# ON THE OPTIMALITY OF THE MAX-DEPTH AND MAX-RANK CLASSIFIERS FOR SPHERICAL DATA

Ondřej Vencálek, Olomouc, Houyem Demni, Cassino,
Amor Messaoud, Tunis, Giovanni C. Porzio, Cassino

Received November 30, 2019. Published online May 25, 2020.

*Abstract.* The main goal of supervised learning is to construct a function from labeled training data which assigns arbitrary new data points to one of the labels. Classification tasks may be solved by using some measures of data point centrality with respect to the labeled groups considered. Such a measure of centrality is called data depth. In this paper, we investigate conditions under which depth-based classifiers for directional data are optimal. We show that such classifiers are equivalent to the Bayes (optimal) classifier when the considered distributions are rotationally symmetric, unimodal, differ only in location and have equal priors. The necessity of such assumptions is also discussed.

*Keywords*: depth-based classifier; von Mises-Fisher distribution; directional data; cosine depth

*MSC 2020*: 62H30, 62G30

## 1. Introduction

Supervised classification techniques enjoy a wide range of applications in many fields. Given a training set of observations and their membership of certain groups, new observations with unknown membership should be accordingly assigned. A fairly large number of classification rules are available in the literature (e.g. [9]).

Within this setting, depth-based classification procedures have been introduced. Depths provide center-outward ordering of points in multidimensional spaces with

This work was partially developed while Houyem Demni was visiting the Department of Mathematical Analysis and Applications of Mathematics, Palacký University in Olomouc. Supported by the University of Cassino and Southern Lazio Mobility Grant D.R.310.2019.

DOI: 10.21136/AM.2020.0331-19 

respect to a given distribution, and their applications often lead to effective robust statistical procedures. As a consequence, depth-based supervised classification techniques are typically able to deal with the presence of outliers or mislabeled observations in the training set [8]. Many depth based classifiers are available, and for a review we refer to the work of Vencálek [25]. On the other hand, depth-based supervised classification procedures have only recently been introduced in the directional data framework [5], [20], [21].

Spherical (or directional) data are data lying on the unit hyper-sphere. They occur naturally when a direction or an angle in space is of interest (e.g. wind direction), but also when data consist of time points and the interest is in cycles (time points on a watch can be treated as angles). In higher dimensions, locations on the Earth and/or any kind of information recordable as unit vectors can be analyzed from a directional data perspective.

Such data can be encountered in many fields of science and technology such as Earth sciences [4], [6], meteorology [3], neurosciences [13] or biology [2] to capture the direction of some phenomena of interest. Other interesting applications of directional data include shape analysis and its use in economics [11], [12].

Spherical data have their own specific features and therefore classical statistical methods need to be adjusted to these kinds of data. In this context, depths have been successfully applied ([1], [14], [16], [21]) and some robustness aspects have also been investigated. For instance, Pandolfo et al. [21] showed that the cosine depth deepest point achieves the highest directional breakdown point in terms of lower bound when compared to the chord and arc distance depth deepest points in the case of von Mises-Fisher distributions. A discussion on the finite-sample maximum bias of the cosine depth deepest point (the spherical mean) and the arc distance deepest point (the spherical median) is instead available in [10].

However, although the recently introduced depth-based classifiers for directional data performed well in simulation studies [5], [20], [21], corresponding theoretical results are still lacking. For all the above reasons, this work investigates properties of depth-based classifiers for directional data. It introduces the conditions under which these classifiers are optimal. That is, they are equivalent to the Bayes classifier, the classifier with the lowest achievable probability of misclassification. Special attention is paid to the case of von Mises-Fisher distributions, since they play a central role among models for directional data.

The paper is organized as follows. Section 2 introduces some basic concepts of directional data and the directional distance-based depth functions. Furthermore, it describes the max-depth and the max-rank classifiers for spherical data. Section 3 includes the main results. It provides a discussion on the assumptions under which the depth-based classifiers are optimal as well as the necessity of such assumptions. Final comments are provided in Section 4.

## 2. Background material

This section reviews basic concepts of directional data and their corresponding depth measures. Furthermore, it introduces data depth based classifiers.

**2.1. Directional data.** In $q$-dimensional space, directions can be depicted as points on the sphere $S^{q-1} = \{\boldsymbol{x} \in \mathbb{R}^q \colon \boldsymbol{x}'\boldsymbol{x} = 1\}$ or as vectors with unit radius and center at the origin. Note that in the two-dimensional case, any direction can be also described by an angle (observations are called *circular data* in this case). In the three-dimensional case, data points can also be described by two angles corresponding to longitude and latitude.

The basic location parameter of spherical data is the *mean direction* $\boldsymbol{\mu} = \mathrm{E}\boldsymbol{X}/\|\mathrm{E}\boldsymbol{X}\|$ (defined if and only if the value in the denominator is positive). A possible measure of variability, denoted traditionally as $\varrho$, is called the *mean resultant length* and is defined as $\varrho = \|\mathrm{E}\boldsymbol{X}\| = (\mathrm{E}\boldsymbol{X}'\mathrm{E}\boldsymbol{X})^{1/2}$.

In this paper, the class of rotationally symmetric distributions is considered. The *distribution* $H$ of a random variable $\boldsymbol{X}$ is said to be *rotationally symmetric about some vector* $\boldsymbol{\mu} \in S^{q-1}$ if and only if the distribution of $\boldsymbol{OX}$ is again $H$ for all $q \times q$ orthogonal matrices $\boldsymbol{O}$ satisfying $\boldsymbol{O\mu} = \boldsymbol{\mu}$. This class of distributions was first studied by Saw [22]. Any distribution which is rotationally symmetric about $\boldsymbol{\mu}$ and absolutely continuous w.r.t. a surface area measure on $S^{q-1}$ has a density of the form $h(\boldsymbol{x}) = g(\boldsymbol{x}'\boldsymbol{\mu})$ for some (univariate) function $g\colon [-1,1] \to \mathbb{R}_0^+$, e.g. [19].

The most widely used distribution on the sphere is the *von Mises-Fisher distribution*, which is also rotationally symmetric, e.g. [14]. The probability density function of a von Mises-Fisher distribution is defined as

$$h(\boldsymbol{x}; \boldsymbol{\mu}, \kappa) = c_{\kappa,q} \exp\{\kappa \boldsymbol{\mu}'\boldsymbol{x}\},$$

where $\boldsymbol{\mu}$ is the mean direction, $\kappa \geqslant 0$ is a concentration parameter, and $c_{\kappa,q} > 0$ is a normalizing constant (depending on parameters $\kappa$ and $q$). Its value is

$$c_{\kappa,q} = \left(\frac{\kappa}{2}\right)^{q/2-1} \frac{1}{\Gamma^{(q/2)} I_{q/2-1}(\kappa)},$$

where $I_v$ is the modified Bessel function of the first kind and order $v$, e.g. [18].

**2.2. Data depth for directional data.** The concept of data depth for directional data was first introduced by Small [23]. Later, it was extended by Liu and

Singh, see [16]. They introduced the *arc distance depth* and at the same time extended the simplicial depth (originally introduced in [15]) to the *directional angular simplicial depth* and the halfspace depth (originally introduced in [24]) to the *directional angular Tukey depth.*

In this paper, the class of depth based on rotational invariant distance is considered. It was introduced by Pandolfo et al. [21].

▷ The *directional distance-based depth* is defined as follows:

Let $d\colon\ S^{q-1} \times S^{q-1} \to \mathbb{R}_0^+$ be a bounded distance on the sphere $S^{q-1}$. Let $H$ be a probability distribution on $S^{q-1}$. The directional $d$-depth of a point $\boldsymbol{x} \in S^{q-1}$ with respect to the distribution $H$ is defined as

$$D(\boldsymbol{x}, H) = d^{\sup} - \mathrm{E}_H(d(\boldsymbol{x}, \boldsymbol{X})), \tag{2.1}$$

where $d^{\sup}$ is the upper bound of the distance between any two points on $S^{q-1}$, $\mathrm{E}_H$ is the expected value and $\boldsymbol{X}$ is a random directional variable from $H$.

▷ Rotational invariance is an important property of distance and subsequently depth.

A distance $d$ is *rotationally-invariant* if $d(\boldsymbol{Ox}, \boldsymbol{Oy}) = d(\boldsymbol{x}, \boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in S^{q-1}$ and all $q \times q$ orthogonal matrices $\boldsymbol{O}$.

Any rotationally-invariant (bivariate) distance $d(\boldsymbol{x}, \boldsymbol{y})$ can be expressed as a univariate function $\delta$ of the scalar product $\boldsymbol{x}'\boldsymbol{y}$, i.e.

$$d(\boldsymbol{x}, \boldsymbol{y}) = \delta(\boldsymbol{x}'\boldsymbol{y}), \tag{2.2}$$

as shown in [21] (Proposition 1). It is easy to see that any directional depth based on rotationally-invariant distance is also rotationally-invariant, i.e. $D(\boldsymbol{x}, H) = D(\boldsymbol{Ox}, H_{\boldsymbol{O}})$ for all $\boldsymbol{x} \in S^{q-1}$ and for all $q \times q$ orthogonal matrices $\boldsymbol{O}$, where $H_{\boldsymbol{O}}$ denotes distribution of $\boldsymbol{OX}$ when $\boldsymbol{X}$ has distribution $H$. See also Theorem 1 in [21].

Let us now recall the three most widely used rotationally-invariant distance-based depth functions: the cosine depth, the arc distance depth, and the chord depth.

- ▷ The *cosine depth* of a point $\boldsymbol{x} \in S^{q-1}$ w.r.t. the distribution $H$ of a random directional variable $\boldsymbol{X}$ is defined as $D_{\cos}(\boldsymbol{x}, H) := 2 - \mathrm{E}_H[(1 - \boldsymbol{x}'\boldsymbol{X})] = 1 + \mathrm{E}_H(\boldsymbol{x}'\boldsymbol{X})$ using the cosine distance $\delta(t) = 1 - t$.
- ▷ The *arc distance depth* of a point $\boldsymbol{x} \in S^{q-1}$ w.r.t. the distribution $H$ of a random directional variable $\boldsymbol{X}$ is defined as $D_{\mathrm{arc}}(\boldsymbol{x}, H) := \pi - \mathrm{E}_H[\arccos(\boldsymbol{x}'\boldsymbol{X})]$ using the arc distance $\delta(t) = \arccos(t)$.
- ▷ The *chord depth* of a point $\boldsymbol{x} \in S^{q-1}$ w.r.t. the distribution $H$ of a random directional variable $\boldsymbol{X}$ is defined as $D_{\mathrm{chord}}(\boldsymbol{x}, H) := 2 - \mathrm{E}_H[\sqrt{2(1 - \boldsymbol{x}'\boldsymbol{X})}]$ using the chord distance $\delta(t) = \sqrt{2(1 - t)}$.

**2.3. Max-depth and max-rank classifiers for directional data.** This section introduces the max-depth and the max-rank classifiers. The above classifiers can be associated with all the available depth functions for directional data within the literature. In this study, the cosine depth is preferred for the following reasons. First, the cosine depth does not require a large computational effort, unlike the other depths. Secondly, both classifiers provide good performance when associated with the cosine depth on hyper-spheres [5], [21]. Finally, the cosine depth (deepest point) can be considered a robust location estimator [21].

Consider now $K$ different distributions $H_1, \ldots, H_K$ on hyper-sphere $S^{q-1}$. A classification rule in the directional framework is a function

$$c\colon\ S^{q-1} \to \{1, \ldots, K\},$$

which assigns points on the hyper-sphere to distributions from which they are likely to come. Here we restrict our attention to the two-class problem ($K = 2$).

**2.3.1. Directional max-depth classifier.** The concept of max-depth classifier for multivariate data was developed by Ghosh and Chaudhuri, [7]. More recently, Pandolfo et al. [21] extended the max-depth classifier to the directional framework.

Let $\boldsymbol{x}$ be the new observation to be classified, and let $D(\boldsymbol{x}, H_i)$, $i = 1, \ldots, K$, be the depth of $\boldsymbol{x}$ with respect to the distributions $H_1, \ldots, H_K$, respectively. The max-depth classification rule is then given by

$$c_m(\boldsymbol{x}) = \arg\max_i D(\boldsymbol{x}; H_i). \tag{2.3}$$

In practice, theoretical distributions are unknown and need to be estimated. Therefore, one uses empirical distribution functions $\widehat{H}_i$ based on data points in the training set instead of theoretical distributions $H_i$.

**2.3.2. Directional max-rank classifier.** The depth distribution classifier known also as the max-rank classifier was introduced by Makinde and Fasoranbaku [17] for multivariate data and then extended to directional data by Demni et al. [5].

The *cumulative distribution function of the depth function* $D(\cdot, H)$, denoted as $F_D(\cdot, H)$, is defined as

$$F_D(\boldsymbol{x}, H) = \mathrm{P}(D(\boldsymbol{X}, H) \leqslant D(\boldsymbol{x}, H)), \tag{2.4}$$

where $\boldsymbol{X}$ is a random directional variable from the distribution $H$.

The *directional depth distribution classification rule* [5] is then defined as

$$c_{\mathrm{edd}}(\boldsymbol{x}) = \arg\max_i F_D(\boldsymbol{x}, H_i).$$

In practice, the unknown distributions $H_i$ are again replaced by their corresponding empirical distributions based on training set observations.

## 3. Properties of the max-depth and max-rank classifiers

The properties of the max-depth and max-rank classifiers are studied in this section. To our best knowledge, the optimality property of the depth-based classifiers has not been investigated elsewhere in the context of directional data.

The optimality of the considered depth-based classifiers was studied by Ghosh and Chaudhuri [7] (the max-depth classifier) and by Makinde and Fasoranbaku [17] (the max-rank classifier) in the context of multivariate (unconstrained) data. Both classifiers were shown to be equivalent to the optimal Bayes classifier (the classifier with the lowest total probability of misclassification) in some situations. More precisely, optimality is achieved if the considered distributions are elliptically symmetric with density strictly decreasing from the center (which implies unimodality of the distributions), differing only in location and having equal prior probabilities. While the assumptions on symmetry and unimodality are not too restrictive in practice, the assumptions on equal dispersions (implied by difference only in location) and equal priors reduce the applicability of the classifiers in practice quite substantially. We show that similar assumptions are needed for optimality also in the case of directional data.

**Theorem 3.1.** *Let $H_1$ and $H_2$ be rotational symmetric unimodal continuous distributions on the sphere $S^{q-1}$ differing only in their mean directions (denoted by $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, respectively), i.e., their densities $h_i(\cdot)$, $i = 1, 2$, can be expressed as $h_i(\boldsymbol{x}) = h(\boldsymbol{\mu}_i'\boldsymbol{x})$, $i = 1, 2$ for all $\boldsymbol{x} \in S^{q-1}$, where $h(\cdot)$ is some strictly increasing function. Let the distributions have equal prior probabilities $p_1 = p_2$. Then for any rotation-invariant distance-based depth, both the max-depth classifier and the max-rank classifier are equivalent to the (optimal) Bayes classifier.*

P r o o f. First, we simplify the form of the Bayes classifier in the considered settings. The Bayes classification rule assigns $\boldsymbol{x}$ to group 1 if and only if

$$p_1 h_1(\boldsymbol{x}) > p_2 h_2(\boldsymbol{x}).$$

In the case of equal priors and rotational symmetric distributions the inequality simplifies to $h(\boldsymbol{\mu}_1'\boldsymbol{x}) > h(\boldsymbol{\mu}_2'\boldsymbol{x})$. Since $h(\cdot)$ is a strictly increasing function, the inequality can be rewritten as

$$\boldsymbol{\mu}_1'\boldsymbol{x} > \boldsymbol{\mu}_2'\boldsymbol{x}.$$

Now we show that the max-depth classifier can be expressed in the same way. This results directly from Theorem 3 of [21] which showed that in the considered situation, depth can be expressed as a strictly increasing function of the cosine distance from the mean direction, i.e. $D(\boldsymbol{x}, H_i) = \varphi(\boldsymbol{\mu}_i'\boldsymbol{x})$, $i = 1, 2$, for some strictly increasing function $\varphi\colon [-1, 1] \to \mathbb{R}_0^+$. Since the function $\varphi$ is the same for both distributions, the inequality $D(\boldsymbol{x}, H_1) > D(\boldsymbol{x}, H_2)$ holds if and only if

$$\boldsymbol{\mu}_1'\boldsymbol{x} > \boldsymbol{\mu}_2'\boldsymbol{x}.$$

Finally, we deal with the max-rank classifier. For the distribution $H_i$, $i = 1, 2$, the cumulative distribution function of depth (2.4) can be expressed as

$$F_D(\boldsymbol{x}, H_i) = \mathrm{P}(D(\boldsymbol{X}, H_i) \leqslant D(\boldsymbol{x}, H_i)) = \int_{S(\boldsymbol{x})} h_i(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = \int_{S(\boldsymbol{x})} h(\boldsymbol{\mu}_i'\boldsymbol{y})\,\mathrm{d}\boldsymbol{y},$$

where $S(\boldsymbol{x}) = \{\boldsymbol{y} \in S^{q-1}\colon \boldsymbol{\mu}_i'\boldsymbol{y} < \boldsymbol{\mu}_i'\boldsymbol{x}\}$. Since we are integrating a non-negative function, the value of the integral increases with expanding the set $S(\boldsymbol{x})$. Therefore, the higher is the product $\boldsymbol{\mu}_i'\boldsymbol{x}$, the higher is the value of the integral, and hence $F_D(\boldsymbol{x}, H_1) > F_D(\boldsymbol{x}, H_2)$ if and only if

$$\boldsymbol{\mu}_1'\boldsymbol{x} > \boldsymbol{\mu}_2'\boldsymbol{x}.$$

□

In the following, we discuss the conditions which guarantee the Bayes optimality. The depth-based classifiers employ rotation-invariant distance-based depth functions and therefore the depth is a function of the cosine distance from the mean direction. To achieve correspondence between depth and density (used in the Bayes classifier), we have to assume that the density is also a function of the cosine distance from the mean direction, i.e. the rotational symmetry of the distribution. We further need assumption of monotonicity of a function $h(\cdot)$ to avoid situations in which the density is low in points close to the mean direction. As already mentioned at the beginning of this section, the other assumptions—on equal variability and equal priors—reduce the applicability of the classifiers in practice quite substantially. Therefore, we investigated the performance of the classifiers in the case of unequal concentrations.

**3.1. The max-depth classifier in a more general case.** The following theorem clarifies the form of the max-depth classifier for the cosine depth in the situation in which the considered distributions may differ not only in location but also in dispersion.

**Theorem 3.2.** *Let $H_1$ and $H_2$ be two distributions on the sphere $S^{q-1}$, having mean directions $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, respectively, and mean resultant lengths $\varrho_1$ and $\varrho_2$, respectively. If the cosine depth is employed, the max-depth classifier* (2.3) *has the form*

$$c(\boldsymbol{x}) = \arg\max_i \varrho_i \boldsymbol{\mu}_i' \boldsymbol{x}, \tag{3.1}$$

*and therefore, the distributions are "separated" by the hyperplane*

$$(\varrho_1 \boldsymbol{\mu}_1 - \varrho_2 \boldsymbol{\mu}_2)' \boldsymbol{x} = 0. \tag{3.2}$$

P r o o f. The theorem directly follows from the form of the cosine depth in this case:

$$D(\boldsymbol{x}, H_i) = 1 + \mathrm{E}_{H_i} \boldsymbol{x}' \boldsymbol{X} = 1 + \varrho_i \boldsymbol{x}' \boldsymbol{\mu}_i.$$

□

The separating hyperplane is determined by the parameters of location ($\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$) and parameters reflecting variability of the distributions ($\varrho_1$ and $\varrho_2$). Clearly, the max-depth classifier does not include (and hence does not account for) information on priors. Also, the whole information on distribution is reduced only to its mean direction and mean resultant length.

In the case of equal mean resultant lengths, the max-depth classifier simplifies to the form $c(\boldsymbol{x}) = \arg\max \boldsymbol{\mu}_i' \boldsymbol{x}$. The separating hyperplane is then determined by the equation $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{x} = 0$. It is a hyperplane orthogonal to the hyperplane determined by vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ which halves the angle between them.

Note that the formula of the max-depth classifier cannot be simplified in this way when using nonlinear transformations of the scalar product $\boldsymbol{\mu}_i' \boldsymbol{x}$ in the depth function even if the transformation is monotone, i.e. for the arc distance depth and the chord depth.

Geometrically, we can imagine the above-described situation as follows. Denote the angle between $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ as $\theta$ ($\cos\theta = \boldsymbol{\mu}_1' \boldsymbol{\mu}_2$). There exists an orthogonal matrix $\boldsymbol{R}$ such that

$$\boldsymbol{R}\boldsymbol{\mu}_1 = \left(\cos\tfrac{\theta}{2}, \sin\tfrac{\theta}{2}, 0, \ldots, 0\right)' =: \boldsymbol{\mu}_1^0$$

and

$$\boldsymbol{R}\boldsymbol{\mu}_2 = \left(\cos\tfrac{\theta}{2}, -\sin\tfrac{\theta}{2}, 0, \dots, 0\right)' =: \boldsymbol{\mu}_2^0.$$

Hence, we can assume that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_1^0$ and $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_2^0$.

In this situation, the cosine depth of a point $\boldsymbol{x} = (x_1, x_2, \dots, x_q)'$ can be expressed in the following form:

$$D(\boldsymbol{x}, H_1) = 1 + \varrho_1\left(x_1 \cos\tfrac{\theta}{2} + x_2 \sin\tfrac{\theta}{2}\right),$$
$$D(\boldsymbol{x}, H_2) = 1 + \varrho_2\left(x_1 \cos\tfrac{\theta}{2} - x_2 \sin\tfrac{\theta}{2}\right).$$

The separating hyperplane is then determined by the equation

$$(\varrho_1 - \varrho_2)\left(\cos\tfrac{\theta}{2}\right)x_1 + (\varrho_1 + \varrho_2)\left(\sin\tfrac{\theta}{2}\right)x_2 = 0,$$

which simplifies to the form $x_2 = 0$ in the case of equal mean resultant lengths.

**3.2. Studied class of spherical distributions.** We studied a broad subclass of unimodal rotational symmetric distributions on the sphere $S^{q-1}$ for which the Bayes classifier can be derived and subsequently compared to the max-depth classifier discussed above.

Let us consider a density function $h(\boldsymbol{x})$ proportional to a sum $v + g(\boldsymbol{\mu}'\boldsymbol{x})$, where $v > 0$ is a positive real constant, $\boldsymbol{\mu} \in S^{q-1}$ mean direction and $g\colon [-1, 1] \to \mathbb{R}$ is an odd strictly increasing function.

Note that the higher the value of the constant $v$, the closer is the distribution to the uniform distribution. Therefore, higher values of $v$ imply higher variability. Parameter $v$ can be thus understood as a measure of variability.

After plugging in the normalizing constant, the density can be expressed as

$$h(\boldsymbol{x}) = \frac{1}{A_q} + \frac{g(\boldsymbol{\mu}'\boldsymbol{x})}{vA_q},$$

where $A_q$ denotes the surface area of the sphere $S^{q-1}$. Assuming that $\boldsymbol{\mu} = (1, 0, \dots, 0)'$, which can be achieved by a rotation of the distribution, one can derive a relation between the variability parameter $v$ and the mean resultant length $\varrho$ of the considered distribution:

$$\varrho = \mathrm{E}X_1 = \int_{S^{q-1}} x_1 h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_{S^{q-1}} x_1\left(\frac{1}{A_q} + \frac{g(x_1)}{vA_q}\right)\mathrm{d}\boldsymbol{x} = \frac{1}{vA_q}G_q,$$

where $G_q = \int_{S^{q-1}} x_1 g(x_1)\,\mathrm{d}\boldsymbol{x}$ is a constant. The density can thus be expressed as a function of its mean direction $\boldsymbol{\mu} \in S^{q-1}$ and mean resultant length $\varrho$ (using above-defined constants $A_q$ and $G_q$) in the following way:

$$h(\boldsymbol{x}) = \frac{1}{A_q} + \frac{1}{G_q}\varrho g(\boldsymbol{\mu}'\boldsymbol{x}). \tag{3.3}$$

Let us now consider a classification problem for two distributions with densities of the above-mentioned form (3.3) with possibly different mean directions and mean resultant lengths, but with the same function $g(\cdot)$, i.e., we assume

$$h_i(\boldsymbol{x}) = \frac{1}{A_q} + \frac{1}{G_q}\varrho_i g(\boldsymbol{\mu}_i'\boldsymbol{x}), \quad i = 1, 2.$$

Assuming equal prior probabilities, the Bayes classifier can be expressed as

$$c(\boldsymbol{x}) = \arg\max_i \varrho_i g(\boldsymbol{\mu}_i'\boldsymbol{x}).$$

If $g$ is the identity, i.e. $g(y) = y$, the Bayes classifier is equivalent to the max-depth classifier if the cosine depth is employed (see Theorem 3.2 above).

We have shown that equal variability (expressed by the mean resultant length) is not a necessary condition for optimality. We found a class of distributions, in which optimality is achieved even if distributions differ in variability (identity or some multiple of identity are the only cases of $g(\cdot)$ in which the Bayes classifier coincides with the max-depth classifier).

**3.3. Bayes classifier in the case of von Mises-Fisher distributions.** The class of distributions studied in the previous section does not include the most well-known distribution on the sphere, namely the von Mises-Fisher (vMF) distribution. In this section, we briefly discuss this important case. Let us consider two different vMF distributions, i.e. distributions with densities

$$h_i(\boldsymbol{x}; \boldsymbol{\mu}_i, \kappa_i) = c_{\kappa_i,q} \exp\{\kappa_i \boldsymbol{\mu}_i'\boldsymbol{x}\}, \quad i = 1, 2.$$

The equation defining the separating subspace for the Bayes classifier given by equality $\pi_1 h_1(\boldsymbol{x}) = \pi_2 h_2(\boldsymbol{x})$ can be rewritten as

$$(\kappa_2\boldsymbol{\mu}_2 - \kappa_1\boldsymbol{\mu}_1)'\boldsymbol{x} = \ln\Big(\frac{\pi_1 c_{\kappa_1,q}}{\pi_2 c_{\kappa_2,q}}\Big). \tag{3.4}$$

As with the max-depth classifier (3.2), the separation subspace is a hyperplane. However, mean directions are multiplied by concentration parameters $\kappa$ here, not by mean resultant lengths $\varrho$. The relationship between these parameters of variability is not straightforward. The following holds:

$$\varrho = \frac{I_{q/2}(\kappa)}{I_{q/2-1}(\kappa)}, \tag{3.5}$$

where $I_v$ is the modified Bessel function of the first kind and order $v$, see Section 9.3.2 of [18]. Note that the ratio (3.5) is strictly increasing in $\kappa$. Moreover, the constant on the right-hand side of (3.4) is non-zero in the case of differing priors and concentration parameters (if the considered ratio is not equal to one by chance).

## 4. Final Remarks

This paper reviewed two depth-based classifiers for directional data, namely the max-depth and max-rank classifiers, and discussed conditions under which they are equivalent to the Bayes (optimal) classifier. Conditions under which optimality is guaranteed include (rotational) symmetry and unimodality of the underlying distributions, with a difference only in location and equal prior probabilities.

These conditions are not necessary and we found a class of rotational symmetric distributions for which the max-depth classifier based on the cosine depth can be optimal even if distributions also differ in variability. On the other hand, such a class does not include the von Mises-Fisher distributions, and the max-depth classifier is generally not optimal when groups present different variability levels. Moreover, it was shown that the above classifiers ignore information on prior probabilities.

### *References*


[1] *C. Agostinelli, M. Romanazzi*: Nonparametric analysis of directional data based on data depth. Environ. Ecol. Stat. *20* (2013), 253–270. MR doi

[2] *E. Batschelet*: Circular Statistics in Biology. Mathematics in Biology. Academic Press, London, 1981. zbl MR

[3] *J. A. Bowers, I. D. Morton, G. I. Mould*: Directional statistics of the wind and waves. Appl. Ocean Research *22* (2000), 13–30. doi

[4] *T. Chang*: Spherical regression and the statistics of tectonic plate reconstructions. Int. Stat. Rev. *61* (1993), 299–316. doi

[5] *H. Demni, A. Messaoud, G. C. Porzio*: The cosine depth distribution classifier for directional data. Applications in Statistical Computing. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Cham, 2019, pp. 49–60. MR doi

[6] *N. I. Fisher*: Smoothing a sample of circular data. J. Struct. Geol. *11* (1989), 775–778. doi

[7] *A. K. Ghosh, P. Chaudhuri*: On maximum depth and related classifiers. Scand. J. Stat. *32* (2005), 327–350. zbl MR doi

[8] *M. Hubert, P. Rousseeuw, P. Segaert*: Multivariate and functional classification using depth and distance. Adv. Data Anal. Classif., ADAC *11* (2017), 445–466. zbl MR doi

[9] *G. James, D. Witten, T. Hastie, R. Tibshirani*: An Introduction to Statistical Learning: With applications in R. Springer Texts in Statistics 103. Springer, New York, 2013. zbl MR doi

[10] *T. Kirschstein, S. Liebscher, G. Pandolfo, G. C. Porzio, G. Ragozini*: On finite-sample robustness of directional location estimators. Comput. Stat. Data Anal. *133* (2019), 53–75. zbl MR doi

[11] *T. Klecha, D. Kosiorowski, D. Mielczarek, J. P. Rydlewski*: New proposals of a stress measure in a capital and its robust estimator. Available at `https://arxiv.org/abs/1802.03756` (2018), 24 pages.

[12] *D. Kosiorowski*: About phase transitions in Kendall's shape space. Acta Univ. Lodz., Folia Oeconomica *206* (2007), 137–155.

[13] *P. Leong, S. Carlile*: Methods for spherical data analysis and visualization. J. Neurosci. Methods *80* (1998), 191–200. doi
[14] *C. Ley, C. Sabbah, T. Verdebout*: A new concept of quantiles for directional data and the angular Mahalanobis depth. Electron. J. Stat. *8* (2014), 795–816. zbl MR doi
[15] *R. Y. Liu*: On a notion of data depth based on random simplices. Ann. Stat. *18* (1990), 405–414. zbl MR doi
[16] *R. Y. Liu, K. Singh*: Ordering directional data: Concepts of data depth on circles and spheres. Ann. Stat. *20* (1992), 1468–1484. zbl MR doi
[17] *O. S. Makinde, O. A. Fasoranbaku*: On maximum depth classifiers: Depth distribution approach. J. Appl. Stat. *45* (2018), 1106–1117. MR doi
[18] *K. V. Mardia, P. E. Jupp*: Directional Statistics. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, 2000. zbl MR doi
[19] *D. Paindaveine, T. Verdebout*: Optimal rank-based tests for the location parameter of a rotationally symmetric distribution on the hypersphere. Mathematical Statistics and Limit Theorems. Springer, Cham, 2015, pp. 249–269. zbl MR doi
[20] *G. Pandolfo, A. D'Ambrosio, G. C. Porzio*: A note on depth-based classification of circular data. Electron. J. Appl. Stat. Anal. *11* (2018), 447–462. MR doi
[21] *G. Pandolfo, D. Paindaveine, G. C. Porzio*: Distance-based depths for directional data. Can. J. Stat. *46* (2018), 593–609. zbl MR doi
[22] *J. G. Saw*: A family of distributions on the $m$-sphere and some hypothesis tests. Biometrika *65* (1978), 69–73. zbl MR doi
[23] *C. G. Small*: Measures of centrality for multivariate and directional distributions. Can. J. Stat. *15* (1987), 31–39. zbl MR doi
[24] *J. W. Tukey*: Mathematics and the picturing of data. Proceedings of the International Congress of Mathematicians. Canad. Math. Congress, Montreal, 1975, pp. 523–531. zbl MR
[25] *O. Vencálek*: Depth-based classification for multivariate data. Austrian J. Stat. *46* (2017), 117–128. doi

*Authors' addresses*: *Ondřej Vencálek* (corresponding author), Faculty of Science of the Palacký University in Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic, e-mail: `ondrej.vencalek@upol.cz`; *Houyem Demni*, University of Cassino and Southern Lazio, Viale dell'Università, Cassino (FR) 03043, Italy; Laboratoire ARBRE, Institut Supérieur de Gestion de Tunis, 41 Avenue de la Liberté, Tunis 2000, Tunisia, e-mail: `houyem66@gmail.com`; *Amor Messaoud*, Laboratoire ARBRE, Institut Supérieur de Gestion de Tunis, Université de Tunis, 41 Avenue de la Liberté, Tunis 2000, Tunisia, e-mail: `amor.messaoud@gmail.com`; *Giovanni C. Porzio*, Department of Economics and Law, University of Cassino and Southern Lazio, Viale dell'Università, Cassino (FR) 03043, Italy, e-mail: `porzio@unicas.it`.