Ondřej Vencálek

k-Depth-nearest Neighbour Method and its Performance on Skew-normal Distributons

Acta Universitatis Palackianae Olomucensis. Facultas Rerum Naturalium. Mathematica, Vol. 52 (2013), No. 2, 121--129

Persistent URL: http://dml.cz/dmlcz/143544

Acta Univ. Palacki. Olomuc., Fac. rer. nat.,
Mathematica **52**, 2 (2013) 121–129

k-Depth-nearest Neighbour Method and its Performance on Skew-normal Distributons*

Ondřej VENCÁLEK

Department of Mathematical Analysis and Applications of Mathematics
Faculty of Science, Palacký University
17. listopadu 12, 771 46 Olomouc, Czech Republic
e-mail: ondrej.vencalek@upol.cz

(Received June 25, 2013)

Abstract

In the present paper we investigate performance of the k-depth-nearest classifier. This classifier, proposed recently by Vencálek, uses the concept of data depth to improve the classification method known as the k-nearest neighbour. Simulation study which is presented here deals with the two-class classification problem in which the considered distributions belong to the family of skewed normal distributions.

Key words: data depth, classification, k-nearest neighbour, skewed normal distribution

2010 Mathematics Subject Classification: 62G30, 62H30

1 Introduction

Concept of data depth provides one possible approach to a nonparametric analysis of multivariate data. The concept and its possible aplications was discussed in detail at the ODAM 2011 conference followed by the paper [7] which provides a review of possible applications of the data depth concept. Let us recall that a depth function is basically any function which provides an ordering (or rather quasiordering) of points in multidimensional real space $\mathbb{R}^d$ with respect to some probability distribution P on this space. Several depth functions have been proposed since 1970s, for example halfspace depth, simplicial depth, projection

*Supported 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).

depth, L_1-depth, zonoid depth or Mahalanobis depth (for more details see [8]). Great variety of ad hoc proposed depth functions led to the general definition of the depth function stated by Zuo and Serfling in 2000 [10].

One possible application of the data depth concept is in classification. Several classifiers based on data depth have been proposed in the last ten years. The present paper is intended to study properties of the classifier proposed by Vencálek in 2013 [9]. The classifier can be shortly described as a two step procedure—an application of a well known k-nearest neighbour method on DD-plot. The procedure, called *k-depth-nearest neighbour*, is described in detail in the Section 2 of the present paper. Main results are included in the Section 3 (Simulation study) and they are commented in the Section 4.

Before describing the k-depth-nearest neighbour procedure let us recall the classification problem and introduce the notation used in this paper. For simplicity we consider two-class classification problem. We consider two groups of objects. Each object can be characterized by d numerical (real) values. We assume that the distribution P_1 of these characteristics in group 1 differs from the distribution of these characteristics in group 2 (P_2). Both distributions are assumed to be continuous. The training set—random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n_1}$ from P_1 and random sample $\boldsymbol{X}_{n_1+1}, \ldots, \boldsymbol{X}_{n_1+n_2}$ from P_2—is available. Empirical distributions based on the training set are denoted $\widehat{P}_1$, $\widehat{P}_2$, respectively.

2 k-depth-nearest neighbor method

Recently, Vencálek [9] proposed a classifier which combines the concept of data depth with the well known classification procedure of k-nearest neighbour. Newly proposed procedure, called *k-depth-nearest neighbour*, is based on the idea that points with similar location with respect to the two considered distributions have similar depths w.r.t. these distributions. The k-depth-nearest neighbor procedure is a two step procedure with the following steps:

step 1: reduction of dimension = computing depths = construction of the DD-plot,

step 2: classification = using classical kNN procedure on the DD-plot.

In the first step we compute the depth of all points in the training set w.r.t. $\widehat{P}_1$ and the depth of all points in the training set w.r.t. $\widehat{P}_2$. Any point $\boldsymbol{X}_i$ ($i \in \{1, \ldots, n_1 + n_2\}$) of the training set is thus characterized by the pair od depths $[D(\boldsymbol{X}_i; \widehat{P}_1), D(\boldsymbol{X}_i; \widehat{P}_2)]$. A two-dimensional graph displaying pairs of depths for all points in the training set is known as the DD-plot (Depth-versus-Depth plot). Any point $\boldsymbol{x} \in \mathbb{R}^d$ corresponds to a point in DD-plot whose first (horizontal) coordinate is equal to $D(\boldsymbol{x}; \widehat{P}_1)$ and its second (vertical) coordinate is equal to $D(\boldsymbol{x}; \widehat{P}_2)$. We use different colours or different symbols in the DD-plot to visualize the goup membership of points from the training set. The first step is basically nothing else than the reduction of dimension. Points from the d-dimensional real space are projected to the DD-plot: subspace of $\mathbb{R}^2$.

In the second step, the classification is performed on the DD-plot. Any classifier can be used in the DD-plot. We use the kNN and benefit from its

local nature. The procedure thus combines global nature of depth (used in the first step) with localal nature of the kNN (used in the second step). Thus, the procedure should be efficient as well as flexible.

Let us make note that a similar classifier was proposed by Li et al [5]. Their classifier separates points in DD-plot by line or a higher degree polynomial function. Such an approach would be more efficient in some cases (for example in the case of two elliptically symmetric distributions that differ in location only), on the other hand our approach should be more flexible.

Vencálek [9] presented results of a simulation study in which performance of the k-depth-nearest neighbour classifier (kDepthNN) was studied for the case of multivariate normal distributions. More precisely P_1 and P_2 were assumed to be two- or five- or ten-dimensional normal distributions which differ in location and/or scatter. kDepthNN procedure was compared to the classical kNN procedure with the folowing results: kDepthNN outperforms classical kNN in the case of different scatters (regardless of the presence or absence of the location shift). Differences in misclassification rates increase with increasing dimension—whereas kNN deteriorates dramatically with increasing dimension, the misclassification rate of the kDepthNN remains close to the Bayes optimal rate.

3 Simulation study

Results recalled above are promising, however simulations presented in [9] were performed only for multivariate normal distributions. Of course, the classification in the case of normal distributions is not challenging since LDA or QDA are known to be appropriate in this case. In the current paper we present results of a simulation study which deals with more challenging problem. We assume P_1 and P_2 be members of the family of skewed normal distributions.

The family of skewed normal distributions, introduced by Azzalini and Della Valle (see [1]), includes normal distributions and some others, that are derived from normal distribution by its "skewing". The formal definition follows:

Definition 1 A d-dimensional random vector $\boldsymbol{X}$ has a *central-skewed-normal distribution* with variance matrix $\boldsymbol{\Sigma}_0$ and skewness-regulating parameter $\boldsymbol{\alpha} \in \mathbb{R}^d$, if its density function is of form

$$f(\boldsymbol{x}) = 2\phi_d(\boldsymbol{x}, \boldsymbol{\Sigma}_0)\Phi(\boldsymbol{\alpha}^T\boldsymbol{x}),$$

where $\phi_d(\cdot, \boldsymbol{\Sigma}_0)$ is a density function of d-dimensional normal distribution with zero mean and variance matrix $\boldsymbol{\Sigma}_0$; $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution. We write $\boldsymbol{X} \sim SN_d(\boldsymbol{0}, \boldsymbol{\Sigma}_0, \boldsymbol{\alpha})$.

A d-dimensional random vector $\boldsymbol{Y}$ has a *skewed-normal distribution* with mean $\boldsymbol{\mu}$, variance matrix $\boldsymbol{\Sigma}$ and skewness-regulating parameter $\boldsymbol{\alpha} \in \mathbb{R}^d$ if

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\omega}^T\boldsymbol{X},$$

where $\boldsymbol{X} \sim SN_d(\boldsymbol{0}, \boldsymbol{\Sigma}_0, \boldsymbol{\alpha})$, and $\boldsymbol{\Sigma} = \boldsymbol{\omega}^T\boldsymbol{\Sigma}_0\boldsymbol{\omega}$. We write $\boldsymbol{Y} \sim SN_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$.

In our simulation study, we consider three two-dimensional examples and three ten-dimensional examples. Parameters of the considered two-dimensional skewed normal distributions are presented in Tab. 9. Levelsets of density of these distributions are plotted in Fig. 2. Optimal (Bayes) classifiers, i.e. classifiers with the minimal average misclassification rates, are also sketched in Fig. 2.

Example		P_1	P_2
1	$\boldsymbol{\mu}$	$(1, 2)$	$(0, -1)$
	$\boldsymbol{\Sigma}$	$diag(1, 7)$	$diag(1, 5)$
	$\boldsymbol{\alpha}$	$(-2, -5)$	$(1, 5)$
2	$\boldsymbol{\mu}$	$(0, -2)$	$(0, 2)$
	$\boldsymbol{\Sigma}$	$diag(1, 5)$	$diag(2, 14)$
	$\boldsymbol{\alpha}$	$(1, 5)$	$(-2, -5)$
3	$\boldsymbol{\mu}$	$(0.4, -0.7)$	$(-1, 0)$
	$\boldsymbol{\Sigma}$	$diag(1, 1)$	$diag(2.25, 2.25)$
	$\boldsymbol{\alpha}$	$(0, 2)$	$(2, 0)$

Table 9: Parameters of two-dimensional skewed normal distributions considered in the simulation study.

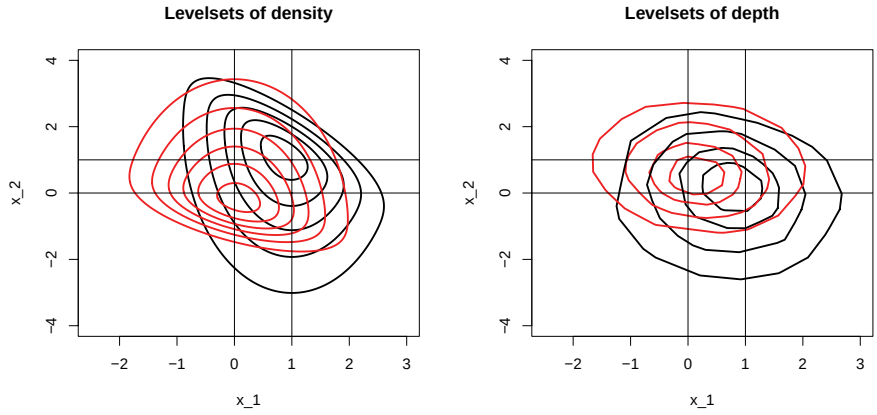

Figure 1: Comparison of levelsets of density (left) and levelsets of depth in the Example 1.

Parameters of the considered ten-dimensional skewed normal distributions are presented in Table 10.

In the simulation study, we generated 100 training sets, each consisting of exactly 100 points from P_1 and 100 points from P_2 (equal priors are considered). Four different classifiers are used to classify points from the training set: classical kNN, kDepthNN which is in the centre of our interest, linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA). The performance

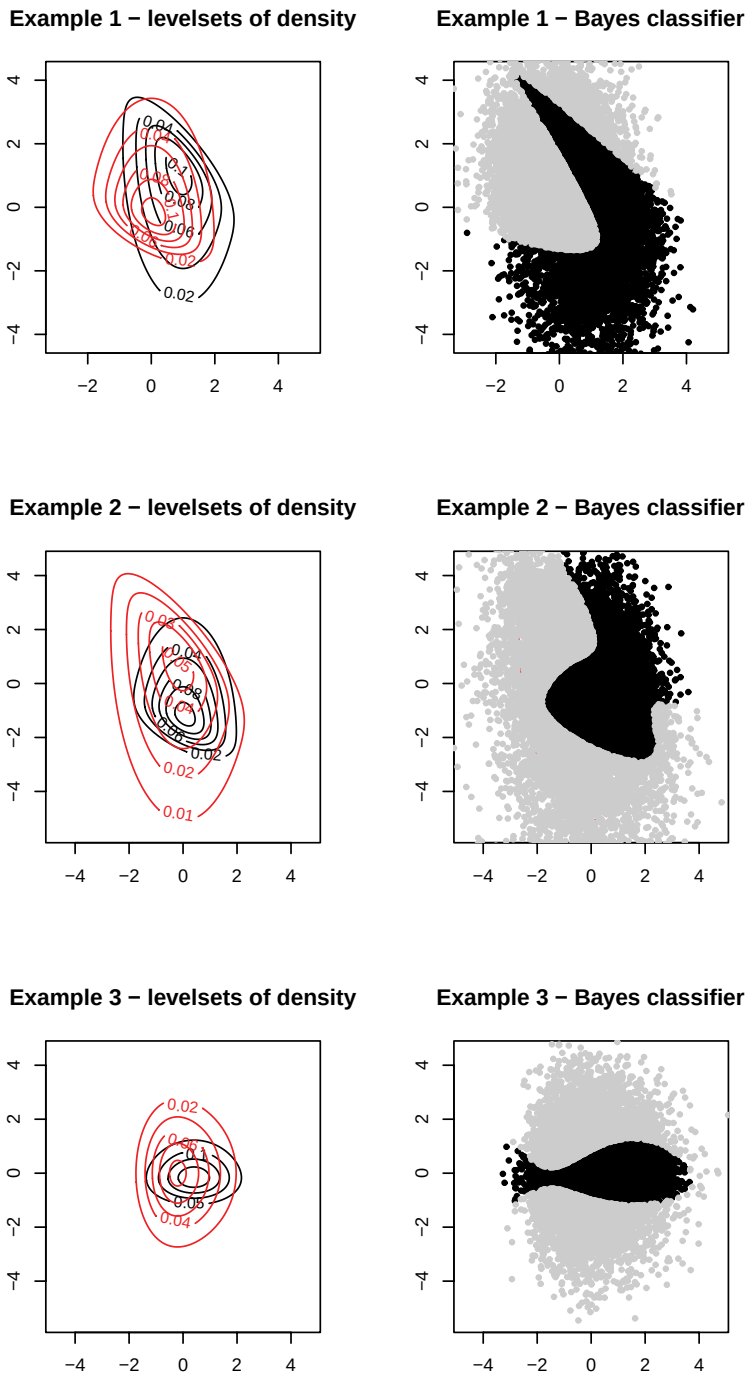

Figure 2: Levelsets of density (left) and Bayes optimal classifiers (right) for Examples 1, 2 and 3.

Example		P_1	P_2
4	$\boldsymbol{\mu}$	$(0,0,0,0,0,1,1,1,1,1)$	$(0,0,1,1,2,2,1,1,0,0)$
	$\boldsymbol{\Sigma}$	$diag(1,3,5,3,7,7,3,5,3,1)$	$2 \cdot diag(1,3,5,3,7,7,3,5,3,1)$
	$\boldsymbol{\alpha}$	$(1,2,3,4,5,5,4,3,2,1)$	$-(1,2,3,4,5,5,4,3,2,1)$
5	$\boldsymbol{\mu}$	$(0,0,0,0,0,0,0,0,0,0)$	$(0,0,0,0,0,0,0,0,0,0)$
	$\boldsymbol{\Sigma}$	$diag(1,7,3,2,7,1,7,3,2,7)$	$diag(1,2,3,2,1,1,2,3,2,1)$
	$\boldsymbol{\alpha}$	$(2,5,1,5,3,2,5,5,1,5,3)$	$(3,4,2,1,1,3,4,2,1,1)$
6	$\boldsymbol{\mu}$	$(0,0,0,0,0,0,0,0,0,0)$	$(0,0.5,1,1,2,2,1,1,0.5,0)$
	$\boldsymbol{\Sigma}$	$diag(1,3,5,3,7,7,3,5,3,1)$	$diag(1,3,5,3,7,7,3,5,3,1)$
	$\boldsymbol{\alpha}$	$(1,2,3,4,5,5,4,3,2,1)$	$-(1,2,3,4,5,5,4,3,2,1)$

Table 10: Parameters of ten-dimensional skewed normal distributions considered in the simulation study.

of classifiers is measured by average misclassification rate—percentage of incorrectly classified points. Number of considered nearest neighbors in the classical kNN and the kDepthNN procedures is chosen in such a way that it minimizes the number of incorrecly classified points.

We use the L_1-depth (also known as the spatial depth) in the simulations. Its main advantage is computational simplicity. The classification of 100 times 200 points takes about one minute on personal computer. Use of this depth function also prevents ties and ambiguity in classification. Let us recall that the L_1-depth of a point $\boldsymbol{x} \in \mathbb{R}^d$ w.r.t. a distribution P is defined as

$$D(\boldsymbol{x}; P) = 1 - \left\| \mathrm{E}_P \frac{\boldsymbol{x} - \boldsymbol{X}}{\|\boldsymbol{x} - \boldsymbol{X}\|} \right\|.$$

For more details see [8, Section 1.2.3].

The simulation was done using statistical software R. Particularly we use the package `sn` [2] which includes functions for manipulating skew-normal and skew-t probability distributions.

Results: Table 11 displays average misclassification rates for examples 1, 2 and 3 (two-dimensional distributions). The first row of the table includes information on the Bayes risks—minimal possible average misclassification rates—for the considered examples. In the first example kNN is the best classifier, in the second and the third example the QDA is the best procedure. In all three examples the kDepthNN is worse than the classical kNN. A possible explanation of such a disappointing result is suggested in the next section.

Average misclassification rates in the ten-dimensional case are summarized in Tab. 12. The first row of the table again includes information on the Bayes risks. In the Example 4 the kDepthNN outperforms clearly the classical kNN and it is the best classifier (of the four considered classifiers). QDA is comparable to the kDepthNN in this case. The kDepthNN classifier outperforms the classical kNN also in the Example 5. However, in this case QDA is much better. Finally in the Example 6 all four classifiers deteriorate considerably from the Bayes optimal classifier.

Method	Average misclassification rate		
	Example 1	Example 2	Example 3
Bayes	32.3	29.6	30.9
kNN	36.4	34.0	34.1
kDepthNN	42.8	35.8	35.9
LDA	38.4	40.3	43.1
QDA	39.5	32.9	31.8

Table 11: Average misclassification rates of the considered classifiers for three two-dimensional examples.

Method	Average misclassification rate		
	Example 4	Example 5	Example 6
Bayes	17.7	11.5	34.0
kNN	39.3	26.5	49.2
kDepthNN	26.1	22.4	49.3
LDA	38.9	38.9	47.7
QDA	26.3	15.5	49.5

Table 12: Average misclassification rates of the considered classifiers for three ten-dimensional examples.

4 Comment on the results of the simulation study

Results of the simulation study are rather disappointing. Considering skewed normal distributions the kDepthNN method seems to be worse than the classical kNN method in two-dimensional case. In ten-dimensional case its performance is unclear—sometimes it gives good results, but sometimes not. Typically if it outperforms the classical kNN method its performance is at best as good as the performance of QDA.

Let us suggest a possible explanation of such a poor performance. The notion of data depth is designed to generalize the notions "median" and "quantiles" of a univariate distribution and define "median" and "quantiles" of a multivariate distribution (see [3]). The deepest point thus corresponds to the multivariate median. Unfortunatelly for the skewed distributions the median does not correspond to the modus—the point with the highest density. The classification based on data depth (median and quantiles) is thus inappropriate in the case of skewed distributions.

Let us come back to the example 1 for a while. The difference between levelsets of density (with modus as the point with maximal density) and levelsets of depth (with multivariate median as the point with maximal depth) can be seen in Fig. 1. It can be seen that the modus of P_1 is above the median of P_1 (vertical shift) whereas the modus of P_2 (distribution that is rather left) is under the median of P_2 (again vertical shift can be observed). As the Bayes classifier is based on the density functions, the discrepancy between levelsets of density and levelsets of depth causes problems in classification.

Let us conclude this section with several suggestions how to overcome the previously described problem. These ideas might become a subject of further investigations:

- It is well known that the performance of kNN methods strongly depends on the used metric. Here we used the Euclidean metric. However, it is worthy to note that the transformation $\mathbb{R}^d \to \mathbb{R}^2$ which forms the DD-plot is quite peculiar. Thus it can be useful to study this transformation in detail and propose another metric which would lead to lower misclassification rate.
- The shape of levelsets of depth strongly depends on the used depth function. We used L_1-depth here. However, it can be useful to use another depth function. Particularly useful might be the projection depth corresponding to so called adjusted outlyingness (see [4]). Adjusted outlyingness measures outlyingness of points (hence also their centrality) w.r.t. some distribution; the adjustment is made with respect to the skewness of the distribution.

5 Conclusion

The present simulation study has uncovered inadequacy of the k-depth-nearest neigbour procedure in the case of classification problem for two skewed normal distributions. The classifier performs sometimes surprisingly well in higher dimensions, but we did not discover the class of problems for which it gives better results than its competitors. Suggestions for further research were presented at the end of the article.

References

[1] Azzalini, A., Della Valle, A.: *The multivariate skew-normal distribution.* Biometrika **83**, 4 (1996), 715–726.

[2] Azzalini, A. R: *package sn: The skew-normal and skew-t distributions (version 0.4-6).* http://azzalini.stat.unipd.it/SN, 2006.

[3] Hlubinka, D.: *Výpravy do hlubin dat.* In: Antoch J., Dohnal G. (eds.): Sborník prací 15. letní školy JČMF ROBUST 2008, *JČMF*, Praha, 2009, 97–130, (in czech).

[4] Hubert, M., van der Veeken, S.: *Fast and robust classifiers adjusted for skewness.* In: Lechevallier Y., Saporta G. (eds.): COMPSTAT 2010: proceedings in computational statistics: 19th symposium held in Paris, France *Springer*, Heidelberg, 2010, 1135–1142.

[5] Li, J., Cuesta-Albertos, J. A., Liu, R. Y.: *DD-classifier: nonparametric classification procedure based on DD-plot.* Journal of the American Statistical Association **104**, 498 (2012), 737–753.

[6] Paindaveine, D., Van Bever, G.: *Nonparametrically consistent depth-based classifiers.* arXiv preprint arXiv:1204.2996, 2012.

[7] Vencálek, O.: *Concept of Data Depth and Its Applications.* Acta Univ. Palacki. Olomuc., Fac. rer. nat., Math. **50**, 2 (2011), 111–119.

[8] Vencálek, O.: *Weighted data depth and depth based classification.* PhD Thesis, *MFF UK*, Praha, 2011.

[9] Vencálek, O.: *New depth-based modification of the k-nearest neighbour method.* Informační bulletin České statistické společnosti, (2013), (to appear).

[10] Zuo, Y., Serfling, R.: *General notion of statistical depth function.* Annals of Statistics **28** (2000), 461–482.