Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v11n2p447

A note on depth-based classification of circular data

By Pandolfo, D'Ambrosio, Porzio

Published: 14 October 2018

Electronic Journal of Applied Statistical Analysis
Vol. 11, Issue 02, October 2018, 447-462
DOI: 10.1285/i20705948v11n2p447

A note on depth-based classification of circular data

Giuseppe Pandolfo*[a], Antonio D'Ambrosio[b], and Giovanni C. Porzio[c]

[a] *University of Naples Federico II, Department of Industrial Engineering, Napoli, Italy*
[b] *University of Naples Federico II, Department of Economics and Statistics, Napoli, Italy*
[c] *University of Cassino and Southern Lazio, Department of Economics and Law, Cassino, Italy*

Published: 14 October 2018

A procedure is developed in order to deal with the classification problem of objects in circular statistics. It is fully nonparametric and based on depth functions for directional data. Using the so-called DD-plot, we apply the k-nearest neighbors method in order to discriminate between competing groups. Three different notions of data depth for directional data are considered: the *angular simplicial*, the *angular Tukey* and the *arc distance*. We investigate and compare their performances through the average accuracy rate by means of simulated and real data sets.

keywords: Angular depths, Supervised circular classification, k-NN, Accuracy rate.

1 Introduction

Circular (or angular) data arise when observations are measured as directions or angles. They play an important role in many fields such as biology, meteorology and physics; a well-known example in biology is about the migration of birds.

Within the literature, circular data are presented and discussed in books by Mardia and Jupp (1972) and Batschelet (1981). They provide a wide survey about specific features and problems we have to face when dealing with them. As a matter of fact, most of the standard methods used "on the line" are not appropriate, and often misleading,

*Corresponding author: giuseppe.pandolfo@unina.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

for observations on the circle. For instance, the standard arithmetic mean fails in such spaces. Thus, specific tools are required, and this is true also for classification.

In this work, the focus is on supervised classification (also called discriminant analysis), where the aim is to assign observations to classes (or groups), given a previous knowledge on their structures obtained through some preliminary observed data.

Note that a single cut-off value, which defines two half-lines, may be enough to discriminate between two different distributions on the line. By contrast, to discriminate on a circle we need to identify at least two angles, which in turns define two different arcs (one for each distribution). This is necessary as no natural ordering exists for this kind of data.

To illustrate how discriminant analysis on a circle works, consider data from a two-armed neurological pilot-study. They are depicted in Figure 1. Two groups of 9 and 10 patients are examined with respect to their orientation abilities, and two different diagnosis are considered. Sample values (measured in degrees) for the two groups are $55°, 77°, 77°, 82°, 86°, 95°, 110°, 130°, 171°$, and $10°, 29°, 52°, 52°, 58°, 65°, 90°, 285°, 325°, 355°$, respectively (Ackermann, 1997). The aim is to identify two sectors so that a new patient can be assigned to one of the groups given his/her orientation ability. In practice, according to some proper rule, two splitting angles must be found.

Group 1	Group 2
55°	10°
77°	29°
77°	52°
82°	52°
86°	58°
95°	65°
110°	90°
130°	285°
171°	325°
-	355°

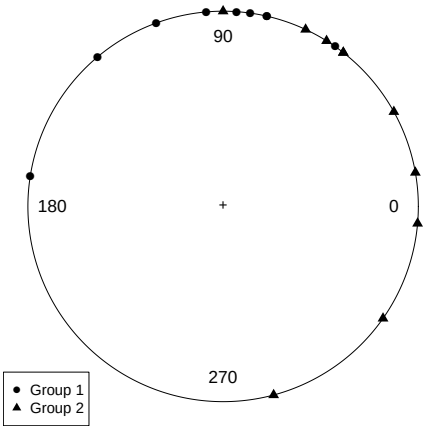

Figure 1: Sample values (degrees) of a two-armed neurological pilot-study two groups of N = 9 + 10 = 19 patients associated with two different diagnosis examined with respect to their abilities of orientation (Ackermann, 1997) (left panel), and the corresponding circular plot (right panel)

To perform circular data classification, many parametric tools are available (see e.g. Morris and Laycock, 1974, El Khattabi and Streit, 1996 and Figueiredo, 2009), with applications in sign language, astronomy, medicine and biology (see e.g. Sengupta and Ugwuowo, 2011 and the list in López-Cruz et al., 2015).

In spite of the interest on this topic, the literature on nonparametric discriminant

analysis for circular data is very sparse. Liu and Singh (1992) discussed methods for ordering directional data and suggested a possible solution for discriminant analysis. That is, an observation can be classified according to its center-outward ordering with respect to two given competing distributions. Later, the optimal rule for data on the line given by Stoller (1954) was adapted to circular data by Ackermann (1997). This method, which looks for the pair of splitting angles which maximizes the probability of correct classification, is asymptotically unbiased and consistent. Unfortunately, this optimal solution requires solving an NP-complex problem, and hence it is quite computationally infeasible.

In brief, given that Liu and Singh suggestion has not been developed, and that Ackermann's proposal is infeasible in practice, this work is aimed at analyzing and comparing new methods for nonparametric supervised classification on circles (and possibly on spheres). More specifically, in line with a literature for data in the d-dimensional Euclidean space $\mathbb{R}^d$, a classifier based on data depth is considered.

The leading idea is to evaluate the depth, i.e. the degree of centrality of a new observation with respect to the previously collected samples (one for each group), and then using its value as a basis for a classification rule. Different notions of angular depth functions and classifications rules are investigated in order to provide suggestions on their proper use under different conditions.

The paper is organized as follows. In Section 2, we provide a brief review of data depth and depth-based classifiers for data in $\mathbb{R}^d$. In Section 3, data depth functions for directional data are introduced. In Section 4, simulated data are used to study the performance of the classifier, and then a real data example is presented. Finally, Section 5 contains some concluding remarks.

2 Nonparametric classification through data depth

A quite large number of tools are available for classification purposes in standard multivariate analysis. Assuming normality, linear discriminant analysis is probably the most used. Other examples of widely used methods are: quadratic discriminant analysis, k-nearest neighbors (k-NN), support vector machines, and decision trees. For an overview see e.g. Hastie et al. (2009). Many others methods, both parametric and nonparametric, are available as well.

Nonparametric classifiers have the important advantage to be more flexible, given that they do not rely on distributional assumptions and are capable of fitting a large number of functional forms. Amongst them, this work focuses on classifiers based on data depth. Considering their full nonparametric nature, they can be used in many different contests. For instance, apart from data in $\mathbb{R}^d$, they have been also exploited for functional data (López-Pintado and Romo, 2006).

In the next sections, first a very short introduction to data depth is provided. Then, how these functions can be used for classification purposes is reviewed.

2.1 Data depth in $\mathbb{R}^d$

The depth of a point $x \in \mathbb{R}^d$ (for $d \geq 1$) is a function that measures the "centrality" or "deepness" of it with respect to a multivariate distribution F or a multivariate data cloud, and it is denoted by $D(x, F)$. A general definition of statistical depth function was introduced by Zuo and Serfling (2000). After them, a statistical depth function is any non-negative and bounded function that satisfies the following four desirable properties:

P1. affine invariance;

P2. maximality at center;

P3. monotonicity relative to the deepest point;

P4. vanishing at infinity.

Different notions of depth function are available within the literature, and reviews are for instance in Liu et al. (1999), Zuo and Serfling (2000), Serfling (2006), and more recently in Liu and Modarres (2011).

Amongst them, the *halfspace depth*, the *simplicial depth*, the *Mahalanobis depth*, and the L^p *depth* are briefly presented here. Different authors suggested their use for classification purposes for $\mathbb{R}^d$ data. They are defined as follows.

***Definition* 1 (Halfspace Depth)** *The halfspace depth (HD) (Tukey, 1975) of a point $x \in \mathbb{R}^d$ with respect to a distribution F is defined as the minimum probability measure of any closed halfspace H containing x.*

$$HD(x, F) := \inf_{H} \left\{ P_F(H) : H \text{ is any closed halfspace in } \mathbb{R}^d, \text{ and } x \in H \right\}.$$

Given a sample, the empirical version of the halfspace depth of a point x is the smallest proportion of data points contained in any closed halfspace having the boundary that passes through x.

***Definition* 2 (Simplicial Depth)** *The simplicial depth (SD) (Liu, 1990) of a point $x \in \mathbb{R}^d$ with respect to a distribution F is defined as*

$$SD(x, F) := P_F\{x \in S[X_1, \ldots, X_{d+1}]\},$$

where $S[X_1, X_2, \ldots, X_{d+1}]$ is a closed simplex in $\mathbb{R}^d$ formed by $(d+1)$ random observations from F.

Given a sample, the empirical version of the simplicial depth of a point x is the proportion of simplices (made up of observations from the sample) that includes x.

***Definition* 3 (Mahalanobis Depth)** *The Mahalanobis depth (MHD) of a point $x \in \mathbb{R}^d$ with respect to a distribution F is defined as*

$$MHD(x, F) := \left[1 + (x - \mu_F)' \Sigma_F^{-1} (x - \mu_F)\right]^{-1},$$

where μ_F and Σ_F^{-1} are the mean vector and the covariance matrix of F, respectively.

The empirical version of the Mahalanobis depth is simply obtained by replacing μ_F and Σ_F with their sample estimates.

***Definition* 4 (L^p Depth)** *The L^p depth (L^pD) (Zuo and Serfling, 2000) of a point $x \in \mathbb{R}^d$ with respect to a distribution F is defined as*

$$L^pD\left(x, F\right) := \left(1 + E\left\|x - X\right\|_p\right)^{-1},$$

where $\|\cdot\|$ is the L^p norm.

Given a sample, the empirical version of the L^p depth of a point x uses the average of the p-norm distance of it from the sample observations.

From a computational perspective, the Mahalanobis and the L^p depths should be preferred. However, the Mahalanobis depth works properly only when distributions are elliptically contoured (for more details see Zuo and Serfling, 2000). On the other hand, unlike the others, the L^p depth is not affine invariant, but an affine invariant version of the L^2 depth can be obtained. The maximality at center and monotonicity relative to the deepest point properties are satisfied by all of them. Moreover, they all vanish at infinity.

2.2 Depth-based classifiers

After the first proposal of Liu and Singh (1992), the use of data depth to perform supervised classification has been suggested and investigated by many authors. Two main approaches have been adopted in the literature: the *maximum depth classifier* and the *Depth vs. Depth* (*DD*) classifier.

The first computes the depth function of a new observation x with respect to all the competing distributions (or groups) $F_1, \ldots, F_m$, and then assign x to the distribution (or group) with respect to which it attains the highest depth value, that is:

$$D\left(x, F_i\right) > D\left(x, F_j\right) \quad i \neq j, \quad \Rightarrow \text{assign x to } F_i,$$

where $D\left(x, F_i\right)$ and $D\left(x, F_j\right)$ are the empirical depths of x w.r.t. the ith and the jth distribution, respectively.

This classifier is asymptotically optimal (i.e., its misclassification rate is minimized) when populations differ only in location, under the assumption that prior probabilities of the possible classes are equal.

This method was used first by Jörnsten (2004) and later also by Hartikainen and Oja (2006) and Mosler and Hoberg (2006). A modified version for unequal prior probabilities which is asymptotically optimal when the underlying population distributions are elliptical and the density functions are strictly decreasing about the center was proposed by Ghosh and Chaudhuri (2005). However, this technique works well only when the halfspace depth function is adopted. Moreover, this approach requires complicated estimation techniques. A maximum depth classifier for skewed unimodal distributions was

proposed by Hubert and Van der Veeken (2010). On the other hand, Dutta and Ghosh (2011) considered a maximum depth classifier based on the L^p depth.

Recently, Li et al. (2012) proposed the so called *DD*-classifier that generalizes the concept of maximum depth classifier by using polynomial separating functions. This method is based on the *DD*-plot (Depth vs Depth plot), introduced by Liu et al. (1999), which is a graphical representation that compares two multivariate distributions or samples based on data depth.

It is a two-dimensional scatterplot where each data point in $\mathbb{R}^d$ is represented with coordinates evaluated with respect to two distributions. The depth of each data point in $\mathbb{R}^d$ is computed with respect to two multivariate distributions. Then, these depth values are used as coordinates in a scatterplot. If the two empirical distributions are quite separated in the sample space, then the corresponding points are separated in the *DD*-plot as well. For this reason, a classification rule can be directly applied in this latter. Accordingly, the following classification rule can be adopted:

$$\begin{cases} D(x, G) > r(D(x, F)) & \Rightarrow \text{assign x to G} \\ D(x, G) \leq r(D(x, F)) & \Rightarrow \text{assign x to F} \end{cases}$$

where F and G are two competing distributions, and $r(\cdot)$ is a real increasing function. As separating function $r(\cdot)$, Li et al. (2012) suggested to look for a polynomial separator. This is chosen in order to minimize the empirical misclassification error rate on the training sample. Note that when $r(\cdot)$ is the identity function, the corresponding classification rule turns out to be the maximum depth classifier.

The max depth classifier is an intuitive method, easy to be implemented. In addition, it is able to deal with classification problems with a large number of groups. On the other hand, the *DD*-classifier is more flexible than the maximum depth classifier, but it requires to find the degree of the polynomial function for which the misclassification rate is minimized.

A recent approach is given by Paindaveine et al. (2015) who proposed a class of depth-based classifiers of a k-nearest neighbors nature. However, this method is not further considered here since not yet extended to circular data.

Finally, an important aspect to be considered regards the choice of the depth function. Indeed, from a classification perspective, it must be noted that for points not belonging to the convex hull of the support of distribution, the halfspace and simplicial depths assign zero depth value. This implies that sample points lying outside the convex hull of the training set have zero empirical depth, thus it is not possible to assign an observation to one of the competing groups. On the contrary, this does not occur by adopting the Mahalanobis and L^p depths since they are always positive for any data point in the sample space.

3 Classifying circular objects through depth functions

In this section, data depth functions for directional data are briefly introduced. Then, a notion of circular *DD*-classifier is provided.

3.1 Angular data depths

Three notions of angular data depth are essentially available within the literature: the *angular simplicial depth*, the *angular Tukey depth* and the *arc distance depth*. All these were introduced by Liu and Singh (1992). The *angular simplicial* and the *angular Tukey* depth functions extend the notion of simplicial and halfspace depths from $\mathbb{R}^d$ to distributions on circles or on spheres. The *arc distance depth* is specific for data on circles or spheres and can be considered the analog of the L^p depth for directional data. Recently, Pandolfo et al. (2017) introduced a class of depth functions for directional data based on angular distance between points on circles or hyperspheres.

A brief description and the definition of these depth functions are listed below. For the sake of simplicity, we discuss only the case on the circle. The same notions are available for hyperspheres in any dimension. Further discussion on depth functions for circular data is available in Agostinelli and Romanazzi (2013).

When we are dealing with data on circular spaces, simplices are replaced with the "shortest curve" that joins the pairs of data points on the circle. Hence, taking two points θ_1 and θ_2 on a circle, the shortest curve joining θ_1 and θ_2 is the $arc\,(\theta_1, \theta_2)$.

***Definition* 5 (Angular simplicial depth)** *The angular simplicial depth is defined as follows:*

$$ASD\,(\theta, H) := P_H\left\{\theta \in arc\,(W_1, W_2)\right\},$$

where θ is any points on the circle and H is the underlying distribution on the circle. W_1 and W_2 are i.i.d. observations of H.

The empirical version of the *angular simplicial depth* is the proportion of all $\binom{n}{2}$ sample arcs including θ.

On the circle, closed halfspaces are replaced with closed semicircles.

***Definition* 6 (Angular Tukey depth)** *Following Small (1987), the angular Tukey depth for a circular distribution H is defined as follows:*

$$ATD\,(\theta, H) := \inf_{\{S:\theta\in S\}}\left\{P_H\,(S)\right\},$$

that is the infimum taken over the set of all closed semicircles S containing θ in their boundaries or in their interiors.

The empirical version of the *angular Tukey depth* of the point θ is the smallest proportion of data points contained in any semicircle including θ.

The *arc distance depth* is computed on a Riemannian manifold.

***Definition* 7 (Arc distance depth)** *The arc distance depth function of a point θ on the circle S_d is defined as follows:*

$$ADD(\theta, H) := \pi - \int l(\theta, \varphi)\, dH(\varphi),$$

where $l(\theta, \varphi)$ is the Riemannian distance between θ and φ. This distance is the length of the shortest arc that joins θ and φ on the great circle determined by θ and φ.

The empirical version of *arc distance depth* is the mean distance of the point θ from all the other observations on the circle.

All these depth functions are invariant under rotation, are maximized at center, and possess monotonicity relative to the deepest point for unimodal symmetric distributions. The angular Tukey depth requires high computational costs and its empirical version is constant and assumes its minimum on the semi circle opposed to that of the deepest point. The angular simplicial depth is less computationally intensive but is constant on empty arcs. The arc distance depth is computationally feasible and it assumes positive values for points all-over the circle.

3.2 Circular *DD*-classifier

To illustrate how circular depth based classifier may work, we depicted *DD*-plots for circular Ackermann's data introduced above (Figure 2). The three plots are given by the three depth functions just described (angular simplicial, angular Tukey and arc distance depths; upper-left, upper-right, and center-bottom pictures, respectively).

The aim of a *DD*-classifier is to define a partition of such a *DD*-space in order to separate the two groups of observations at their best. That is, a proper depth based classifier should be able to discriminate between the two groups (denoted by triangles and circles) in the pictures.

The max depth classifier correspond to a partition of such space defined according to the 45° line (the solid line in the plots).

According to the discussion in Section 2, a polynomial separating function of order s can be adopted. However, this scheme is really time-consuming. Indeed, the number of different polynomials of order s that can serve as a classification rule is $\binom{N}{s}$, where N is the sample size. This is the number of possible ways to select s points from N, and each of the selections has an associated order s polynomial which interpolates between these s points and the origin $(0,0)'$ of the *DD*-plot. Consequently, as N increases, the complexity of the estimation grows at the rate N^s. For this reason, alternatively, we suggest to consider a k-NN approach.

This nonparametric method is used for classification (and also for regression), and it classifies a given observation θ_0 to the class with the highest estimated probability considering the k points in the training set that are closest to θ_0. Then, it estimates the conditional probability for class i as the fraction of k points in the neighborhood of θ_0.

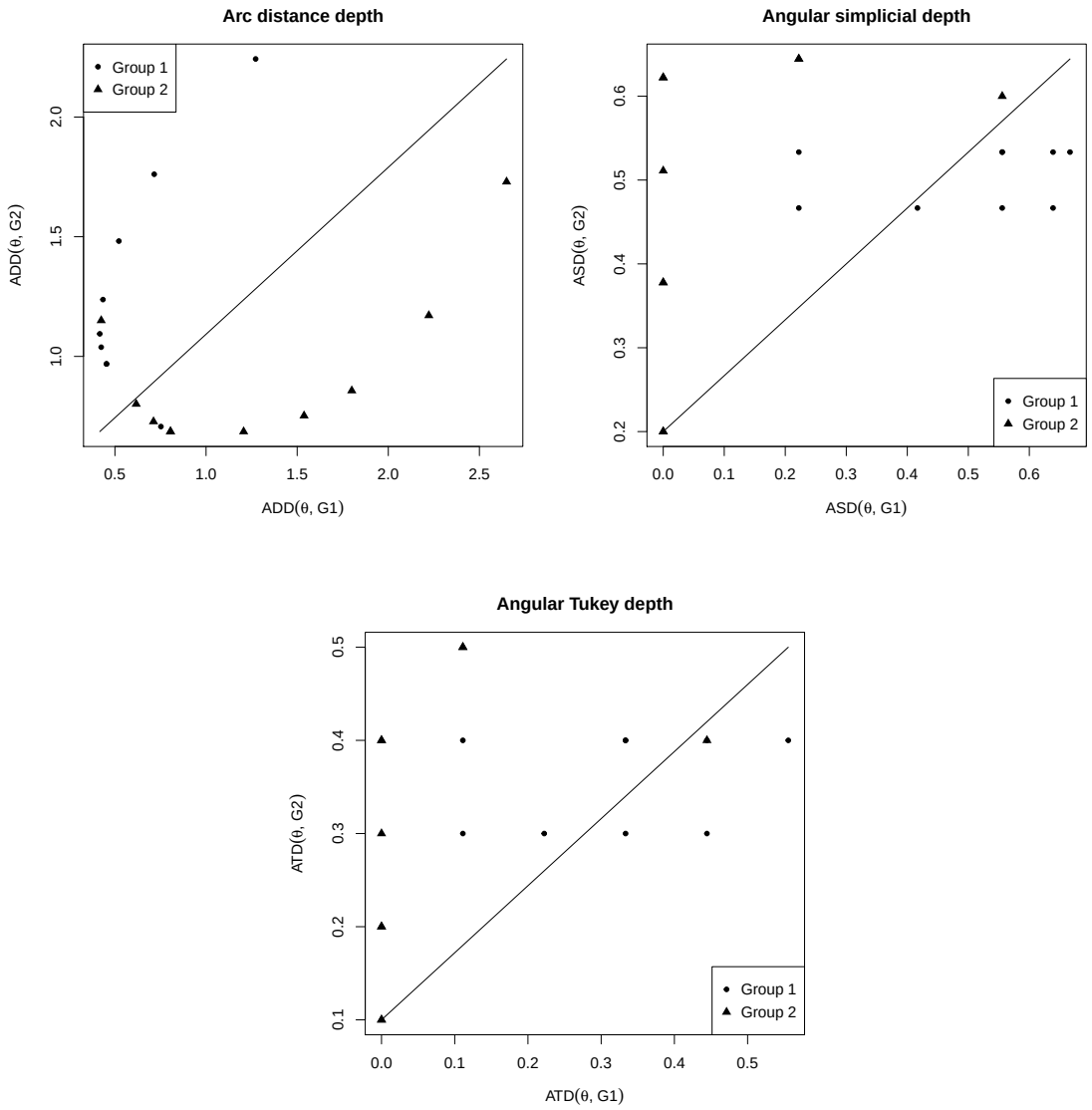

Figure 2: *DD*-plots of Ackermann's data using the Arc distance, the angular simplicial and the angular Tukey depth functions.

Thus, an object is classified by the majority vote of its neighbors, that is, to the class most common among its k nearest neighbors.

4 Simulation study

A simulation study was ran in order to evaluate the performance of the *DD*-classifier for circular data and comparing results obtained through the three data depth functions

described in section 3. Simulations have been run with respect to the reference distribution for circular data, that is, the von Mises distribution. The probability density function of the von Mises distribution is defined as follows:

$$F(\theta; \mu, c) = \frac{1}{2\pi I_0(\nu)} e^{\nu \cos(\theta - \mu)}, \qquad 0 \leq \theta < 2\pi,$$

where the mean direction $0 \leq \mu < 2\pi$ and concentration $\nu \geq 0$ are the parameters, and $I_0(c)$ in the normalizing constant is the modified Bessel function of the first kind and order zero.

To illustrate how the concentration parameter ν works, the circular density plots of von Mises distributions for four different concentration parameters (0, 3, 5 and 10) are displayed in Figure 3. The density lines are obtained through a kernel density estimation. As one can see, for $\nu = 0$ the distribution on the circle is uniform and when ν increases the distribution becomes more and more concentrated about the mean direction μ.

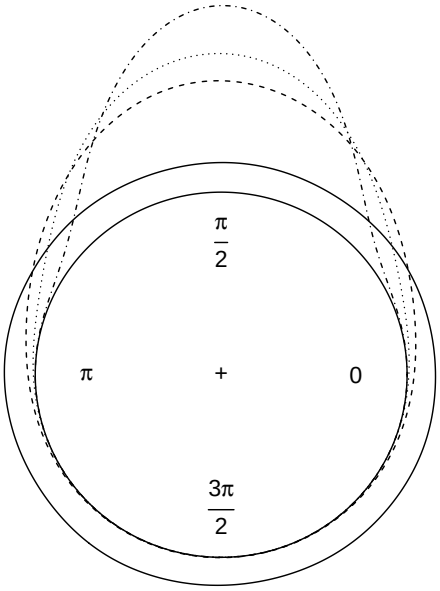

Figure 3: The kernel density functions of several von Mises distributions with mean direction, $\mu = \frac{\pi}{2}$, and various concentration parameters: 0 (solid line), 3 (dashed line), 5 (dotted line) and 10 (dot-dashed line).

In our analysis we consider discrimination between two groups. The observations are generated from von Mises distributions with different location parameters $\mu_1 \neq \mu_2$ and equal concentration parameters $\nu = \nu_1 = \nu_2$. Partially following Figueiredo (2009), three different values of the concentration parameter ν are considered, that is, $\nu \in \{3, 5, 10\}$. In addition, different distances between the mean directions of the two groups

are considered, that is $d(\mu_1, \mu_2) = \pi/6, \pi/3, \pi/2$ and $2/3\pi$.

The analysis is limited to equal prior probabilities cases (50% for each group). For each simulation setting, $N = 2000$ observations are generated (1000 for each group), and the 75% of the data are assigned to the training set and the 25% to the test set. For each combination of parameters, we conducted 100 independent replications. The same simulated data have been analyzed by using a *DD*-plot classifier approach considering the three notions of data depth functions for circular data presented in section 3, that is, the angular simplicial, the angular Tukey and the arc distance (hereafter, for convenience's sake these depth functions are simply called simplicial, Tukey and Arc distance), and the Naive Bayes classifier for von Mises distributions (vMNB) classifier introduced first by Figueiredo (2009) and later further investigated by López-Cruz et al. (2015). This latter, with a binary class and one predictive angular variable reduces to a linear classifier in $\mathbb{R}^2$. Specifically, in the case of equal concentration parameters and different mean directions, the vMNB can be seen as a linear classifier finding the line that goes through the points on the circumference defined by ϕ' and ϕ'' defined as follows:

$$\phi' = \frac{1}{2}(\mu_1 + \mu_2),$$
$$\phi'' = \frac{1}{2}(\mu_1 + \mu_2) + \pi.$$

Hence the two discriminant regions are simply two semicircles.

For all *DD*-classifiers, the leave one-out cross-validation was used to determine the optimal k. The performances of the considered classifiers are then analyzed and compared in terms of accuracy rate which is computed as follows:

$$Accuracy = \frac{TP + TN}{N},$$

where TP and TN denote the true positive hits and the true negative correct rejections, respectively.

The classification procedures have been performed in an `R` environment. Circular data are simulated through the `rvonmises` in the `circular` package. The function `localdepth` in the `localdepth` package is used to compute the simplicial, while the function `sdepth` in the `depth` package is used to compute the Tukey. On the other hand, for the computation of the Arc distance, a specific function has been written by the authors. The k-NN procedure is performed with the `R` packages `FNN` and `class`.

Table 1 shows the mean accuracy and the standard deviation of the compared classifiers for each setup. As expected, when ν increases (for each setup) the estimated accuracy increases for all the compared classifiers. Furthermore, the accuracy increases as the distance between the two groups increases. The performances of the von Mises naive Bayes classifier are in line with its theoretical predictions discussed by Figueiredo (2009). However, the performances of the considered *DD*-classifiers deserve a great attention; indeed they all perform remarkably well showing accuracies which extremely similar to those of the von Mises Bayes classifier in each setup.

$d(\mu_1, \mu_2)$	ν	Classifier			
		DD_{ADD}	DD_{ASD}	DD_{ATD}	$vMNB$
$\pi/6$	3	0.667 (0.029)	0.656 (0.070)	0.668 (0.036)	**0.690** (0.028)
	5	0.732 (0.038)	0.734 (0.026)	0.743 (0.024)	**0.747** (0.033)
	10	0.822 (0.022)	0.823 (0.022)	0.825 (0.028)	**0.828** (0.026)
$\pi/3$	3	0.786 (0.030)	0.795 (0.020)	0.781 (0.019)	**0.797** (0.024)
	5	0.857 (0.024)	0.860 (0.023)	0.858 (0.022)	**0.863** (0.026)
	10	0.943 (0.011)	0.943 (0.010)	0.943 (0.010)	**0.946** (0.008)
$\pi/2$	3	0.879 (0.021)	**0.885** (0.024)	0.874 (0.026)	0.883 (0.023)
	5	0.953 (0.014)	**0.956** (0.012)	0.955 (0.014)	0.955 (0.011)
	10	0.992 (0.007)	**0.994** (0.006)	**0.994** (0.006)	0.993 (0.007)
$2/3\pi$	3	0.942 (0.014)	0.942 (0.019)	0.935 (0.017)	**0.943** (0.020)
	5	0.987 (0.005)	**0.990** (0.005)	0.988 (0.006)	0.988 (0.006)
	10	0.998 (0.004)	0.999 (0.002)	0.999 (0.002)	**0.999** (0.001)

Table 1: Average accuracy rate and standard deviation reported in parentheses (over 100 replications) of the considered classifiers under different distributional settings. The best results are marked in bold.

All the angular depth functions show similar performances in terms of average accuracy rate, thus no clear preference can be stated. As shown in Section 3.2, different depths can give rise to different data structures in the DD-plot. Hence, following the advice of Li et al. (2012), one can use a cross validation approach to choose the depth that yields the highest accuracy rate.

4.1 Real data examples

In this section, we apply the classifiers considered in this paper to two real data sets.

4.2 Back to Ackermann's data

Here we apply the proposed classification approach to Ackermann's data depicted in Figure 1. We used the leave-one-out cross validation to find the optimal k for the DD-classifier based on each of the three directional depth functions.

Table 2 shows the accuracy rates of the classifiers. More in detail, in this case the classifiers based on the simplicial performs better than the other ones and has an accuracy rate equal to 0.947. The Arc distance and the Tukey depths based classifiers have an accuracy rate equal to 0.895 just like the von Mises naive Bayes classifier.

To give some insight, in Figure 4 is depicted the decision boundary of a 2-nearest neighbor classifier on the DD-plot obtained through the angular simplicial depth for the Ackermann's data.

Classifier			
DD_{ADD}	DD_{ASD}	DD_{ATD}	$vMNB$
0.895	**0.947**	0.895	0.895

Table 2: Accuracy rate of the classifiers for the Ackermann's data. The best result are marked in bold.

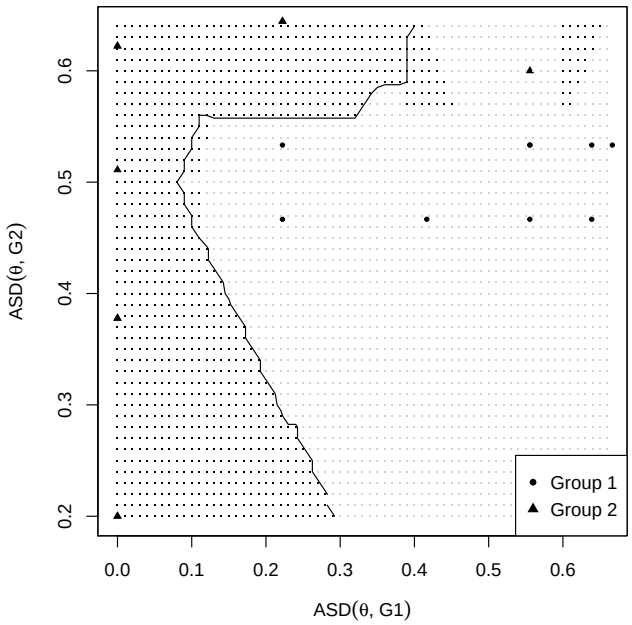

Figure 4: The decision boundary of a 2-nearest neighbor classifier on the DD-plot obtained through the angular simplicial depth for the Ackermann's data.

4.3 Megaspores data

The proposed classifiers and the von Mises naive Bayes classifier are evaluated through the Megaspores data set (included in Oriana software), where megaspores are classified into two classes according to the angle of their wall elements. The two groups of megaspores used in this study are called Selaginellalean and Isoetalean. The data set counts 960 instances, where 360 are Selaginellalean (37.5%) and 600 are Isoetalean (62.5%). Table 2 reports the mean accuracy of the classifiers obtained by a 10-fold cross validation averaged over 10 runs. The leave one-out cross-validation was used to determine the optimal k for the DD-classifiers based on data depths. Results show that all the classifiers have comparable performances.

As one can see, the performances of the compared classifiers are approximately equal with the von Mises naive Bayes and the angular Tukey depth based classifiers achieving slightly better accuracy.

Classifier			
DD_{ADD}	DD_{ASD}	DD_{ATD}	$vMNB$
0.758 (0.020)	0.760 (0.020)	**0.765** (0.021)	**0.765** (0.036)

Table 3: Average accuracy rate and standard deviation reported in parentheses of the considered classifiers for the megaspores data set using ten runs of a stratified 10-fold cross validation. The best result are marked in bold.

5 Conclusions

In this paper, we extended the depth vs. depth classification method to circular data. The idea of depth provides a criterion to order a circular sample from center-outward and gives a new way in classifying angular objects. The performances in terms of average accuracy rate of the DD-classifiers are investigated by means of simulated data drawn from the reference distribution for circular data, that is the von Mises distribution, and two real data examples in neurology and biology, and compared with the von Mises naive Bayes classifier. The results indicate that the DD-classifiers perform remarkably similar to the von Mises naive Bayes classifier. Hence, an interesting new way to perform nonparametric classification in the context of circular data seem to appear. Further developments for other distributions and for spherical data are under study and will be reported elsewhere.

Acknowledgements

Authors would like to thank the two anonymous reviewers, whose comments contributed to improve the quality of the manuscript. For Giuseppe Pandolfo and Antonio D'Ambrosio, this work has been partially supported by the H2020-EU.3.5.4. project "Moving Towards Adaptive Governance in Complexity: Informing Nexus Security (MAGIC)", Grant agreement number 689669.

References

Ackermann, H. (1997). A note on circular nonparametrical classification. *Biometrical journal*, 39(5):577–587.

Agostinelli, C. and Romanazzi, M. (2013). Nonparametric analysis of directional data based on data depth. *Environmental and ecological statistics*, 20(2):253–270.

Batschelet, E. (1981). *Circular statistics in biology*. Academic press.

Dutta, S. and Ghosh, A. K. (2011). On classification based on L^p depth with an adaptive choice of p. *Preprint*.

El Khattabi, S. and Streit, F. (1996). Identification analysis in directional statistics. *Computational Statistics & Data Analysis*, 23(1):45–63.

Figueiredo, A. (2009). Discriminant analysis for the von mises-fisher distribution. *Communications in Statistics-Simulation and Computation*, 38(9):1991–2003.

Ghosh, A. K. and Chaudhuri, P. (2005). On maximum depth and related classifiers. *Scandinavian Journal of Statistics*, 32(2):327–350.

Hartikainen, A. and Oja, H. (2006). On some parametric, nonparametric and semiparametric discrimination rules. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 72:61.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning*. Springer series in statistics New York, NY, USA.

Hubert, M. and Van der Veeken, S. (2010). Robust classification for skewed data. *Advances in Data Analysis and Classification*, 4(4):239–254.

Jörnsten, R. (2004). Clustering and classification based on the l1 data depth. *Journal of Multivariate Analysis*, 90(1):67–89.

Li, J., Cuesta-Albertos, J. A., and Liu, R. Y. (2012). DD-classifier: Nonparametric classification procedure based on dd-plot. *Journal of the American Statistical Association*, 107(498):737–753.

Liu, R. Y. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics*, pages 405–414.

Liu, R. Y., Parelius, J. M., Singh, K., et al. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference,(with discussion and a rejoinder by liu and singh). *The annals of statistics*, 27(3):783–858.

Liu, R. Y. and Singh, K. (1992). Ordering directional data: concepts of data depth on circles and spheres. *The Annals of Statistics*, pages 1468–1484.

Liu, Z. and Modarres, R. (2011). Lens data depth and median. *Journal of Nonparametric Statistics*, 23(4):1063–1074.

López-Cruz, P. L., Bielza, C., and Larrañaga, P. (2015). Directional naive bayes classifiers. *Pattern Analysis and Applications*, 18(2):225–246.

López-Pintado, S. and Romo, J. (2006). Depth-based classification for functional data. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 72:103.

Mardia, K. V. and Jupp, E. (1972). *Statistics of directional data*. Academic press, London.

Morris, J. E. and Laycock, P. (1974). Discriminant analysis of directional data. *Biometrika*, 61(2):335–341.

Mosler, K. and Hoberg, R. (2006). Data analysis and classification with the zonoid depth. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 72:49.

Paindaveine, D., Van Bever, G., et al. (2015). Nonparametrically consistent depth-based classifiers. *Bernoulli*, 21(1):62–82.

Pandolfo, G., Paindaveine, D., and Porzio, G. (2017). Distance-based Depths for Directional Data. *ArXiv e-prints*.

Sengupta, A. and Ugwuowo, F. I. (2011). A classification method for directional data

with application to the human skull. *Communications in Statistics - Theory and Methods*, 40(3):457–466.

Serfling, R. (2006). Depth functions in nonparametric multivariate inference. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 72:1.

Small, C. G. (1987). Measures of centrality for multivariate and directional distributions. *Canadian Journal of Statistics*, 15(1):31–39.

Stoller, D. S. (1954). Univariate two-population distribution-free discrimination. *Journal of the American Statistical Association*, 49(268):770–777.

Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians, Vancouver, 1975*, volume 2, pages 523–531.

Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Annals of statistics*, pages 461–482.