Liet. matem. rink., T. 40, spec. nr., 281–286

Bootstrap methods in selection of the discriminant subspace

Gintautas JAKIMAUSKAS (MII), Ričardas KRIKŠTOLAITIS (VDU)
e-mail:gnt@ktl.mii.lt; ricardas_krikstolaitis@fc.vdu.lt

1. Introduction

We consider the problem of estimating a posteriori probabilities from the multi-dimensional sample supposed to satisfy multidimensional Gaussian mixture model. It is known (see, e.g. [6] and [7]) that projection to lower dimension subspace can reduce errors of estimates of a posteriori probabilities. In practice an investigator faces a dilemma: classification of the initial sample or projection to lower dimension subspace and then classification of the projected sample. One of possible methods to select whether to project the data to lower dimension subspace or not is the bootstrap method. We simulate realizations with some preliminary parameters obtained from the sample, then compare errors of estimates of a posteriori probabilities, assuming that preliminary parameters are true parameters of the sample. At this point we make a decision to project data to lower dimension subspace or not. If so, we apply projection pursuit algorithm to the sample using obtained dimension of the discriminant subspace. One of the main difficulties is to obtain sufficiently good preliminary parameters. Results presented in this paper show that we can use completely automatic procedure for obtaining parameters for the bootstrap procedure.

Theoretical background of this problem is given, e.g., in [8]–[10]. We are thankful to prof. R. Rudzkis who gave the idea and many constructive and valuable remarks.

The introduction presents already known methods. Description of the EM algorithm and the projection pursuit algorithm is given, e.g., in [8]–[9].

Main definitions. Let we have q independent d-dimensional Gaussian random variables Y_i with different distribution densities $\varphi(\cdot; M_i, R_i) \stackrel{\text{def}}{=} \varphi_i$, where means M_i and covariance matrices R_i, $i = 1, 2, \ldots, q$, are unknown. Let ν be random variable (r.v.) independent of Y_i, $i = 1, 2, \ldots, q$, and taking on values $1, 2, \ldots, q$ with unknown probabilities $p_i > 0$, $i = 1, 2, \ldots, q$, respectively. In this paper we assume that number of classes q is known. We observe d-dimensional r.v. $X = Y_\nu$. Each observation belongs to one of q classes depending on r.v. ν. Distribution density of r.v. X is therefore a Gaussian mixture density

$$f(x) = \sum_{i=1}^{q} p_i \varphi_i(x) \stackrel{\text{def}}{=} f(x, \theta), \quad x \in \mathbf{R}^d, \tag{1}$$

where $\theta = (p_i, M_i, R_i,\ i = 1, 2, \ldots, q)$ is an unknown multidimensional parameter. Probabilities $p_i = \mathbf{P}\{\nu = i\}$ are *a priori* probabilities for r.v. X to belong to ith class.

We will consider the general classification problem of estimating a *posteriori* probabilities $\pi(i,x) = \mathbf{P}\{\nu = i \mid X = x\}$ from the sample $\{X_1, X_2, \ldots, X_N\} \stackrel{\text{def}}{=} X^N$ of i.i.d. random variables with distribution density (1). Under assumptions above

$$\pi(i,x) = \pi_\theta(i,x) = \frac{p_i \varphi_i(x)}{f(x,\theta)}, \quad i = 1, 2, \ldots, q, \quad x \in \mathbf{R}^d. \tag{2}$$

The most common method to estimate a posteriori probabilities is based on the EM-algorithm (see, e.g., [8]).

Let $V = \operatorname{cov}(X, X)$ be the covariance matrix of r.v. X. Define the scalar product of arbitrary vectors $u, h \in \mathbf{R}^d$ as $(u, h) = u^T V^{-1} h$ and denote by u_H the projection of arbitrary vector $u \in \mathbf{R}^d$ to a linear subspace $H \subset \mathbf{R}^d$. Discriminant space H is defined as a linear subspace $H \subset \mathbf{R}^d$ with the property $\mathbf{P}\{\nu = i \mid X = x\} = \mathbf{P}\{\nu = i \mid X_H = x_H\}$, $i = 1, 2, \ldots, q$, $x \in \mathbf{R}^d$, and the minimal dimension. Denote $k = \dim H$. It is known that for Gaussian mixture densities (1) with equal covariance matrices we have $k < q$. Clearly, if $k < q$ and H is known, then it is better to estimate a posteriori probabilities from the projected sample rather than initial sample. Unfortunately, in practice H is not known and must be estimated and we get additional estimation errors. As shown in [6] and [7] in many cases despite additional errors the projected sample allows to decrease errors of estimation of a posteriori probabilities. But till now we do not have clear decision rules whether to project the sample or not.

2. Computer simulation results

We present computer simulation results of the statistical procedure of the selection one of two methods of estimating of a posteriori probabilities:

1) Method based on application of the EM algorithm to the initial sample from $\mathbf{R}^d$. This method is implemented in software created in Institute of Mathematics and Informatics (see [5]). This software does not require user intervention, because initial parameter estimates are selected from the sample;

2) Two stage estimation method, where in the first stage we estimate k-dimensional space H from the sample (see [9], [10]). In the second stage a posteriori probabilities are estimated using first method to the projected sample.

Computer simulation was done as follows. We simulate the sample X^N with the selected mixture model and the selected sample size (in our case $N = 300$). As basic mixture model we selected 5-dimensional Gaussian mixture model with three clusters with means $(-3, -a, 0, 0, 0)$, $(0, 2a, 0, 0, 0)$, $(3, -a, 0, 0, 0)$, equal probabilities and unit covariance matrices. At the next step we obtain parameters for bootstrap using the completely automatic procedure, which starts from no information about cluster structure. Bootstrap begins with simulating selected number (in our case 10 realizations) of independent realizations with obtained parameters now supposed to be known. To each realization we apply the procedure of calculating accuracy of estimation of a posteriori probabilities without projection, with projection to one-dimensional subspace and with projection to two-dimensional subspace (as in [7]). In all examples we assume that covariance matrices are equal and use this in all procedures.

In this paper we present four examples (all results are not covered by this paper) of computer simulation results that demonstrate applicability of the automatic parameter estimation procedure and bootstrap methods. For comparison we present (Examples 1, 2) results when parameters for bootstrap are obtained from theoretical parameters that were used for simulation of the sample X^N. In Examples 3 and 4 we make "true" bootstrap.

We studied accuracy of estimation of a posteriori probabilities, number of Bayesian classification errors (i.e., classification using estimated parameters vs. classification using theoretical parameters) and true classification errors (i.e., Bayesian classification using estimated or theoretical parameters vs. known true class numbers of the sample). Accuracy of estimation of a posteriori probabilities is measured as mean absolute distance $l(\hat{\pi}^N, \pi^N)$ between the estimated a posteriori probabilities $\hat{\pi}^N$ and the theoretical a posteriori probabilities π^N. We compare distance $l(\hat{\pi}^N, \pi^N)$ and $l(\hat{\pi}^N_H, \pi^N)$ where $\hat{\pi}^N$ are obtained from MLE in the initial space and $\hat{\pi}^N_H$ are obtained from MLE in the discriminant subspace H. Number of Bayesian classification errors is measured as percentage of differences in Bayesian classification comparing classification using known theoretical parameter versus classification using estimated parameter.

In Example 1 (see Figs. 1–3, in all Figs. on x axis we have parameter a) we make "false" bootstrap (series are simulated with theoretical parameter) and project data to x and y axes. We can compare accuracy of estimation of a posteriori probabilities, number of Bayesian classification errors with those given in [6], where we asssumed that the covariance matrices are non-equal. We get significantly less estimation errors, but the tendencies are the same.

In Example 2 (see Figs. 4–6) we make bootstrap with parameters obtained from the theoretical parameters that were used for simulation of the sample X^N.

In Example 3 (see Figs. 7–9) we make "true" bootstrap and project data to the directions obtained by projection pursuit algorithm.

In Example 4 (see Figs. 10–12) we use Gaussian mixture model with slightly non-equal covariance matrices. We replace corresponding (depending on cluster number) element on the diagonal of unit covariance matrix by the value of 1.5. But despite of it in all calculations we assume that covariance matrices are equal.

Performed tests show that the automatic procedure along with bootstrap procedure can be used to make a decision whether to project the data to lower dimension subspace or not. The advantages depend on mixture model. In general, we can make a decision to project data to lower dimension subspace.

Fig. 1. **Mean absolute error (average of 100 realizations)**

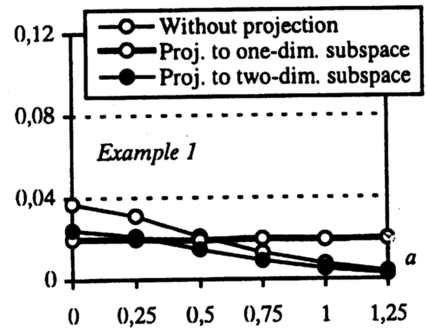


Fig.2. **Number of classification errors (avg. of 100 realizations)**

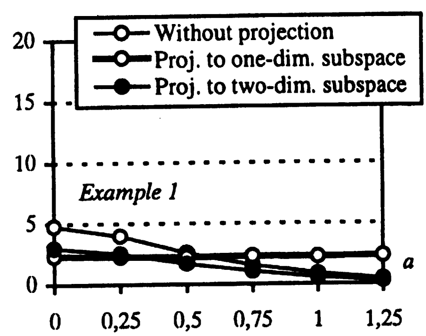


Fig. 3. **True classification errors (average of 100 realizations)**

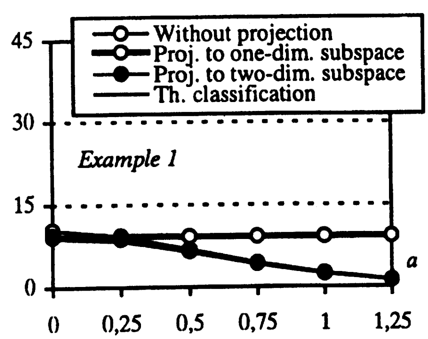


Fig. 4. **Mean absolute error (average of 100 realizations)**

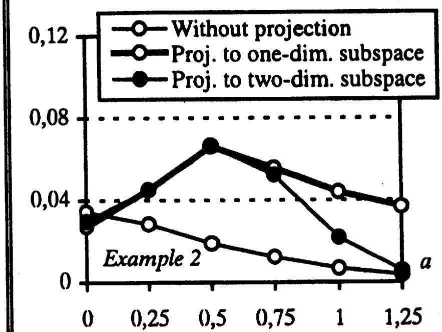


Fig.5. **Number of classification errors (avg. of 100 realizations)**

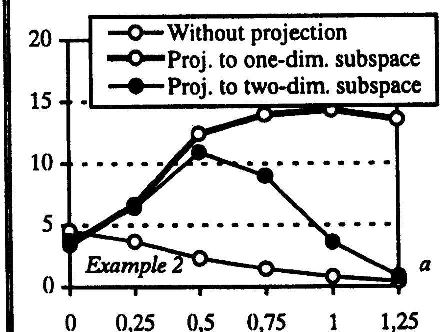


Fig. 6. **True classification errors (average of 100 realizations)**

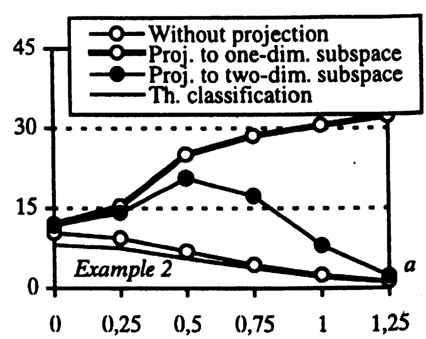


Fig. 7. **Mean absolute error (average of 100 realizations)**

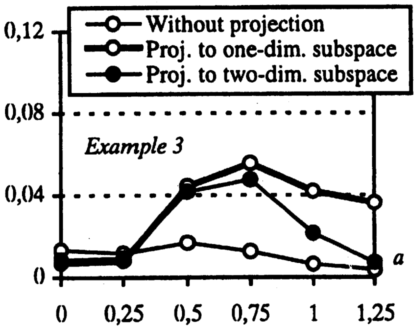


Fig. 10. **Mean absolute error (average of 20 realizations)**

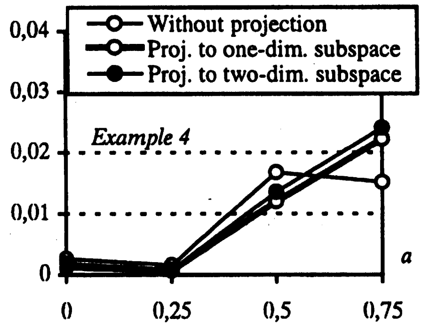


Fig.8. **Number of classification errors (avg. of 100 realizations)**

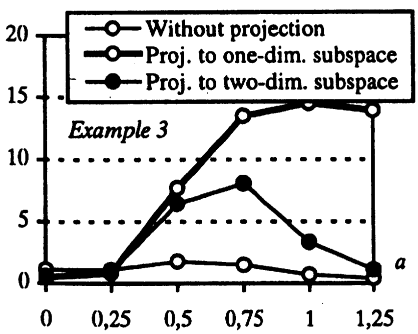


Fig. 11. **Number of classification errors (avg. of 20 realizations)**

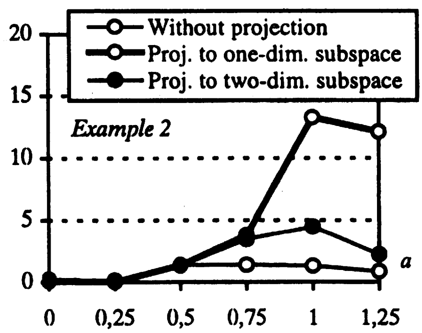


Fig. 9. **True classification errors (average of 100 realizations)**

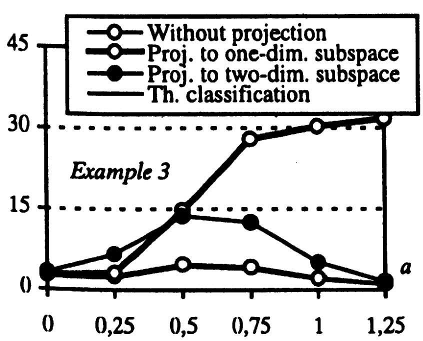
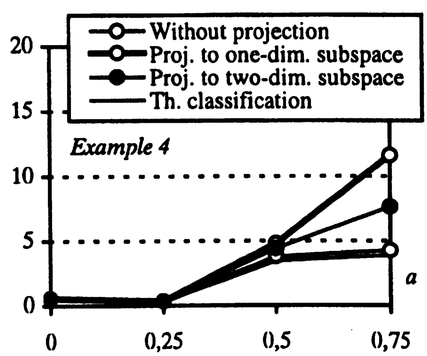


Fig. 12. **True classification errors (average of 20 realizations)**



References

[1] S.A. Aivazyan, Mixture approach to clustering via maximum likelihood, criteria of model complexity and projection pursuit, In: *Data Science, Classification and Related Methods, Abstracts of 5th IFCS Conference*, IFCS, Cobe, **1**, 36 (1996).

[2] S.A. Aivazyan, V.M. Buchstaber, I.S. Yenyukov and L.D. Meshalkin, *Applied Statistics. Classification and Reduction of Dimensionality*, Finansy i Statistika, Moscow (1989) (in Russian).

[3] J.H. Friedman: Exploratory projection pursuit, *J. Amer. Statist. Assoc.*, **82**, 249–266 (1987).

[4] J.H. Friedman and J.W.Tukey, A projection pursuit algorithm for exploratory data analysis, *IEEE Trans. Comput.*, **C–21**, 881–889 (1974).

[5] G. Jakimauskas, Efficiency analysis of one estimation and clusterization procedure of one-dimensional Gaussian mixture, *Informatica*, **8**(3), 331–343 (1997).

[6] G. Jakimauskas and R. Krikštolaitis, Appropriateness of a projection pursuit in classification, *LMD mokslo darbai*, MII, Vilnius, **3**, 370–374 (1999).

[7] G. Jakimauskas and R. Krikštolaitis, Influence of projection pursuit on classification errors: computer simulation results, *Informatica*, **11**(2), 115–124 (2000).

[8] R. Rudzkis and M. Radavičius, Statistical estimation of a mixture of Gaussian distributions, *Acta Applicandae Mathematicae*, **38**, 37–54 (1995).

[9] R. Rudzkis and M. Radavičius, Projection pursuit in Gaussian mixture models preserving information about cluster structure, *Liet. Matem. Rink.*, **37**(4), 550–563 (1997) (in Russian).

[10] R. Rudzkis and M. Radavičius, Characterization and statistical estimation of a discriminant space for Gaussian mixtures, *Acta Applicandae Mathematicae*, **58**, 279–290 (1999).

Bootstrap metodai parenkant diskriminantinės erdvės dimensiją

G. Jakimauskas, R. Krikštolaitis

Nagrinėtas apriorinių tikimybių statistinio įvertinimo uždavinys, kai stebėjimai tenkina daugiamačio Gauso mišinio modelį. Tiriamas bootstrap metodo taikymo tikslingumas, parenkant vieną iš dviejų metodų: EM algoritmo taikymą pirminiams duomenims arba projektuotiems į mažesnės dimensijos erdvę.