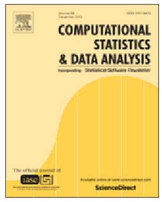
ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda



Depth-weighted Bayes classification[☆]

Ondrej Vencalek[a,*], Oleksii Pokotylo[b]

[a] Palacky University Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic
[b] Universität zu Köln, Universitätsstr. 22a, SSC Building, 50937, Köln, Germany

ARTICLE INFO

Article history:
Received 8 February 2017
Received in revised form 18 January 2018
Accepted 19 January 2018
Available online xxxx

Keywords:
Bayes classifier
Data depth
Nonparametric
Rank
Supervised learning

ABSTRACT

Two procedures for supervised classification are proposed. These are based on data depth and focus on the centre of each class. The classifiers add either a depth or a depth rank term to the objective function of the Bayes classifier. The cost of misclassifying a point depends not only on a class where it belongs, but also on its centrality with respect to this class. The classification of points that are more central is enforced while outliers are downweighted. The proposed objective function can also be used to evaluate the performance of other classifiers instead of the usual average misclassification rate. Use of the depth function increases robustness of the new procedures against the large inclusion of contaminated data that often impede the Bayes classifier. Properties of the new methods are investigated and compared with those of the Bayes classifier. Theoretical results are derived for elliptically symmetric distributions, while comparison for non-symmetric distributions is conducted by means of a simulation study. Comparisons are conducted for both theoretical classifiers and their empirical counterparts. The performance of the newly proposed classifiers is also compared to the performance of several standard methods in some real life situations.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Consider a classification problem that consists of creating a rule for assigning new observations to one of two or more distributions. We assume that the distributions are known and their supports are overlapping. There is no existing rule that offers zero probability of misclassification. A natural request is to classify at least the "typical" points correctly. For example, we seamlessly accept the wrong classification of the point $x = 4.417$, which comes from $P_1 = N(0, 1)$, as $P(|x| > 4.417 | x \sim P_1) \approx 10^{-5}$. However, points which are "close to zero" should be assigned to this distribution.

The requirement for the correct classification of "typical" points is not necessarily met when using the (Bayes) classifier which guarantees the minimal probability of misclassification, especially in the case of imbalanced data. For example, if $P_1 = N(0, 1), P_2 = N(1, 1)$ and the prior probabilities of these distributions are $\pi_1 = 0.7, \pi_2 = 0.3$, then the point $x = 1$ will be assigned to P_1 by the Bayes classifier. Although it is the centre (mode, median, mean) of the distribution P_2, it is more likely that the point $x = 1$ originates from P_1 than from P_2. In such situations, the Bayes classifier may be additionally weighted to achieve the desired misclassification rate for the minor class; however, in this case, its outliers are also overweighted, which leads to misclassification of the major class in their neighbourhood.

What is meant by the term "typical"? In the previous example, the point $x = 1$ has central position w.r.t. the distribution $N(1, 1)$. Hence, points that are close to the centre are typical whereas outliers represent rather atypical cases. But still, the

☆ The manuscript is supplemented by online material.

* Corresponding author.
E-mail address: ondrej.vencalek@upol.cz (O. Vencalek).

https://doi.org/10.1016/j.csda.2018.01.011
0167-9473/© 2018 Elsevier B.V. All rights reserved.

UNCORRECTED PROOF

Please cite this article in press as: Vencalek O., Pokotylo O., Depth-weighted Bayes classification. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.01.011.

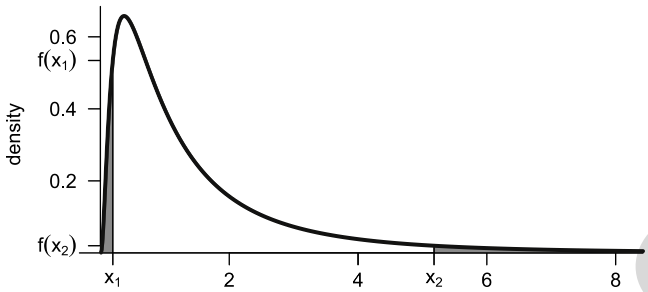


Fig. 1. The lognormal distribution. Points x_1 and x_2 determine areas containing 5% of the extreme values and thus have the same outlyingness, although the density is much higher in x_1.

term centre should be discussed in more detail. This term is clear for symmetric distributions where the centre is the point of symmetry. Note that the point of symmetry defines the median for univariate variables. The median can be considered to be the centre of the distribution even if the distribution is not symmetrical. The notion of the quantile, on the other hand, can be used to define outlyingness. The whole concept can be generalised for multivariate distributions using the notion of data depth, see e.g. Mosler, (2013), Serfling (2006) and Zuo and Serfling (2000) or Liu et al. (1999). A depth function provides a measure of centrality. The point with the highest depth is a multivariate analogy to the median, while points far from the centre have small depth.

It is important to be aware of the similarities as well as the differences between the Bayes classifier, which uses the density function and any other approach based on a depth function. In a simple but fundamental case of unimodal elliptically symmetric distribution, the level sets of the depth function correspond to those of the density function. However, the correspondence disappears as soon as the assumption of unimodality or symmetry is not fulfilled. It can be argued that the assumption of unimodality is justified in the context of classification, because a class which is a mixture can be decomposed into several unimodal subclasses. However, there is no justification for the assumption of symmetry, and the difference between the depth and the density may be substantial when the distribution is skewed. For example, consider a lognormal distribution whose logarithm is a standard $N(0,1)$ (see Fig. 1). Let $q_{0.05}$ be the 5% quantile of $N(0, 1)$. Note that $-q_{0.05}$ is the 95% quantile of $N(0, 1)$. Consider points $x_1 = \exp(q_{0.05})$ and $x_2 = \exp(-q_{0.05})$. Both x_1 and x_2 determine areas containing 5% of the extreme values (low in the case of x_1 and high in the case of x_2), so that they have exactly the same (non)-central location and their depth (more precisely: halfspace depth) is equal. However, there is a substantial difference in density (denoted f) between these points: $f(x_1) = 0.53$, $f(x_2) = 0.02$, indeed $f(x_1)$ is almost 27 times greater than $f(x_2)$. Therefore, from a classical point of view, the correct classification of the point x_1 is much more important than the correct classification of the point x_2, but is of equal importance when considering the centrality of the points.

In recent years, several classifiers using different notions of data depth have been proposed, such as the maximum-depth classifier and its improved version (Ghosh and Chaudhuri, 2005), the robust classifier using projection depth (Hubert and Van der Veeken, 2010; Dutta and Ghosh, 2012), the DD-plot classifier (Li et al., 2012), the *k*NN in the *DD*-plot (Vencalek, 2014), the DD-alpha procedure (Lange et al., 2014), the nonparametrically consistent classifier (Paindaveine and Van Bever, 2015), the classifier using local depth (Dutta et al., 2016), or the classifier suggested in Hubert et al. (2017). A comprehensive overview of depth-based classifiers can be found in Vencalek (2017).

Although the conceptualisation of these classifiers is different from that of the Bayes classifier, they have often been compared. In some cases, e.g. for elliptically symmetric distributions, the average misclassification rate (an empirical version of the probability of misclassification) of the depth-based classifiers has been shown to be asymptotically equal to the error rate of the Bayes classifier (Ghosh and Chaudhuri, 2005). Depth-based classifiers are not generally constructed to minimise the total probability of misclassification (or the average misclassification rate). In this paper, we discuss another measure of performance that can be used to evaluate classifiers such as those which are depth-based.

A discussion of alternative measures of performance of classifiers leads directly to the introduction of depth-weighted and depth-rank weighted classifiers. Instead of the global weighting of the classes, we focus on the depth weights of the misclassification cost, whereby the outlying points receive less weight than the central ones. An analysis of the properties of the newly proposed classifiers is the main objective of this paper.

This paper is structured as follows. Section 2 presents the Bayes classifier and introduces the depth-weighted classifier. Section 3 discusses the relationship between the newly proposed classifier and the Bayes classifier, focusing on their potential similarities as well as their potential differences. The choice of depth function is then discussed in Section 4 and leads on to the introduction of the rank-weighted classifier. The main results of a broad simulation study conducted to explore the behaviour of the newly proposed classifiers are reported in Section 5, a detailed description of which is provided in the online supplementary material. The robustness of the method is then assessed in Section 6, with concluding comments presented in Section 7.

UNCORRECTED PROOF

Please cite this article in press as: Vencalek O., Pokotylo O., Depth-weighted Bayes classification. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.01.011.

Download English Version:

https://daneshyari.com/en/article/6868713

Download Persian Version:

https://daneshyari.com/article/6868713

Daneshyari.com