

Research Article

Depth-Based Classification for Distributions with Nonconvex Support

Daniel Hlubinka¹ and Ondrej Vencalek²

¹ Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Sokolovska 83, 186 75 Praha 8, Czech Republic

² Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacky University in Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic

Correspondence should be addressed to Ondrej Vencalek; ondrej.vencalek@upol.cz

Received 30 April 2013; Accepted 13 August 2013

Academic Editor: Zhidong Bai

Copyright © 2013 D. Hlubinka and O. Vencalek. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Halfspace depth became a popular nonparametric tool for statistical analysis of multivariate data during the last two decades. One of applications of data depth considered recently in literature is the classification problem. The data depth approach is used instead of the linear discriminant analysis mostly to avoid the parametric assumptions and to get better classifier for data whose distribution is not elliptically symmetric, for example, skewed data. In our paper, we suggest to use weighted version of halfspace depth rather than the halfspace depth itself in order to obtain lower misclassification rate in the case of “nonconvex” distributions. Simulations show that the results of depth-based classifiers are comparable with linear discriminant analysis for two normal populations, while for nonelliptic distributions the classifier based on weighted halfspace depth outperforms both linear discriminant analysis and classifier based on the usual (nonweighted) halfspace depth.

1. Introduction

Data classification or discrimination is an old statistical problem which has been discussed, studied, and applied since more than century. Therefore we will recall the problem only very briefly. The goal of the classification is to allocate a new observation into one of two (or more) groups. The rule for assessing a new observation to one of the possible groups is created by analysis of available observations with known group assignment (so-called training set).

The most popular classification method is based on normality assumption, and it is known as the linear discriminant analysis (LDA). LDA is easy to use and quite successful in many cases, in particular for elliptically symmetric distributions.

Naturally there are also nonparametric methods of classification. One of the motivations for replacing the LDA method is to avoid the normality assumption and therefore, hopefully, get better results for nonelliptic distributions. Recently it was proposed to use data depth as the basis for new nonparametric classifiers, see, for example, [1] or [2].

Let us recall the essential facts about the data depth. A depth of a point \mathbf{x} with respect to a probability measure P is a nonnegative number $D(\mathbf{x}, P)$ which measures the “centrality” of \mathbf{x} with respect to P . In other words, depth should reflect position of the point with respect to the probability distribution, or in the sample version, position of the point with respect to the observed data cloud. Recall also that for a multivariate data there is no natural linear ordering available. The depth, as a nonnegative (and bounded) number allows to define a depth-based ordering of multivariate data. Points of a high depth are “central” while points of a low depth are “outliers”; hence we speak about the “centre-outward” ordering.

There are many possibilities how to define data depth. The most popular and known are the halfspace (Tukey) depth [3] and the simplicial depth [4]. Other depths include the zonoid depth [5] or L_1 depth [6] among many others. Complete overview of the data depth, its properties, computational aspects, and applications may be found in Liu et al. [7].

The main advantage of LDA is its computational simplicity. On the other hand the assumption of normality is in some situations inadequate, and depth-based methods are more

appropriate, for example, for nonconvex distributions (in the sense of nonconvex level sets of density or even nonconvex support of distribution) or for functional data (observations in function space) classification [8]. Unfortunately the computation of most data depth functions (even of the most popular halfspace and simplicial) is very slow, in particular in higher dimensions. The L_1 depth is computationally much easier but it has several disadvantages for classification use. For example, points far from convex hull of all training sets still have positive L_1 depth. Therefore we restrict ourselves to the halfspace depth and its generalization; the simplicial depth and its generalization would give very similar conclusions.

The paper is divided into two main parts. In the first part we introduce the max-depth classification rule. Then the halfspace depth function is defined, and the maximal depth classifier for the halfspace depth is discussed. Some disadvantages of the classifier suggest changing some of the depth properties in order to improve the depth-based classifier. We propose to use the generalized halfspace depth which is defined, and it is shown that thanks to its “local” property it may be considered as a better choice of depth for depth-based classification than the original halfspace depth.

Section 3 is devoted to simulations. The simulation study shows that considering the generalized halfspace depth for the maximal depth classifier is usually better than the halfspace depth itself for any distribution. In particular, it is shown that the classification based on the generalized halfspace depth [9] gives comparable results as the halfspace depth classification and LDA for normal populations while it is *much better* for “nonconvex” distributions.

2. Classification Based on a Depth Function

In what follows we consider $K \geq 2$ populations $\mathcal{P}_1, \dots, \mathcal{P}_K$ with unknown absolutely continuous d -variate probability distributions $P_i, i = 1, \dots, K$, respectively. The corresponding empirical probability measures $\hat{P}_i, i = 1, \dots, K$ may be easily calculated from the training sets $\mathbf{Y}_{i,j}$, where $i = 1, \dots, K$, and $Y_{i,1}, \dots, Y_{i,m_i}$ are iid observations from the distribution P_i . Having a new observation $\mathbf{X} = \mathbf{x}$ (independent of the training sets and following unknown one of the distributions P_i) the problem is to find a rule (measurable map) $d : \mathbf{x} \in \mathbb{R}^d \mapsto \{1, \dots, K\}$ allocating the observation to a population $\mathcal{P}_{d(\mathbf{x})}$. In other words classification rule d means that the unknown distribution of \mathbf{X} is estimated by $P_{d(\mathbf{x})}$. As a usual measure of quality of classification the misclassification rates are considered in what follows, that is, the probability $P_i[d(\mathbf{X}) \neq i]$ if \mathbf{X} follows the distribution P_i .

2.1. Maximal Depth Classifier. The basic depth-based rule for classification is simple and natural. Consider a depth function D . Compute the depth of a new observation \mathbf{x} with respect to all empirical distribution functions $\hat{P}_1, \dots, \hat{P}_K$ and allocate \mathbf{x} to the population to which it has maximal depth:

$$d(\mathbf{x}) = \arg \max_{i=1, \dots, K} D(\mathbf{x}, \hat{P}_i). \quad (1)$$

The classifier is known as the maximal depth classifier. This idea was used by Jörnsten [1], Hartikainen and Oja [10], or Mosler and Hoberg [11].

2.2. Halfspace Depth. The halfspace depth is the most popular and the most frequently used depth function. Briefly speaking the halfspace depth of a point \mathbf{x} is the minimal probability of any closed halfspace which contains this point. More formal definition follows.

Definition 1. Let P be a probability measure on \mathbb{R}^d . The halfspace depth of the point $\mathbf{x} \in \mathbb{R}^d$ with respect to P is defined as

$$\text{HD}(\mathbf{x}, P) = \inf_{\mathbf{u}: \|\mathbf{u}\|=1} P\{\mathbf{y} : \mathbf{u}^T(\mathbf{y} - \mathbf{x}) \geq 0\}. \quad (2)$$

The sample version of the halfspace depth is defined as $\text{HD}(\mathbf{x}, \hat{P})$, where \hat{P} is the empirical probability measure. The maximal possible depth is $1/2$ for the symmetric distributions ($1/2$ is attained at the centre of symmetry in such case); for general distributions the maximal depth may be much lower. Points that are not in a convex hull of the support of distribution (convex hull of observations) have zero depth.

Halfspace depth has many desirable properties; see discussion of Zuo and Serfling [12]. The halfspace depth is in particular strongly consistent, affine invariant, vanishing at infinity, and maximal at centre of symmetry (if it exists); there exists unique deepest point for any absolutely continuous distributions, the depth is decreasing along rays from the deepest point, depth level sets are always convex, and so forth.

Ghosh and Chaudhuri [2] have proved that the maximal depth classifier based on the halfspace depth is asymptotically optimal (it has the lowest possible average misclassification rate) if the considered distributions are elliptically symmetric with the density function strictly decreasing in every direction from its centre of symmetry, differ only in location and rotation and have equal prior probabilities. We will show that for nonelliptic distributions the halfspace depth classifier is far from being optimal.

2.3. Problem of Distributions with Nonconvex Density Level sets. One particularly important property of the halfspace depth is its *quasiconcavity*. Quasiconcavity means that level sets (and equivalently central regions) of halfspace depth are always convex; that is, for any $a \geq 0$ and for any probability measure P the set $\{\mathbf{x} \in \mathbb{R}^d : \text{HD}(\mathbf{x}, P) \geq a\}$ is convex (or empty). However, this property becomes a *disadvantage* for classification if the level sets of density function are not convex.

Let us show an example of two bivariate distributions with *disjoint supports* such that LDA has misclassification rate equal almost to $1/2$ (the estimated normal distributions are almost the same) and the central regions of halfspace depth have large intersections; hence the misclassification rate is relatively high.

Example 2. Consider two bivariate uniform distributions P_1 and P_2 on fixed bounded sets (supports). The (nonconvex)

supports of the distributions are disjoint, and both are symmetric about the origin:

$$\begin{aligned} \text{supp}(P_1) &= \left(\left\{ x, y : \frac{x}{y} > 1 \right\} \right. \\ &\quad \left. \cup \left\{ x, y : -1 < \frac{x}{y} < 0 \right\} \right) \\ &\quad \cap \{ x, y : x^2 + y^2 < 1 \}, \\ \text{supp}(P_2) &= \{ x, y : x^2 + y^2 < 1 \} \setminus \text{supp}(P_1). \end{aligned} \tag{3}$$

The supports are plotted as grey areas in Figure 1.

Although the supports are disjoint and the classification may seem to be an easy task it is not true at all. The halfspace depth central regions are quite similar and with large intersections rather than disjoint. In Figure 1 the contours of the 0.25, 0.50 and 0.75 central regions (multivariate analogy of univariate quantiles) are shown. Clearly the discrimination cannot be good in this situation. In Section 3 it will be shown that the LDA is even worse in this case (almost as worse as random classifier) since the estimated means and variance matrices are almost equal in most simulations. The problem lies obviously in the construction of supports $\text{supp}(P_1)$ and $\text{supp}(P_2)$. The convexity of both halfspace depth level sets and normal distribution density level sets is in big contrast with the nonconvexity of the underlying distributions P_1 and P_2 . Therefore there is a natural question to find a depth function which may reflect this nonconvexity.

2.4. Generalised Halfspace Depth. Motivated by Example 2 we may try to use a generalised version of halfspace depth proposed by Hlubinka et al. [9]. The main idea of the generalisation is to use weighted probability in the halfspace rather than the probability of halfspace itself. The generalized halfspace depth may be also considered as *localized* version of the halfspace depth. Recall that the depth reflects *global position* of a point with respect to the distribution. The localized depth should also include more local properties of the distribution beside the global position. We will see that adding the locality may improve performance of the max-depth classifier substantially.

Definition 3. Let P be a probability measure on \mathbb{R}^d . The weighted halfspace depth of the point $\mathbf{x} = (x_1, \dots, x_d)$ with respect to P is defined as

$$\text{GD}(\mathbf{x}, P) = \inf_{A \in \mathcal{O}(d)} \frac{E_P \omega(A(X - \mathbf{x}))}{E_P \omega'(A(X - \mathbf{x}))}, \tag{4}$$

where $\mathcal{O}(d)$ is a space of all $d \times d$ orthogonal matrices, $\omega : \mathbb{R}^d \rightarrow \mathbb{R}^+$ is a weight function such that $\omega(\mathbf{x}) = 0$ for $x_d < 0$, and $\omega'(x_1, \dots, x_d) = \omega(x_1, \dots, x_{d-1}, -x_d)$. We set $0/0 = 1$.

Let us first clarify the connection between the weighted halfspace depth and the halfspace depth itself. If we consider weight function equal to 1 on the whole halfspace, that is,

$\omega(\mathbf{x}) = 1$ for all $\mathbf{x} \in \mathbb{R}^d : x_d \geq 0$, we get a depth function, which can be easily “converted” to the halfspace depth, since

$$\text{HD}(\mathbf{x}, P) = \frac{\text{GD}(\mathbf{x}, P)}{1 + \text{GD}(\mathbf{x}, P)} \tag{5}$$

in this case. As this transformation is monotone, the ordering based on the weighted halfspace depth is the same as the ordering based on the halfspace depth in this case. While the halfspace depth of a point \mathbf{x} can be interpreted as the minimal probability of any closed halfspace containing this point, the weighted halfspace depth is the minimal ratio of probabilities of complementary halfspaces. The difference between the weighted halfspace depth and the halfspace depth might be observed only when other weights are considered. Weights are used to “localize” the depth—greater weight might be assigned to points which are “near” the considered point \mathbf{x} . One particularly useful choice of the weights will be presented in the following paragraphs.

The weighted halfspace depth loses some properties of the halfspace depth. The depth is still strongly consistent for “regular” weight functions [9]. It is affine invariant for translation and rotation only, not for scaling; hence, it is generally recommended to use some of the transformation techniques [13] to obtain affine invariant version. There is unique deepest point for symmetric absolutely continuous distributions but for general distributions more deepest points may exist (but once more, it can be even advantage for classification). The central regions of generalised halfspace depth need not to be convex; they even need not to be connected sets. This is mainly due to the fact that a weight function may emphasise the “local” properties of the underlying distribution.

One of the weight functions which is simple to use and has the “localisation” property is the *band weight function* defined in d -variate case as

$$\omega(x_1, \dots, x_d) = \begin{cases} 1, & \text{if } \sum_{i=1}^{d-1} x_i^2 < c, \ x_d > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{6}$$

where $c > 0$ is half of the *bandwidth*. We shall call the generalised halfspace depth with the band weight function *the band depth*. From the geometrical point of view, we restrict our attention from the whole halfspace only to a band (where the weights is nonzero). The band contains points with the first $d - 1$ coordinates close to the first $d - 1$ coordinates of the considered point \mathbf{x} .

The choice of the bandwidth c may be done based on the training set. Clearly, the smaller is the value c the more “localised” the band depth is while for higher c the band depth is approximating the halfspace depth. Namely, in Example 2 the band depth is equivalent to the halfspace depth if $c \geq 2$. The main problem of choosing c is, however, in the computation of the depth. Depth computation is relatively slow (compared to LDA) as it is necessary to calculate the depth for each observation. Obviously, for optimal choice of c the calculation of band depth must be repeated for different values c ; hence the time needed for classification grows rapidly.

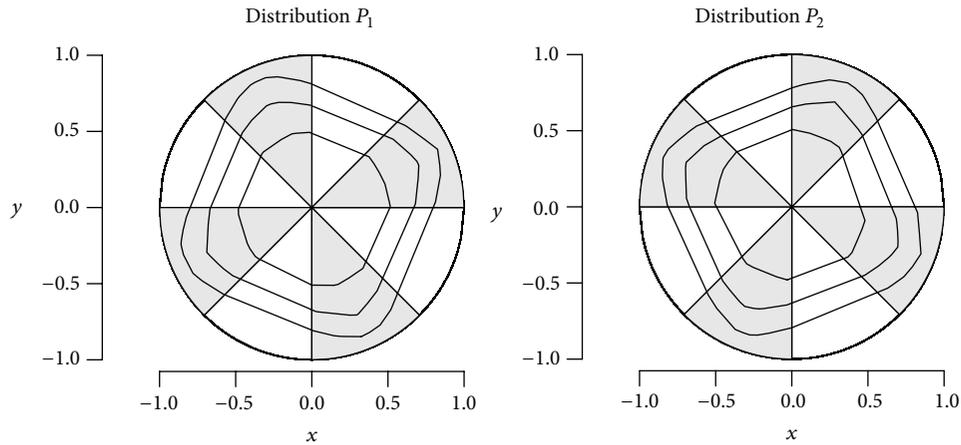


FIGURE 1: Supports of two bivariate uniform distributions (represented by grey areas). Estimated (solid curves) are halfspace depth central regions of 25, 50, and 75 percent of deepest points.

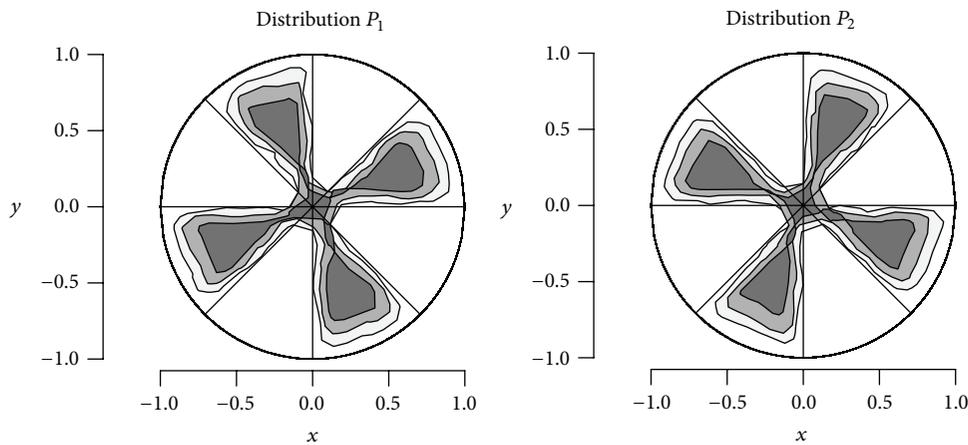


FIGURE 2: Estimated central regions of 25, 50, and 75 percent of deepest points when considering generalized (band) halfspace depth.

In Figure 2 we can see that the shapes of the band depth central regions “mimic” the shape of the supports. In particular the value $c = 0.15$ is used here. We have tried several choices, and then $c = 0.15$ was chosen to have sufficiently “smooth” borders of central regions, and simultaneously the central regions do not cover much areas without observations of the training sets. Comparing contours on Figures 1 and 2 it is not surprising that the results of band depth-based classification are much better than the corresponding results for halfspace depth-based classifier.

2.5. Observations of Zero Depth. Most of the depth functions assign zero depth to points outside the convex hull of support of distribution. In empirical version it means that points outside the convex hull of a training set have zero depth. In the case of weighted halfspace depth there may exist points in the convex hull of support with zero depth; see Figure 2.

Let us consider a situation that for a new observation \mathbf{x} it holds $D(\mathbf{x}, \hat{P}_i) = 0$ for all $i = 1, \dots, K$; the observation has zero depth with respect to all training sets. What should

be the classification $d(\mathbf{x})$ in this case? There are several possibilities how to solve this situation.

- (1) Other depth function (e.g., Mahalanobis depth or L_1 depth) which is positive for any point may be used. This method was used, for example, by Mosler and Hoberg [11] who combined zonoid and Mahalanobis depth.
- (2) $d(\mathbf{x})$ may be chosen to minimise the Euclidean distance from the (nearest) deepest point.
- (3) $d(\mathbf{x})$ may be chosen to minimise the Euclidean distance from a specific central region (regions).
- (4) Central regions may be inflated, and $d(\mathbf{x})$ is chosen such that the inflation needed to cover \mathbf{x} is the smallest one.

Rules 1 and 2 are almost useless in Example 2 since origin is the theoretical deepest point of both populations, and the Mahalanobis depth gives almost the same value for both training sets to any point. Method 4 works well for star-shaped central regions (as in Example 2).

TABLE 1: Simulated situations.

| | Population 1 | Population 2 |
|-------------|---|---|
| Situation 1 | $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 16 & 0 \\ 0 & 4 \end{pmatrix}\right)$ | $N\left(\begin{pmatrix} 6 \\ 0 \end{pmatrix}, \begin{pmatrix} 16 & 0 \\ 0 & 4 \end{pmatrix}\right)$ |
| Situation 2 | uniform on $x^2 + y^2 < 1$ & $x/y \in (-1, 0) \cup (1, \infty)$ | uniform on $x^2 + y^2 < 1$ & $x/y \in (-\infty, -1) \cup (0, 1)$ |
| Situation 3 | uniform on $((-3, -1) \cup (1, 3)) \times (0, 2)$ | uniform on $(-1, 1) \times (0, 6)$ |

3. Simulation Study

Although the computation of depth-based classification is slower comparing to the LDA in some situations it gives much better results. In what follows three different situations are considered to show strong and weak sides of the discussed classification procedures. The maximal depth classifier which uses the band depth is compared to the classifier which uses the halfspace depth and to the classical LDA.

In all examples two subpopulations \mathcal{P}_1 and \mathcal{P}_2 of a total population \mathcal{P} are considered. It is further considered that the sizes of the respective training sets are equal. First the training sets consisting of 100 observations are considered. Later the size of each training set is considered to be 1000.

Observations of zero depth are allocated randomly to one of the populations. This rule is very naïve, and the misclassification rates of depth-based classifications are therefore overestimated in what follows: in particular, the overestimation is high for the smaller training sets.

Let us now introduce three situations which are considered in the simulation study.

- (i) In the first simulation a mixture of two bivariate normal distributions of the same variance-covariance matrix and different means is considered. Therefore the LDA should be the best in this case.
- (ii) The second situation is taken from Example 2. It can be suggested from Figures 1 and 2 that the classifier based on the band depth is more promising than the other two classifiers.
- (iii) For the third simulation uniform distributions were chosen again. The distribution of population \mathcal{P}_1 is supported by two disjoint rectangles $(-3, -1) \times (0, 2)$ and $(1, 3) \times (0, 2)$. The mean value of distribution P_1 lies therefore outside the support. The support of distribution P_2 of population \mathcal{P}_2 is one rectangle $(-1, 1) \times (0, 6)$. Hence, support of P_2 contains both $\mu_1 = (0, 3)$, the mean value of P_1 , and $\mu_2 = (0, 1)$ the mean value of P_2 while the support of P_1 contains neither μ_1 nor μ_2 . The misclassification rates may therefore differ in both populations.

The formal description of the three different pairs of populations is given in Table 1.

In the next three paragraphs we discuss the results for the three considered classification problems. The summary of the simulation studies is presented in Table 2 and in Figure 3.

TABLE 2: Mean misclassification rates (in percent) for population 1 (upper) and population 2 (lower).

| | LDA | Halfspace classif. | Band depth classif. |
|-------------|------|--------------------|---------------------|
| Situation 1 | 22.6 | 24.3 | 26.5 |
| $N = 100$ | 21.9 | 24.9 | 27.3 |
| Situation 1 | 22.4 | 22.6 | 23.3 |
| $N = 1000$ | 22.7 | 22.9 | 23.1 |
| Situation 2 | 44.2 | 29.6 | 18.8 |
| $N = 100$ | 43.5 | 29.4 | 18.9 |
| Situation 2 | 47.5 | 28.4 | 7.6 |
| $N = 1000$ | 48.2 | 29.2 | 7.2 |
| Situation 3 | 0.1 | 6.0 | 12.0 |
| $N = 100$ | 22.2 | 30.7 | 10.9 |
| Situation 3 | 0.1 | 0.8 | 1.6 |
| $N = 1000$ | 23.3 | 28.7 | 3.1 |

Situation 1 (two normal distributions). The bandwidth c of weight function (6) is set to $c = 6$. This value was chosen ad hoc to show the difference between the band depth and the halfspace depth, since the optimal value is $c = \infty$ (the halfspace depth is proven to be optimal in this situation).

For small training sets the misclassification is in general higher for depth-based classification. This is partially given by relatively higher number of observations with zero depth than for large training sets. For larger training sets the misclassification rates are almost equal for the three methods.

The simulation has shown that the depth-based classifiers are comparable to optimal classification procedures (LDA) in standard situations.

Situation 2 (uniform distribution on circle). We have already stated that the band depth is expected to give the best results in this situation. Not surprisingly the misclassification rate of the LDA is close to $1/2$ with extremely high variability, and the variability is not decreasing with the size of training sets. This is mainly due to the fact that estimates of mean values and variance matrices of P_1 and P_2 , respectively, are very similar in most simulations.

The classification based on the band depth is superior to both LDA and the halfspace depth classification. The superiority is in particular clear for larger training sets (1000 points each), and the situation for smaller training sets is biased mainly due to chosen naïve allocation of points of zero depth, and the misclassification rate is highly overestimated in particular for the band depth.

Situation 3 (uniform distribution on disjoint rectangles). This is a very different situation since it is not symmetric from the subpopulations point of view. The convex hull of support of distribution P_1 contains one-third of the support of distribution P_2 of population \mathcal{P}_2 . On the other hand the (convex) support of distribution P_2 and support of distribution P_1 are disjoint. Therefore it may be expected that many points of population \mathcal{P}_2 will be allocated to population \mathcal{P}_1 while the opposite case will be quite rare. It is indeed true for the LDA and the halfspace depth classification.

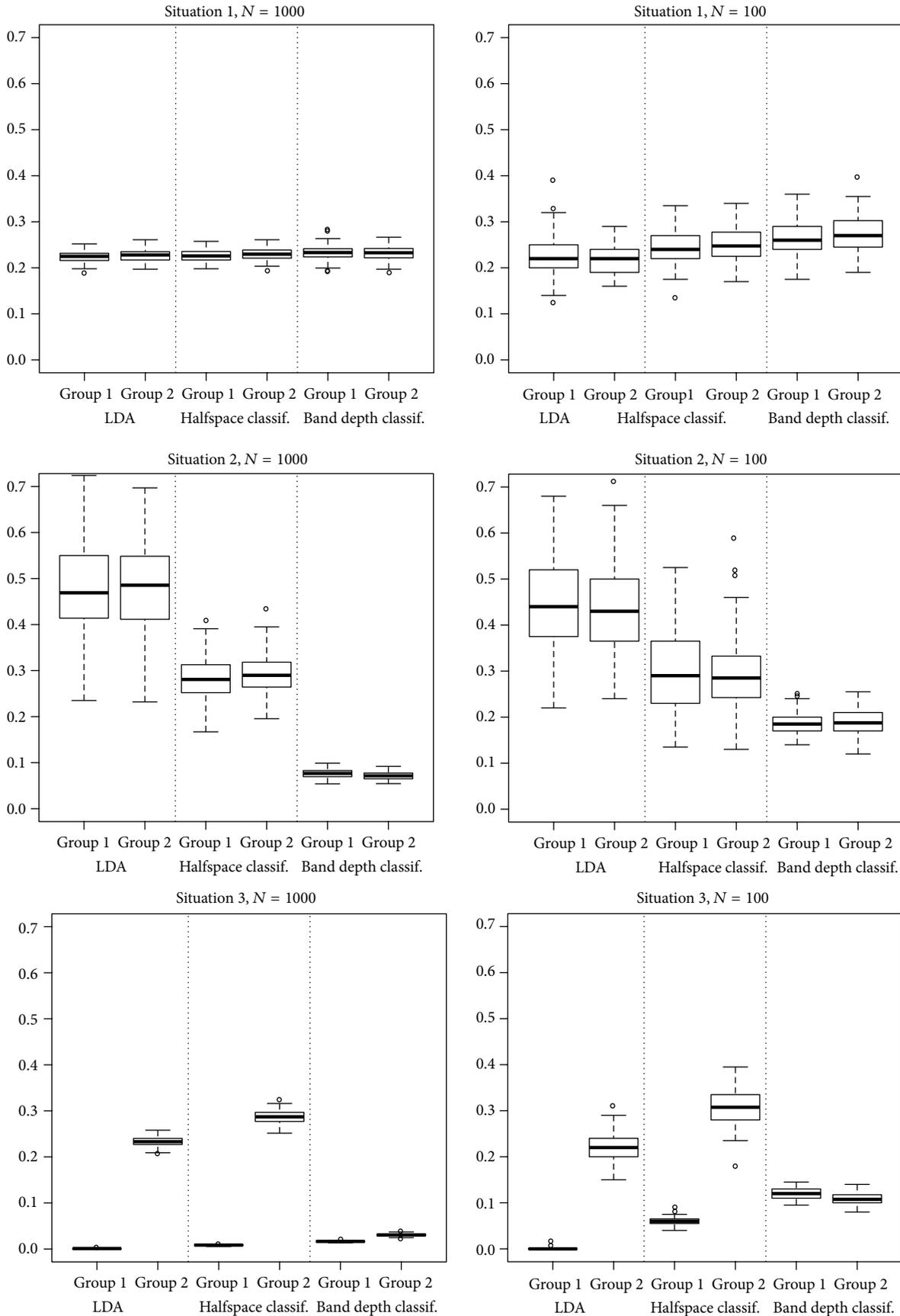


FIGURE 3: Misclassification rates for different situations and methods.

The classification based on the band depth has misclassification rate similarly small for both populations. The main reason is that the band depth of points *inside* the convex hull of support of \mathcal{P}_1 but *outside* the support itself is zero or very close to zero. For the halfspace depth almost 1/3 of points of population \mathcal{P}_2 has higher depth in population \mathcal{P}_1 , and this corresponds to the misclassification rate of the halfspace depth.

Conclusion. The misclassification rate of the LDA which is low for normal distributions may be very high for distributions with nonconvex supports (or for distributions with nonconvex levelsets of density). Nonparametric approach using the halfspace depth may be in such situation better (Situation 2) but may be worse (Situation 3).

In general it may be recommended to use the generalised halfspace depth instead of the halfspace depth if possible. The advantage of the generalised halfspace depth (simple band depth was used here) is shown on extreme Situations 2 and 3 above to emphasise its properties. Hence it may be concluded that the generalised halfspace depth is much better for distributions with nonconvex level sets. Situation 3 was chosen to show that for different types of distributions of populations \mathcal{P}_i the misclassification rate is similar for each population when the generalised halfspace depth is used while for the LDA and classification based on the halfspace depth the misclassification rates differ a lot between the considered populations.

The main disadvantage of the generalised halfspace depth is the need to choose the weight function and its respective parameter. It seems that band weight function (6) is good enough but it need not to be optimal, and even if we choose the band weight function the problem of bandwidth c remains.

Acknowledgment

The authors gratefully acknowledge the support by the Operational Program Education for Competitiveness — European Social Fund (Project CZ.1.07/2.3.00/20.0170 of the Ministry of Education, Youth and Sports of the Czech Republic).

References

- [1] R. Jörnsten, “Clustering and classification based on the L_1 data depth,” *Journal of Multivariate Analysis*, vol. 90, no. 1, pp. 67–89, 2004.
- [2] A. K. Ghosh and P. Chaudhuri, “On maximum depth and related classifiers,” *Scandinavian Journal of Statistics*, vol. 32, no. 2, pp. 327–350, 2005.
- [3] J. Tukey, “Mathematics and picturing data,” in *Proceedings of the 1975 International Congress of Mathematics*, vol. 2, pp. 523–531, August 1975.
- [4] R. Y. Liu, “On a notion of data depth based on random simplices,” *Annals of Statistics*, vol. 18, pp. 405–414, 1990.
- [5] G. Koshevoy and K. Mosler, “Zonoid trimming for multivariate distributions,” *Annals of Statistics*, vol. 25, no. 5, pp. 1998–2017, 1997.
- [6] Y. Vardi and C. Zhang, “The multivariate L_1 -median and associated data depth,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 4, pp. 1423–1426, 2000.
- [7] R. Y. Liu, R. Serfling, and D. L. Souvaine, Eds., *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, American Mathematical Society, Providence, RI, USA, 2006.
- [8] D. Hlubinka and S. Nagy, “Functional data depth and classification,” Submitted.
- [9] D. Hlubinka, L. Kotík, and O. Vencálek, “Weighted halfspace depth,” *Kybernetika*, vol. 46, no. 1, pp. 125–148, 2010.
- [10] A. Hartikainen and H. Oja, “On some parametric, nonparametric and semiparametric discrimination rules,” in *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, R. Liu, R. Serfling, and D. L. Souvaine, Eds., pp. 61–70, American Mathematical Society, Providence, RI, USA, 2006.
- [11] K. Mosler and R. Hoberg, “Data analysis and classification with the zonoid depth,” in *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, R. Liu, R. Serfling, and D. L. Souvaine, Eds., pp. 49–59, American Mathematical Society, Providence, RI, USA, 2006.
- [12] Y. Zuo and R. Serfling, “General notions of statistical depth function,” *Annals of Statistics*, vol. 28, no. 2, pp. 461–482, 2000.
- [13] B. Chakraborty and P. Chaudhuri, “On a transformation and retransformation technique for constructing an affine equivariant multivariate median,” *Proceedings of the American Mathematical Society*, vol. 124, no. 8, pp. 2539–2547, 1996.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

