Probability Density Estimation Using a Gaussian Clustering Algorithm[1]

J. Ćwik and J. Koronacki

Institute of Computer Science, Polish Academy of Sciences,
Ordona 21, 01-237 Warsaw, Poland

Appeared in
Neural Computing and Applications (1996): 149-160

[1]This work was supported in part by the State Committee for Scientific Research (KBN) under grant PB 0589/P3/94/06. It was completed while the second author was on leave to the Department of Statistics, Rice University, Houston, Texas.

Abstract – A version of the Tråvén's [10] Gaussian clustering algorithm for normal mixture densities is studied. Unlike in the case of the Tråvén's algorithm, no constraints on covariance structure of mixture components are imposed. Simulations suggest that the modified algorithm is a very promising method of estimating arbitrary continuous d-dimensional densities. In particular, the simulations have shown that the algorithm is robust against assuming the initial number of mixture components to be too large.

Index terms: nonparametric density estimation, Gaussian clustering neural network

1 Introduction

Tråvén [10] proposed to use a Gaussian clustering algorithm for normal mixture densities with finitely many components to estimate an arbitrary density on a d-dimensional Euclidean space, R^d, $1 \leq d < \infty$. (While Tråvén's ultimate goal was in fact to use his estimator for pattern recognition purposes, we shall confine ourselves to the problem of estimating just one density.) It should be noted that, actually, any density can be approximated by a normal mixture to within a preassigned accuracy, say, in the L_1 norm. Perhaps the easiest way to see this is to refer to the well-known properties of nonparametric kernel estimators of probability densities (see, e.g., Devroye and Györfi [2]). Indeed, under very general conditions, kernel estimators with Gaussian kernels converge to true densities almost surely in L_1. It is another matter that one should not hope for good estimation results if a density to be estimated is not continuous and the number of random samples (sample size) is not very large. By the same token, estimation by finite mixtures can work well only if the density is continuous on R^d.

As in [10], let us assume that random samples $\mathbf{x}_1, \mathbf{x}_2 \ldots \mathbf{x}_N$ come from a mixture distribution with density

$$p(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i p(\mathbf{x}|\omega_i) \tag{1}$$

where, for $i = 1, \ldots, n$,

$$p(\mathbf{x}|\omega_i) = p(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = |2\pi\boldsymbol{\Sigma}_i|^{-1/2} \exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)) \tag{2}$$

$$\sum \alpha_i = 1, \quad \alpha_i \geq 0 \tag{3}$$

$\mathbf{x} \in R^d$ and α_i, $\omega_i \equiv (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1, \ldots n$, are unknown parameters of $p(\mathbf{x})$. Thus, in the given model, the unknowns are the mixing weights α_i and the parameters of Gaussian component densities, mean vectors $\boldsymbol{\mu}_i$ and covariance matrices $\boldsymbol{\Sigma}_i$, $i = 1, \ldots, n$. The number of mixture components, n, is assumed to be fixed. This assumption is of course very restrictive, in particular if one aims at estimating arbitrary densities. In what follows, we suggest how to overcome this restriction in practice.

In his algorithm, Tråvén [10] assumed that all the n component densities are radially symmetric, i.e., that their covariance matrices are of the form

$$\boldsymbol{\Sigma}_i = \sigma_i^2 \mathbf{I}, \quad i = 1, \ldots, n \tag{4}$$

where $\mathbf{I}$ denotes the identity matrix and the σ_i's are positive constants. Tråvén showed that, despite this assumption is apparently stringent, his algorithm can successfully be used for pattern recognition purposes. However, the situation is different if one's goal is not to distinguish between class conditional densities but to reliably estimate one density. Indeed, while assumption (4) can be considered justified for very small sample sizes N

and it makes the algorithm not only computationally simpler but also clearly having a form of a neural network, it prevents the algorithm from being a truly reliable estimator of an arbitrary probability density. In general, such an estimator can be recommended only for estimating densities which are sufficiently close to normal mixtures whose components have trivial covariance structure.

In Section 2, we present a version of the Tråvén's Gaussian clustering (GC) algorithm assuming no constraints on the covariance structure of components of mixture (2). Otherwise, the two algorithms are practically the same and, in particular, so is their relationship to other self-orginizing networks and, more generally, other pattern recognition algorithms (as discussed in detail in [10]).

The main goal of this research has been to perform a rather extensive simulation study of the two GC algorithms and to compare them with their non-recursive predecessor, the EM algorithm (see Section 2 below and, for a detailed discussion of the algorithm, e.g., [3], [6], [9]). Simulation results, summarized in Section 3, suggest that the algorithm, as proposed in this report, is indeed very promising as a method of estimating arbitrary continuous densities on R^d.

References

[1] Ćwik J, Koronacki J. Multivariate density estimation: A comparative study. Technical Report, Institute of Computer Science, Polish Academy of Sciences, 1995

[2] Devroye L, Györfi L. Nonparametric density estimation: The L_1 view. Wiley, 1985

[3] Everitt BS, Hand DJ. Finite mixture distributions. Chapman and Hall, 1981

[4] Hartigan JA, Wong MA. A k-means algorithm. Applied Statistics 1979; 28:100-108 (also in: Griffiths P, Hill ID (eds.). Applied statistics algorithms. Ellis Horwood, 1981)

[5] MacQueen JB. Some methods for classification and analysis of multivariate observations. Proc. Fifth Berkeley Symp. Math. Stat. Prob., 1967

[6] Redner RH, Walker HF. Mixture densities, maximum likelihood and the EM algorithm. SIAM Review 1984; 26: 195-239

[7] Ruppert D, Cline DBH. Bias reduction in kernel density estimation by smoothed empirical transformations. Ann. Statist. 1994; 22: 185-210

[8] Scott DW. Multivariate density estimation: Theory, practice, and visualization. Wiley, 1992

[9] Titterington DM, Smith AFM, Makov UE. Statistical analysis of finite mixture distributions. Wiley, 1985

[10] Tråvén HGC. A neural network approach to statistical pattern classification by semiparametric estimation of probability density functions. IEEE Trans. Neural Networks 1991; 2: 366-377

[11] Wand MP, Jones MC. Comparison of smoothing parametrizations in bivariate kernel density estimation. J. American Statist. Assoc. 1993; 88: 520-528

[12] Wand MP, Jones MC. Comparison of smoothing parametrizations in bivariate kernel density estimation. Technical Report, Department of Statistics, Rice University, 1991