

5-1-2017

A Comparison of Depth Functions in Maximal Depth Classification Rules


Olusola Samuel Makinde

Federal University of Technology, Akure, Nigeria, osmakinde@futa.edu.ng

Adeyinka Damilare Adewumi

Federal University of Technology Akure, Nigeria, adewumiadeyinkad@yahoo.com

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Makinde, O. S. & Adewumi, A. D. (2017). A comparison of depth functions in maximal depth classification rules. *Journal of Modern Applied Statistical Methods*, 16(1), 388-405. doi: 10.22237/jmasm/1493598120

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in *Journal of Modern Applied Statistical Methods* by an authorized editor of DigitalCommons@WayneState.

A Comparison of Depth Functions in Maximal Depth Classification Rules

Olusola Samuel Makinde
Federal University of Technology
Akure, Nigeria

Adeyinka Damilare Adewumi
Federal University of Technology
Akure, Nigeria

Data depth has been described as alternative to some parametric approaches in analyzing many multivariate data. Many depth functions have emerged over two decades and studied in literature. In this study, a nonparametric approach to classification based on notions of different data depth functions is considered and some properties of these methods are studied. The performance of different depth functions in maximal depth classifiers is investigated using simulation and real data with application to agricultural industry.

Keywords: classification rules, data depth, error rates, non-parametric approach, symmetry

Introduction

Classification is a practical subject in statistics. It aims at assigning an unclassified observation to one of several groups or populations on the basis of some measurement. Anderson (1984) described classification problem as a problem of statistical decision-making. However, classical multivariate analysis has relied heavily on the assumption of normality in data presentation and analysis. Among the classification methods that rely heavily on distribution assumption are Bayes rule (Welch, 1938), linear discriminant analysis and quadratic linear discriminant analysis (Anderson, 1984), and independence rule (Dudoit, Fridlyand & Speed, 2002). Research has shown that most of the data acquired nowadays do not satisfy normality assumption. Similarly, some parametric approaches are prone to the effect of outlying observations. This gives nonparametric approach to classification an edge over parametric methods.

Olusola Samuel Makinde is a Lecturer in the Department of Statistics. Email at osmakinde@futa.edu.ng. Adeyinka Damilare Adewumi is a graduate of the Department of Statistics. Email at adewumiadewinkad@yahoo.com.

Other methods in literature include support vector machine (Vapnik, 1998; Cortes & Vapnik, 1995), nearest neighbour rule (Cover & Hart, 1967), classification rules based on distance functions (Chan & Hall, 2009; Hall, Titterington & Xue, 2009), classifiers based on distribution functions of rank outlyingness (Makinde & Chakraborty, 2015).

Data depth is a way to measure the depth or outlyingness of a given point with respect to a multivariate data cloud or its underlying distribution (Liu, Singh & Parelius, 1999). It gives rise to a natural centr-outward ordering of the sample points in \mathbb{R}^d . This ordering gives rise to new and easy ways to quantify many complex multivariate features of the underlying distribution, including location, quantiles, scale, skewness and kurtosis. Liu (1990) introduced a notion of simplicial depth and corresponding estimators of location, and formulated a quality index with simplicial depth, Mahalanobis depth and majority depth. Koshevoy & Mosler (1997) introduced a notion of zonoid depth while Fraiman, Meloche & García-Escudero (1999) introduced a likelihood type depth function. Rousseeuw & Hubert (1999) introduced a notion of regression depth. Liu, Singh & Parelius (1999) considered some examples of depth functions and developed methodology for their practical applications.

Classification rule based on data depth is considered in the current study. Data depth is formally defined based on Zuo & Serfling (2000a) and examples of depth functions are presented. In reality, an important question that arises in almost all fields where supervised learning is employed is that which of the depth functions should be employed. Classification rules based on the depth functions are defined and properties of the classification rules are presented. Evaluation of the classification rule, accounting for performance of various depth functions are presented based on numerical examples.

Notions of Statistical Depth Functions

Definition 1 (Zuo & Serfling, 2000a). Let the mapping $D(.,.) : \mathbb{R}^d \times \mathcal{F} \rightarrow \mathbb{R}$ be bounded and non-negative, and satisfy:

- i. $D(\mathbf{A}\mathbf{x} + \mathbf{b}, F_{\mathbf{A}\mathbf{x}+\mathbf{b}}) = D(\mathbf{x}, F_{\mathbf{x}})$ holds for any random vector $\mathbf{X} \in \mathbb{R}^d$ and any $d \times d$ nonsingular matrix \mathbf{A} , and any d dimensional vector \mathbf{b} .
- ii. $D(\boldsymbol{\theta}, F) = \sup_{\mathbf{x} \in \mathbb{R}^d} D(\mathbf{x}, F)$ holds for any $F \in \mathcal{F}$ having centre $\boldsymbol{\theta}$.
- iii. For any $F \in \mathcal{F}$ having deepest point $\boldsymbol{\theta}$, $D(\mathbf{x}, F) \leq D(\boldsymbol{\theta} + \alpha(\mathbf{x} - \boldsymbol{\theta}), F)$ holds for $\alpha \in [0, 1]$; and

DEPTH FUNCTIONS IN MAXIMAL DEPTH CLASSIFICATION RULES

- iv. $D(\mathbf{x}, F) \rightarrow 0$ as $\|\mathbf{x}\| \rightarrow \infty$, for each $F \in \mathcal{F}$.

Then $D(\cdot, F)$ is called a statistical depth function.

From **Definition 1**, the first property describes invariance of depth function under general affine transformation of the data. That is, the depth of any observation in \mathbb{R}^d should not depend on the scale of the underlying measurement or underlying coordinate system. The second property implies that depth value attains its maximum value at the point of symmetry for symmetric distributions. The third property implies that the depth value decreases monotonically as vector \mathbf{x} moves away from its most central point while the fourth property implies that the depth value of \mathbf{x} vanishes (tend to zero) as Euclidean norm of \mathbf{x} approaches infinity.

The depth functions in literature include

1. **Mahalanobis Depth (MhD)**. Mahalanobis (1936); Liu & Singh (1993) defined the depth of an observation \mathbf{x} with respect to the distribution F as

$$MhD(\mathbf{x}, F) = [1 + O(\mathbf{x}, \boldsymbol{\mu}_F, \boldsymbol{\Sigma}_F)]^{-1}$$

where $O(\mathbf{x}, \boldsymbol{\mu}_F, \boldsymbol{\Sigma}_F) = (\mathbf{x} - \boldsymbol{\mu}_F)' \boldsymbol{\Sigma}_F^{-1} (\mathbf{x} - \boldsymbol{\mu}_F)$, $\boldsymbol{\mu}_F$ and $\boldsymbol{\Sigma}_F$ are the mean vector and dispersion matrix of F respectively. The sample version of *MhD* is obtained by replacing $\boldsymbol{\mu}_F$ and $\boldsymbol{\Sigma}_F$ with their estimates.

2. **Zonoid Depth (ZD)**. Dyckerhoff et al. (1996) defined a zonoid depth as

$$ZD(\mathbf{x}, F) = \sup\{\alpha : \mathbf{x} \in D_\alpha(\mathbf{X}_1, \dots, \mathbf{X}_n)\}$$

where $D_\alpha(\mathbf{X}_1, \dots, \mathbf{X}_n) = \sum_{i=1}^n \lambda_i \mathbf{X}_i$, $\sum_{i=1}^n \lambda_i = 1$, $\lambda_i \geq 0$, and $\alpha \lambda_i \leq \frac{1}{n}$ for all i .

3. **Half-Space Depth (HD)**. Tukey (1975) defined half-space depth of a point $\mathbf{x} \in \mathbb{R}^d$ with respect to F as the minimum probability mass carried by any closed half-space containing \mathbf{x} . Mathematically,

$$HD(\mathbf{x}, P) = \inf_H [P(H)]$$

where H is a closed halfspace in \mathbb{R}^d and $\mathbf{x} \sim H$.

4. **Oja Depth (OD).** Oja (1983) defined the depth of $\mathbf{x} \in \mathbb{R}^d$ with respect to F as

$$OD(F;\mathbf{x}) = [1 + O(\mathbf{x},F)]^{-1}$$

where $O(\mathbf{x},F) = E_F(\text{Volume}(S[\mathbf{x}, \mathbf{X}_1, \dots, \mathbf{X}_d]))$, $S[\mathbf{x}, \mathbf{X}_1, \dots, \mathbf{X}_d]$ is a closed simplex with vertices \mathbf{x} and d random observations $\mathbf{X}_1, \dots, \mathbf{X}_d$ from F .

5. **Simplicial Depth (SD).** Liu (1990) defined simplicial depth of $\mathbf{x} \in \mathbb{R}^d$ with respect to F as

$$SD(F;\mathbf{x}) = P(\mathbf{x} \in S[\mathbf{x}, \mathbf{X}_1, \dots, \mathbf{X}_{d+1}])$$

where $S[\mathbf{x}, \mathbf{X}_1, \dots, \mathbf{X}_{d+1}]$ is a closed simplex formed by $(d+1)$ random observation from F . The sample version of $SD(F;\mathbf{x})$ is obtained by replacing F in $SD(F;\mathbf{x})$ by F_n .

6. **Projection Depth (PD).** Donoho & Gasko (1982) defined the depth of \mathbf{x} with respect to F as the worst case outlyingness of \mathbf{x} with respect to one dimensional median in any one-dimensional projection.

$$PD(F;\mathbf{x}) = (1 + O(\mathbf{x},F))^{-1}$$

where $O(\mathbf{x},F) = \sup_{\|\mathbf{u}\|=1} \frac{\mathbf{u}'\mathbf{x} - \text{Med}(F_u)}{MAD(F_u)}$, F_u is the distribution $\mathbf{u}'\mathbf{X}$, $\text{Med}(F_u)$ is the median of F_u , $MAD(F_u)$ is the median absolute deviation of F_u and $X \sim F$. The sample version of $PD(F;\mathbf{x})$ is obtained by replacing the median and MAD with their sample estimates.

7. **Likelihood Depth (LD).** Fraiman, Meloche & García-Escudero (1999) defined the depth of \mathbf{x} with respect to F simply as its probability density, that is, $LD(F;\mathbf{x}) = f(\mathbf{x})$, and the empirical version

can be any consistent density estimate at \mathbf{x} , for example, the kernel density estimate.

8. **Spatial Depth (SPD).** Serfling (2002) defined spatial depth of any observation \mathbf{x} with respect to F as

$$SBD(\mathbf{x}, F) = 1 - \left\| E_F \left(\frac{\mathbf{x} - \mathbf{X}}{\|\mathbf{x} - \mathbf{X}\|} \right) \right\|$$

where $X \sim F$.

9. **Simplicial volume depth (SVD).** Zuo & Serfling (2000a, b) expressed SVD of an observation \mathbf{x} with respect to F as

$$SVD(\mathbf{x}, \delta, F) = \left\{ 1 + E_F \left(\frac{\nabla(\mathbf{x}, \mathbf{X}_1, \dots, \mathbf{X}_d)}{|\Sigma_F|^{\frac{1}{2}}} \right)^\delta \right\}^{-1}$$

where $\mathbf{X}_1, \dots, \mathbf{X}_d$ are independent and identically distributed observations from F , $\nabla(\mathbf{x}, \mathbf{X}_1, \dots, \mathbf{X}_d)$ is the volume of the d -dimensional simplex formed by \mathbf{x} and Σ_F is the scatter matrix of the distribution F .

10. **Majority Depth (MJD).** Liu & Singh (1993) defined the depth of \mathbf{x} with respect to F as the probability that \mathbf{x} belongs to the major side (i.e. the half-space with the larger probability measure) of a random hyperplane passing through the data points in \mathbb{R}^d .

Other depth functions include regression depth (Rousseeuw & Hubert, 1999). Gao (2003) defined another depth function based on square of spatial outlyingness function. Few of these depth functions satisfy all the four properties in Definition 1 while others satisfy some of the properties. See Zuo & Serfling (2000a; 2000b) for detail.

Classification Rule

The goal of any classification rule is to find a rule or tool that enables us to assign an observation $\mathbf{x} \in \mathbb{R}^d$ to one of the several competing groups (or classes). One can define a classification rule based on depth functions. It is easy to observe that data depth gives an idea on how outlying an observation \mathbf{x} is with respect to the distribution F . If \mathbf{x} is a central observation, its depth value will be large. On the other hand, if \mathbf{x} is an extreme observation, its depth value will be small. Thus a small depth value may suggest a deviation of \mathbf{x} from F .

Ghosh & Chaudhuri (2005) proposed a classification rule based on simple idea of assigning a new observation to any of the J competing classes, for which it attains maximal depth value. It is expressed as:

$$D(F_k, \mathbf{x}) = \arg \max_{1 \leq j \leq J} D(F_j, \mathbf{x}) \tag{1}$$

where F_k is the distribution of k^{th} class and $1 \leq j \leq J$.

Let us consider two classes for simplicity. Suppose π_j has multivariate distribution with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$, $j = 1, 2$. For $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, the classification rule in (1) can be expressed as

$$\text{Assign } \mathbf{x} \text{ to } F \text{ if } D(F, \mathbf{x}) > D(G, \mathbf{x}), \text{ and to } G \text{ if otherwise.} \tag{2}$$

It is straightforward to show that a depth function can be expressed in terms of probability density function of the competing distribution. This result is presented by a Lemma below:

Lemma 1. Let F_j be spherically symmetric distributions with density functions of the form

$$f_j(\mathbf{x}) = |\boldsymbol{\Sigma}_j|^{-\frac{1}{2}} h(\mathbf{x} - \boldsymbol{\theta}_j)' (\mathbf{x} - \boldsymbol{\theta}_j)$$

for some strictly decreasing, continuous, non-negative scalar function h . Then for any of the depth functions OD and SPD,

$$f_j(\mathbf{x}) = \omega(D(F_j, \mathbf{x}))$$

DEPTH FUNCTIONS IN MAXIMAL DEPTH CLASSIFICATION RULES

for some increasing function ω .

Suppose a random vector \mathbf{X} in \mathbb{R}^d is elliptically distributed such that its density is of the form $f(\mathbf{x}) = |\Sigma|^{-1/2} h((\mathbf{x} - \boldsymbol{\theta})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\theta}))$, then $D(F, \mathbf{x})$ can be expressed as a function of $(\mathbf{x} - \boldsymbol{\theta})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\theta})$. This result is presented formally by a Lemma below:

Lemma 2. Let F_j be elliptically symmetric distributions with density functions of the form

$$f_j(\mathbf{x}) = |\Sigma_j|^{-1/2} h\left((\mathbf{x} - \boldsymbol{\theta}_j)' \Sigma_j^{-1} (\mathbf{x} - \boldsymbol{\theta}_j)\right)$$

for some strictly decreasing, continuous, non-negative scalar function h . Then for any of the depth functions detailed earlier, except OD and SPD,

$$f_j(\mathbf{x}) = \omega\left(D(F_j, \mathbf{x})\right),$$

where Σ_j is not a constant multiple of identity matrix for some increasing function ω .

The optimal rule, Bayes rule, assigns an observation to the class or distribution with highest posterior probability. That is, assign \mathbf{x} to j^{th} class if $p_j f_j(\mathbf{x})$ is the highest, where p_j is the prior probability of the j^{th} class. Based on the results of Lemmas 1 and 2, it is straightforward to show that maximum depth classifiers are Bayes rules under necessary conditions.

Theorem 1. Suppose the conditions of Lemmas 1 and 2 hold on all the depth functions defined earlier. Then the classifier defined in (1) is Bayes rule if competing distributions have equal covariance matrices and prior probabilities.

In practice, a depth function may not be completely known and so need to be estimated based on sample and then define the empirical version of the classification rule based on the empirical depth function. The empirical depth function based on sample is denoted by $D(F_n, \mathbf{x})$. To show the consistency of empirical depth functions, it is desirable to establish the almost sure convergence of empirical depth functions to its population counterpart.

Theorem 2. Suppose $D(F_n, \mathbf{x})$ is an empirical depth function based on $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$. Let $D(F, \mathbf{x})$ be a population depth function of any random vector \mathbf{x} . Then for any \mathbf{x} in the support of F ,

$$\sup_{\mathbf{x}} |D(F_n, \mathbf{x}) - D(F, \mathbf{x})| \rightarrow 0, \quad n \rightarrow \infty.$$

The almost sure convergence of half-space depth has been established in Donoho & Gasko (1992), simplicial depth in Liu (1990). Liu & Singh (1993) has shown almost sure convergence of Mahalanobis depth and majority depth while Zuo & Serfling (2000b) proved convergence of projection depth. The almost sure convergence of spatial depth follows from Koltchinskii's (1997) work on the convergence of the empirical spatial rank function to its population version. Convergence of the empirical classification rule to population version follows from Theorem 2.

Evaluation of Classification methods

One way of evaluating the performance of a classifier is to compute its associated misclassification probability. In a two class classification problem, one can define a misclassification probability as

$$\Delta = p_1 P(D(F, \mathbf{x}) < D(G, \mathbf{x}) | \mathbf{x} \in F) + p_2 P(D(F, \mathbf{x}) > D(G, \mathbf{x}) | \mathbf{x} \in G)$$

The empirical version of the probability of misclassification or error rate, denoted by $\hat{\Delta}$, can be defined as

$$\begin{aligned} \hat{\Delta} = & \frac{p_1}{n} \sum_{i=1}^n I \left\{ D(\hat{F}, \mathbf{x}_i) < D(\hat{G}, \mathbf{x}_i) | \mathbf{x} \in F \right\} \\ & + \frac{p_2}{m} \sum_{i=1}^m I \left\{ D(\hat{F}, \mathbf{x}_i) > D(\hat{G}, \mathbf{x}_i) | \mathbf{x} \in F \right\} \end{aligned}$$

Under the conditions of Theorems 1 and 2, it is straightforward to show that $\hat{\Delta}$ is a Bayes risk.

Simulation Study

As illustration of the performance of maximum depth classification methods, consider the following example. Let populations π_1 and π_2 be bivariate spherically symmetric with centres of symmetry $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and covariance matrices, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, respectively. Assume that the prior probabilities of π_1 and π_2 are equal. Suppose $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ is a random sample from π_1 and $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m$, a random sample from π_2 . New random vectors $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_m$ from π_1 and $\mathbf{Z}_{m+1}, \mathbf{Z}_{m+2}, \dots, \mathbf{Z}_{2m}$ from π_2 are generated and sample sizes n and m are taken to be 100. $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are chosen to be $(0 \ 0)^T$ and $(\delta \ 0)^T$ respectively for $\delta \in [-2, 2]$ and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \mathbf{I}_2$. The simulation size is taken to be 1000. Different depth functions are considered for some competing distributions. The distributions are bivariate normal distributions and bivariate Laplace distributions. For computation of likelihood depth, Gaussian kernel is used with turning parameter ($=0.3$). *R* Package `fda.usc` is used for computing projection and likelihood depth. *R* Package `depth` is used for computing Oja depth, simplicial depth and half-space depth while *R* Package `dda1pha` is used for computing simplicial volume depth, Mahalanobis depth and Zonoid depth.

Estimates of misclassification probabilities are less in bivariate normally distributed samples than bivariate Laplace samples, as shown in Figure 1. It is observed from Figures 2 and 3 that maximal depth classification rule based on half-space depth outperforms others when the distinction between competing distributions is not wide. That is, when $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \rightarrow \mathbf{0}$. The distinction between competing distributions becomes clear as $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ moves away from $\mathbf{0}$ and the performance of various depth functions becomes equivalent. It is noted that exact computation of half-space depth and simplicial depth functions is feasible only in \mathbb{R}^3 and \mathbb{R}^2 respectively. Cuesta-Albertos & Nieto-Reyes (2010) suggested a modified version of half-space depth for functional data, as extension of multivariate set-up. The performance of empirical likelihood depth based on kernel estimator of probability density function depends on the choice of kernel function and turning parameter. It is observed that spatial depth and Oja depth are not invariant under general affine transformation. Makinde (2017) considered various affine invariant versions of spatial rank, a related notion to spatial depth. Robustness of spatial rank (a straightforward extension of spatial depth) against deviation from notion of elliptical symmetry is demonstrated in Makinde and Chakraborty (2015).

Maximum depth classification rule is compared with some classification methods, which include linear discriminant analysis (LDA), k -nearest neighbor

rule (kNN) and support vector machine (SVM); using the above setting for $\delta = 1, 2$. Table 1 below presents performance of classifiers. It is observed from the table that maximum depth classifiers based on half-space depth has the best performance among the depth based procedures, linear discriminant analysis, k -nearest neighbor rule and support vector machine. It has the least mean error rates when the competing distributions are normal and Laplace. Next to half-space depth among the depth functions for maximum depth classification rule is zonoid depth.

However, zonoid depth is not robust against outlying observations in the data cloud. LDA performs well compared with kNN and SVM. It is noted that linear discriminant analysis is Bayes (optimal) rule when competing distributions are multivariate normal. Hence maximum depth classifiers based on half-space depth is a better alternative to the known parametric classification methods, e.g. LDA.

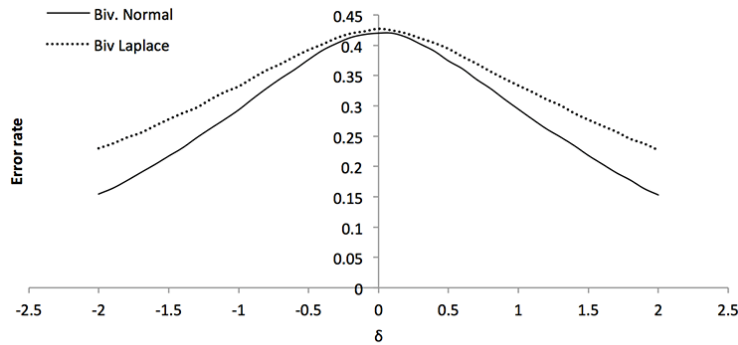


Figure 1. Comparison of error rates associated with half space depth for normally distributed samples and Laplace distributed samples.

DEPTH FUNCTIONS IN MAXIMAL DEPTH CLASSIFICATION RULES

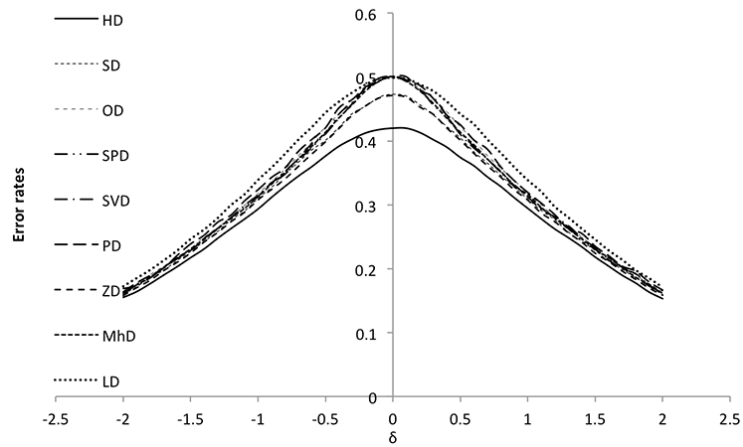


Figure 2. Comparison of depth functions in classification based on error rates for normally distributed samples

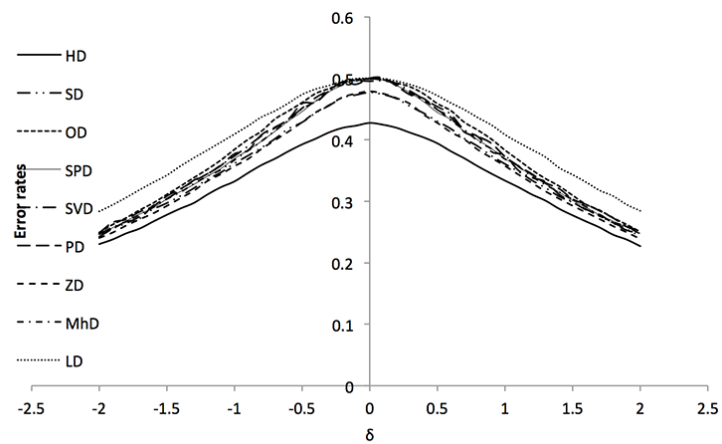


Figure 3. Comparison of depth functions in classification based on error rates Laplace distributed samples.

Table 1. Comparison of mean error rates of classifiers when competing distributions differ in location.

Distribution	δ	<i>Maximal Depth Classifiers</i>											
		HD	SD	OD	SPD	SVD	PD	ZD	MhD	LD	LDA	kNN	SVM
Biv normal	1	0.295	0.309	0.318	0.313	0.313	0.319	0.307	0.316	0.334	0.315	0.356	0.316
	2	0.153	0.160	0.165	0.162	0.161	0.167	0.159	0.161	0.167	0.161	0.181	0.167
Biv Laplace	1	0.334	0.361	0.383	0.369	0.375	0.373	0.357	0.369	0.413	0.377	0.410	0.381
	2	0.227	0.243	0.250	0.248	0.248	0.253	0.239	0.246	0.287	0.246	0.273	0.257

Table 2. Comparison of mean error rates of classifiers when competing distributions differ in location and scale.

	<i>Maximal Depth Classifiers</i>											QDA	kNN	SVM
	HD	SD	OD	SPD	SVD	PD	ZD	MhD	LD					
Biv normal	0.382	0.386	0.500	0.387	0.389	0.386	0.389	0.389	0.166	0.142	0.209	0.148		
Biv Laplace	0.410	0.417	0.500	0.418	0.421	0.419	0.421	0.418	0.255	0.214	0.282	0.214		

Table 3. Comparison of computation time of classifiers for bivariate Laplace distributions.

	<i>Maximal Depth Classifiers</i>											QDA	kNN	SVM
	HD	SD	OD	SPD	SVD	PD	ZD	MhD	LD					
Time (seconds)	0.12	0.12	0.14	0.52	15.84	5.97	0.34	0.32	1.39	0.08	0.05	0.31		

Only populations which are separated by location are considered so far. Table 2 presents a comparison of proportions of misclassification of depth based procedures, quadratic discriminant analysis (kNN) and SVM when competing populations have different location vectors and covariance matrices.

Suppose the mean vectors and covariance matrices of π_1 and π_2 are $(\mu_1 = (0 \ 0)^T, \Sigma_1 = \mathbf{I}_2)$ and $(\mu_2 = (2 \ 0)^T, \Sigma_2 = 9\mathbf{I}_2)$, respectively. It is well known that QDA is an optimal rule when competing populations are normally distributed and differ in location and scale. Hence it has a least mean error rate (= 0.142) for normal distributions. Maximum depth classifier based on likelihood depth has the least mean error rate (= 0.166) among the depth classifiers, which is competitive with QDA and SVM (with mean error rate = 0.148). Maximum depth classifier based on Oja depth has the worst performance in this case. For bivariate Laplace distributions, Maximum depth classifier based on likelihood depth has the least mean error rate (= 0.255) among the depth classifiers, which is competitive with QDA (with mean error rate = 0.214), SVM (with mean error rate = 0.214) and kNN (with mean error rate = 0.282). Mean error rates of other depth classifiers are a bit high.

DEPTH FUNCTIONS IN MAXIMAL DEPTH CLASSIFICATION RULES

Presented in Table 3 is a comparison of computation time in seconds of each classifier when competing distributions are bivariate Laplace for one repetition. It is shown in Table 3 that QDA and kNN have the least computation time. However, computation time of maximum depth classifiers based on half-space depth, simplicial depth, Oja depth, zonoid depth and spatial depth are competitive with those of parametric classifiers.

Analysis of Real Data

A real dataset is also analysed to illustrate the performances of depth functions in maximal depth classification methods. Maximal depth classifiers are applied on mineral ions variability data. The data was extracted from a project experiment on crop science and production at the Institute for Agricultural Research and Training (IAR&T) project titled “inter- and intra-maturity group differences in physiological quality of maize seeds” (Olasoji, 2014). The data contains measurements of mean amount of mineral ions (Na, Ca, K and P) leaked after 24 hours from soaked maize seeds at different maturity groups (early, intermediate and lately). Each observation consists of four attributes, which are mean mineral ions (Na, Ca, K and P). Each group consists of 36 observations. A random sample of size 30 and a test sample of size 6 are chosen. The experiment is repeated 100 times; quantile, mean and standard deviation of the proportions of misclassification associated with each of the classifiers are computed.

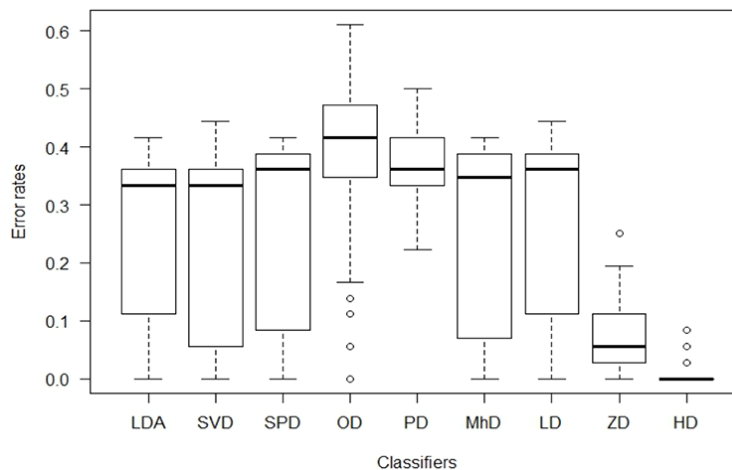


Figure 4. Box plot of proportions of misclassification associated with some classifiers for real data example

Table 4. Quantiles, means and standard deviations of proportions of misclassification of some classifiers for real data example.

	<i>Maximal Depth Classifiers</i>								
	HD	OD	SPD	SVD	PD	ZD	MhD	LD	LDA
Minimum	0.0000	0.0000	0.0000	0.0000	0.2222	0.0000	0.0000	0.0000	0.0000
25% Quantile	0.0000	0.3541	0.0833	0.0556	0.3333	0.0278	0.0764	0.1111	0.1111
Mean	0.0069	0.3861	0.2697	0.2364	0.3717	0.0725	0.2683	0.2767	0.2619
Median	0.0000	0.4167	0.3611	0.3333	0.3611	0.0556	0.3472	0.3611	0.3333
75% Quantile	0.0000	0.4722	0.3889	0.3611	0.4167	0.1111	0.3889	0.3889	0.3611
Maximum	0.0556	0.6111	0.4444	0.4444	0.5000	0.2500	0.4167	0.4167	0.4167
Standard deviation	0.0150	0.1478	0.1575	0.1618	0.0623	0.0600	0.1533	0.1462	0.1472

Presented in Figure 4 is a comparison of maximum depth classifiers with linear discriminant analysis based on the proportions of misclassification using box plot. The figure shows that the maximum depth classifiers based on half-space depth and zonoid depth has the least proportions of misclassification while the maximum depth classifiers based on Oja depth and projection depth has highest proportions of misclassification.

Presented in Table 4 is the quantile, mean and standard deviation of the proportions of misclassification associated with each of the competing classifiers. Maximum depth classifier based on half-space depth has the least mean proportion of misclassification as shown in the table. Use of spatial depth, simplicial volume depth, Mahalanobis depth and likelihood depth in maximum depth classifiers perform equivalently to LDA, while maximum depth classifiers based on half-space depth and zonoid depth outperform LDA. Simplicial depth values could not be computed as $d = 4 > 2$. For computation of half-space depth, an approximate algorithm implemented in R Package depth is used.

Conclusion

The maximum depth classifiers based on the training samples when any of the half-space depth, projection depth, simplicial depth, spatial depth, Oja depth, and majority depth is used, do not depend on any distributional assumptions or do not require any estimation of model parameters. That gives maximum depth classifiers an importance over parametric methods. Maximum depth classifiers are easily lent to multiclass cases. We have noted in our real data examples that the maximum depth classifiers are quite competitive with similar classifiers, especially when any of half-space or zonoid depth is used.

Acknowledgment

The authors will like to thank Dr. J. Olasoji of the Institute for Agricultural Research and Training (IAR&T), Ibadan for making mineral ions variability data available.

References

- Anderson, T. W. (1984). *An introduction to multivariate statistical analysis*. NY: John Wiley & Sons, Inc.
- Chan, Y. & Hall, P. (2009). Scale adjustments for classifiers in high-dimensional, low sample size settings. *Biometrika*, 96(2), 469 – 478. doi: [10.1093/biomet/asp007](https://doi.org/10.1093/biomet/asp007)
- Cortes, C. & Vapnik, V. N. (1995). Support-vector networks. *Machine Learning*, 20(3), 273 – 297. doi: [10.1007/bf00994018](https://doi.org/10.1007/bf00994018)
- Cover, T. M. & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21 – 27. doi: [10.1109/tit.1967.1053964](https://doi.org/10.1109/tit.1967.1053964)
- Cuesta-Albertos, J. A. & Nieto-Reyes, A. (2010). Functional classification and the random Tukey depth. Practical issues. *Advances in Intelligent and Soft Computing*, 77, 123 – 130. doi: [10.1007/978-3-642-14746-3_16](https://doi.org/10.1007/978-3-642-14746-3_16)
- Donoho, D. L. & Gasko, M. (1992). Breakdown properties of multivariate location parameters and dispersion matrices. *The Annals of Statistics*, 20(4), 1803 – 1827. doi: [10.1214/aos/1176348890](https://doi.org/10.1214/aos/1176348890)
- Dudoit S., Fridlyand, J. & Speed, T. P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, 97(457), 77 – 87. doi: [10.1198/016214502753479248](https://doi.org/10.1198/016214502753479248)
- Dutta, S. & Ghosh, A. K. (2012). *On classification based on L_p depth with an adaptive choice of p* . Technical Report No. R5/2011, Statistics and Mathematics Unit. Indian Statistical Institute, Kolkata, India
- Dyckerhoff, R., Mosler, K. & Koshevoy, G. (1996). Zonoid Data Depth: Theory and Computation. In: Prat, A. (Ed.) *COMPSTAT: Proceedings in Computational Statistics 12th Symposium held in Barcelona, Spain, 1996*. Berlin: Physica-Verlag HD. doi: [10.1007/978-3-642-46992-3_26](https://doi.org/10.1007/978-3-642-46992-3_26)

- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2), 179 – 188. doi: 10.1111/j.1469-1809.1936.tb02137.x
- Fraiman, R., Meloche, J., García-Escudero, L. A., et al. (1999). Multivariate L-estimation. *Test*, 8(2), pp. 255 – 317. doi: 10.1007/bf02595872
- Gao, Y. (2003). Data depth based on spatial rank. *Statistics & Probability Letters*, 65(3), 217 – 225. doi: 10.1016/j.spl.2003.06.003
- Ghosh, A. K. & Chaudhuri, P. (2005). On maximum depth and related classifiers. *Scandinavian Journal of Statistics*, 32(2), 327 – 350. doi: 10.1111/j.1467-9469.2005.00423.x
- Hall, P., Titterton, D.M. & Xue, J. (2009). Median based classifiers for high dimensional data. *Journal of the American Statistical Association*. 104(488), 1597 – 1608. doi: 10.1198/jasa.2009.tm08107
- Hubert, M. & Van Driessen, K. (2004). Fast and robust discriminant analysis. *Computational Statistics and Data Analysis*, 45(2), 301 – 320. doi: 10.1016/s0167-9473(02)00299-2
- Koltchinskii, V. I. (1997). M-estimation, convexity and quantiles. *The Annals of Statistics*, 25(2), 435 – 477. doi: 10.1214/aos/1031833659
- Koshevoy, G. & Mosler, K. (1997). Zonoid trimming for multivariate distributions. *The Annals of Statistics*, 25(5), 1998 – 2017. doi: 10.1214/aos/1069362382
- Li, J., Cuesta-Alberstos, J. A. & Liu, R. Y. (2012). DD-classifier: nonparametric classification procedure based on DD-plot. *Journal of the American Statistical Association*, 107(498), 737 – 753. doi: 10.1080/01621459.2012.688462
- Liu, R. Y. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics*, 18(1), 405 – 414. doi: 10.1214/aos/1176347507
- Liu, R. Y. & Singh, K. (1993). A quality index based on multivariate data depth and multivariate rank tests. *Journal of the American Statistical Association*, 88(421), 252 – 260. doi: 10.1080/01621459.1993.10594317
- Liu, R. Y., Parelius, J. M. & Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *The Annals of Statistics*, 27(3), 783 – 858. doi: 10.1214/aos/1018031260
- Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2(1), 49–55.

- Makinde, O. S. (2015). *On some classification methods for high dimensional and functional data*. PhD Thesis, University of Birmingham
- Makinde, O. S. (2017). Multivariate rank outlyingness and correlation effects. *Journal of Modern Applied Statistical Methods*, 16(1), 246-260. doi: 10.22237/jmasm/1493597580
- Makinde, O. S. & Chakraborty, B. (2015). On some classifiers based on distribution functions of multivariate ranks. In Nordhausen, K and Taskinen, S. (Eds). *Modern Nonparametric, Robust and Multivariate Methods, Festschrift in Honour of Hannu Oja*. NY: Springer, 249 – 264. doi: 10.1007/978-3-319-22404-6_15
- Oja, E. (1983). *Subspace methods of pattern recognition*. Letchworth, England: Research Studies Press.
- Olasoji, J. (2014). Inter- and intra-maturity group differences in physiological quality of maize seeds. Unpublished raw data.
- Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, 79(388), 871 – 880. doi: 10.1080/01621459.1984.10477105
- Rousseeuw, P. J. & Hubert, M. (1999). Regression depth (with discussion). *Journal of the American Statistical Association*, 94(446), 388 – 433. doi: 10.1080/01621459.1999.10474129
- Rousseeuw, P. J. & Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3), 212 – 223. doi: 10.1080/00401706.1999.10485670
- Serfling, R. (2002). A depth function and a scale curve based on spatial quantiles. In Y. Dodge (Ed.). *Statistical Data Analysis Based on the L1-Norm and Related Methods*, pp. 25–38. Basel: Birkhauser. doi: 10.1007/978-3-0348-8201-9_3
- Tukey, J. W. (1975). Mathematics and the picturing of data. Proceeding of the International Congress of Mathematicians, Vancouver, 523–531.
- Vapnik, V. N. (1998). *Statistical learning theory*. NY: John Wiley and Sons
- Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29(3-4), 350 – 362. doi: 10.2307/2332010
- Zuo, Y. & Serfling, R. (2000a). General notion of statistical depth function. *The Annals of Statistics*, 28(2), 461 – 482. doi: 10.1214/aos/1016218226

Zuo, Y. & Serfling, R. (2000b). Structural properties and convergence results for contours of sample statistical depth function. *The Annals of Statistics*, 28(2), 483 – 499. doi: 10.1214/aos/1016218227