, R., Kwek, S., & Japkowicz, N. (2004). Applying support vector machines to imbalanced datasets. In *Proceedings of 15th European conference on machine learning, Pisa* (pp. 39–50).

Alcalá, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., et al. (2011). Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Multiple-Valued Logic and Soft Computing*, *17*(2–3), 255–287.

Alpaydın, E. (1999). Combined 5×2 CV F test for comparing supervised classification learning algorithms. *Neural Computation*, *11*(8), 1885–1892.

Angiulli, F. (2012). Prototype-based domain description for one-class classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(6), 1131–1144.

Arora, S., & Lund, C. (1996). *Approximation algorithms for NP-hard problems*. Boston, MA: PWS Publishing.

Bereg, S., Cabello, S., Díaz-Báñez, J. M., Pérez-Lantero, P., Seara, C., & Ventura, I. (2012). The class cover problem with boxes. *Computational Geometry*, *45*(7), 294–304.

Bien, J., & Tibshirani, R. (2011). Prototype selection for interpretable classification. *The Annals of Applied Statistics*, *5*(4), 2403–2424.

Cannon, A. H., & Cowen, L. J. (2004). Approximation algorithms for the class cover problem. *Annals of Mathematics and Artificial Intelligence*, *40*(3–4), 215–223.

Ceyhan, E. (2005). An investigation of proximity catch digraphs in delaunay tessellations. PhD thesis, Johns Hopkins University, Baltimore, MD.

Ceyhan, E. (2010). Extension of one-dimensional proximity regions to higher dimensions. *Computational Geometry*, *43*(9), 721–748.

Ceyhan, E., & Priebe, C. E. (2005). The use of domination number of a random proximity catch digraph for testing spatial patterns of segregation and association. *Statistics & Probability Letters*, *73*(1), 37–50.

Ceyhan, E., Priebe, C. E., & Marchette, D. J. (2007). A new family of random graphs for testing spatial segregation. *Canadian Journal of Statistics*, *35*(1), 27–50.

Ceyhan, E., Priebe, C. E., & Wierman, J. C. (2006). Relative density of the random *r*-factor proximity catch digraph for testing spatial patterns of segregation and association. *Computational Statistics & Data Analysis*, *50*(8), 1925–1964.

Chvatal, V. (1979). A greedy heuristic for the set-covering problem. *Mathematics of Operations Research*, *4*, 233–235.

Deng, X., & Zhu, B. (1999). A randomized algorithm for the Voronoi diagram of line segments on coarse-grained multiprocessors. *Algorithmica*, *24*(3–4), 270–286.

DeVinney, J., Priebe, C., Marchette, D., & Socolinsky, D. (2002). Random walks and catch digraphs in classification. In *Proceedings of the 34th symposium on the interface, volume 34: Computing science and statistics, Montreal, Quebec*.

DeVinney, J. G. (2003). The class cover problem and its application in pattern recognition. PhD thesis, Johns Hopkins University, Baltimore, MD.

Devroye, L., Gyorfi, L., & Lugosi, G. (1996). *A probabilistic theory of pattern recognition*. New York: Springer.

Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, *10*(7), 1895–1923.

Downey, R. G., & Fellows, M. R. (2013). *Fundamentals of parameterized complexity* (Vol. 4). New York: Springer.

Dua, D., & Graff, C. (2019). UCI machine learning repository. Irvine, CA: University of California, School of Information and Computer Science. http://archive.ics.uci.edu/ml. Accessed 10 July 2019.

Eveland, C. K., Socolinsky, D. A., Priebe, C. E., & Marchette, D. J. (2005). A hierarchical methodology for class detection problems with skewed priors. *Journal of Classification*, *22*(1), 17–48.

Frey, P. W., & Slate, D. J. (1991). Letter recognition using holland-style adaptive classifiers. *Machine Learning*, *6*(2), 161–182.

Gao, B. J., Ester, M., Xiong, H., Cai, J. Y., & Schulte, O. (2013). The minimum consistent subset cover problem: A minimization view of data mining. *IEEE Transactions on Knowledge and Data Engineering*, *25*(3), 690–703.

Hammer, P., Liu, Y., Simeone, B., & Szedmák, S. (2004). Saturated systems of homogeneous boxes and the logical analysis of numerical data. *Discrete Applied Mathematics*, *144*(1–2), 103–109.

Jameson, A., Possenti, A., Stappers, B. W., Levin, L., Bailes, M., Burgay, M., et al. (2010). The high time resolution universe pulsar survey ? I. system configuration and initial discoveries. *Monthly Notices of the Royal Astronomical Society*, *409*(2), 619–627.

Jaromczyk, J. W., & Toussaint, G. T. (1992). Relative neighborhood graphs and their relatives. *Proceedings of the IEEE*, *80*(9), 1502–1517.

Karr, A. F. (1992). *Probability* (1st ed.). New York: Springer.

Lawson, C. L. (1986). Properties of n-dimensional triangulations. *Computer Aided Geometric Design*, *3*(4), 231–246.

Lyon, R. J., Stappers, B., Cooper, S., Brooke, J., & Knowles, J. (2016). Fifty years of pulsar candidate selection: from simple filters to a new principled real-time classification approach. *Monthly Notices of the Royal Astronomical Society*, *459*(1), 1104–1123.

Manukyan, A., & Ceyhan, E. (2016). Classification of imbalanced data with a geometric digraph family. *Journal of Machine Learning Research*, *17*(189), 1–40.

Marchette, D. J. (2004). *Random graphs for statistical pattern recognition*. Hoboken: Wiley.

Mehta, M., Rissanen, J., & Agrawal, R. (1995). Mdl-based decision tree pruning. In *Knowledge discovery and data mining* (pp. 216–221).

Morello, V., Barr, E., Bailes, M., Flynn, C., Keane, E., & van Straten, W. (2014). Spinn: A straightforward machine learning solution to the pulsar candidate selection problem. *Monthly Notices of the Royal Astronomical Society*, *443*(2), 1651–1662.

Narasimhan, H., Pan, W., Kar, P., Protopapas, P., & Ramaswamy, H. G. (2016). Optimizing the multiclass f-measure via biconcave programming. In *2016 IEEE 16th international conference on data mining (ICDM)* (pp. 1101–1106). IEEE.

Parekh, A. K. (1991). Analysis of a greedy heuristic for finding small dominating sets in graphs. *Information Processing Letters*, *39*, 237–240.

Pękalska, E., Duin, R. P., & Paclík, P. (2006). Prototype selection for dissimilarity-based classifiers. *Pattern Recognition*, *39*(2), 189–208.

Priebe, C. E., DeVinney, J. G., & Marchette, D. J. (2001). On the distribution of the domination number for random class cover catch digraphs. *Statistics & Probability Letters*, *55*(3), 239–246.

Priebe, C. E., Marchette, D. J., DeVinney, J., & Socolinsky, D. (2003a). Classification using class cover catch digraphs. *Journal of Classification*, *20*(1), 3–23.

Priebe, C. E., Solka, J. L., Marchette, D. J., & Clark, B. T. (2003b). Class cover catch digraphs for latent class discovery in gene expression monitoring by DNA microarrays. *Computational Statistics & Data Analysis*, *43*(4), 621–632.

Rissanen, J. (1989). *Stochastic complexity in statistical inquiry theory*. River Edge: World Scientific Publishing Co. Inc.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, *13*(7), 1443–1471.

Seidel, R. (1995). The upper bound theorem for polytopes: An easy proof of its asymptotic version. *Computational Geometry*, *5*(2), 115–116.

Serafini, P. (2014). Classifying negative and positive points by optimal box clustering. *Discrete Applied Mathematics*, *165*, 270–282.

Takigawa, I., Kudo, M., & Nakamura, A. (2009). Convex sets as prototypes for classifying patterns. *Engineering Applications of Artificial Intelligence*, *22*(1), 101–108.

Toussaint, G. T. (1980). The relative neighborhood graph of a finite planar set. *Pattern Recognition*, *12*(4), 261–268.

Toussaint, G. T. (2002). Proximity graphs for nearest neighbor decision rules: Recent progress. In *Proceedings of the 34th symposium on the interface, Montreal, Quebec* (Vol. 34).

Ungar, A. A. (2010). *Barycentric calculus in euclidean and hyperbolic geometry: A comparative introduction*. Singapore: World Scientific Publishing Co. Pte. Ltd.

Vazirani, V. V. (2001). *Approximation algorithms*. New York: Springer.

Wang, W., Xu, Z., Lu, W., & Zhang, X. (2003). Determination of the spread parameter in the Gaussian kernel for classification and regression. *Neurocomputing*, *55*(3–4), 643–663.

Warren, J. (1996). Barycentric coordinates for convex polytopes. *Advances in Computational Mathematics*, *6*(1), 97–108.

Watson, D. F. (1981). Computing the *n*-dimensional Delaunay tessellation with application to Voronoi polytopes. *The Computer Journal*, *24*(2), 167–172.

West, D. B. (2000). *Introduction to graph theory* (2nd ed.). Englewood Cliffs: Prentice Hall.

Woźniak, M., Graña, M., & Corchado, E. (2014). A survey of multiple classifier systems as hybrid systems. *Information Fusion*, *16*, 3–17.