Compressive Gaussian Mixture Estimation

Anthony Bourrier, Rémi Gribonval, Patrick Perez

Compressive Gaussian Mixture Estimation

Anthony Bourrier*†, Rémi Gribonval†, Patrick Pérez*
*Technicolor, 975 Avenue des Champs Blancs, 35576 Cesson-Sévigné, France
†INRIA Rennes-Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes, France

***Abstract*—We propose a framework to estimate the parameters of a mixture of isotropic Gaussians using empirical data drawn from this mixture. The difference with standard methods is that we only use a *sketch* computed from the data instead of the data itself. The sketch is composed of empirical moments computed in one pass on the data. To estimate the mixture parameters from the sketch, we derive an algorithm by analogy with Iterative Hard Thresholding, used in compressed sensing to recover sparse signals from a few linear projections. We prove experimentally that the parameters can be precisely estimated if the sketch is large enough, while using less memory than an EM algorithm if the data is numerous. Our approach also preserves the privacy of the initial data, since the sketch doesn't provide information on the individual data.**

I. Introduction and related work

Fitting a probability mixture model to data vectors is a widespread technique in machine learning. However, it usually requires extensive access to the data, with a prohibitive cost in terms of memory and complexity when the vectors become numerous and/or high-dimensional. The typical way of handling this issue is by subsampling the data and/or projecting the vectors in a lower-dimensional space.

Here, we consider a framework where we compute several empirical moments in one pass over the data. The moments are concatenated into a *sketch* $\hat{\mathbf{z}}$, which is a compressed representation of the data. Sketches are a common tool in large database processing algorithms and have also been used for histogram fitting [4].

To solve the estimation problem from the sketch, we cast it as an inverse problem in a similar way as [1], [3] and we search for probability mixtures which have a sketch close to $\hat{\mathbf{z}}$. Our algorithm is derived by analogy with Iterative Hard Thresholding [2] (IHT).

II. Problem statement and framework

We consider a set $F \subset L^1(\mathbb{R}^n)$ of probability densities and consider a mixture p of k densities taken in F, that is $p = \sum_{s=1}^{k} \alpha_s f_s$, where $\forall s, f_s \in F$, $\alpha_s \geq 0$ and $\sum_{s=1}^{k} \alpha_s = 1$. Given a set $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subset \mathbb{R}^n$ of vectors drawn *i.i.d.* from p, we want to estimate α_s and f_s.

The family F we consider here is a set of isotropic Gaussians:

$$F = \left\{ f_{\boldsymbol{\mu}} : \mathbf{x} \mapsto \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} \exp\left(\frac{\|\mathbf{x} - \boldsymbol{\mu}\|_2^2}{2\sigma^2} \right), \boldsymbol{\mu} \in \mathbb{R}^n \right\}. \quad (1)$$

The sketch of a density mixture q is defined as $\mathbf{A}q = (h_j(q))_{j=1}^m$, where $h_j(q) = \int_{\mathbb{R}^n} q(\mathbf{x}) e^{-i\langle \mathbf{x}, \boldsymbol{\omega}_j \rangle} d\mathbf{x}$ for m frequency vectors $\boldsymbol{\omega}_j$. This operator corresponds to a sampling of m complex values of the characteristic function of q. An estimate of $\mathbf{A}p$ can be computed from the data through the quantities $\hat{h}_j(\mathcal{X}) = \frac{1}{N} \sum_{i=1}^{N} e^{-i\langle \mathbf{x}_i, \boldsymbol{\omega}_j \rangle}$, by defining $\hat{\mathbf{z}} = \left(\hat{h}_j(\mathcal{X}) \right)_{j=1}^m$.

The parameter estimation problem is cast as the minimization of the following objective function:

$$\hat{p} = \underset{q \in \Sigma_k}{\operatorname{argmin}} \ \|\hat{\mathbf{z}} - \mathbf{A}q\|_2^2, \quad (2)$$

where $\Sigma_k := \{ f = \sum_{s=1}^{k} \alpha_s f_s | \forall s, f_s \in F \wedge \alpha_s \geq 0 \}$. Note that we do not constrain the weights to sum to 1.

III. Algorithm

To solve this problem, we propose an iterative algorithm which updates an estimate $\hat{p}$ parametrized by a vector $\boldsymbol{\alpha} \in \mathbb{R}^k$ of positive weights and a support $\hat{\Gamma} = \{\hat{\boldsymbol{\mu}}_1, \dots, \hat{\boldsymbol{\mu}}_k\} \subset \mathbb{R}^n$. Each iteration is divided in three steps:

1) M local minima of the function $\boldsymbol{\mu} \mapsto -\langle \mathbf{A}f_{\boldsymbol{\mu}}, \hat{\mathbf{r}} \rangle$ are sought and added to the current support. These local minima correspond to elements of F which, added to the current estimate with positive weights, decrease the objective function.
2) $\hat{\mathbf{z}}$ is projected on this support with a positivity constraint on the coefficients. Only the k highest coefficients and the corresponding vectors of the support are kept.
3) A gradient descent algorithm is applied to decrease the objective function with respect to the weights and support vectors.

IV. Experiments

We conducted experiments on vectors drawn from a mixture of k isotropic Gaussians with $\sigma = 1$ and compared our method to an EM algorithm, drawing weights and means randomly.

To measure the reconstruction quality, we measured empirical versions of symmetrical Kullback-Leibler (KL) divergence and Hellinger distance, both measuring the discrepancy between two probability densities (the lower the better). The following table shows the results for $n = 20, k = 10, m = 1000$. The compressed method achieves results similar to EM at a lower memory cost for numerous data.

N	Compressed		
	KL div.	Hell.	Mem.(Mb)
10^3	0.68 ± 0.28	0.06 ± 0.01	0.6
10^4	0.24 ± 0.31	0.02 ± 0.02	0.6
10^5	0.13 ± 0.15	0.01 ± 0.02	0.6
N	EM		
	KL div.	Hell.	Mem.(Mb)
10^3	0.68 ± 0.44	0.07 ± 0.03	0.24
10^4	0.19 ± 0.21	0.01 ± 0.02	2.4
10^5	0.13 ± 0.21	0.01 ± 0.02	24

Acknowledgment

This work was supported in part by the European Research Council, PLEASE project (ERC-StG-2011-277906).

References

[1] K. Bertin, E. Le Pennec, and V. Rivoirard. Adaptive Dantzig density estimation. *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, 47(1):43–74, 2011.
[2] T. Blumensath and M. Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274, 2009.
[3] F. Bunea, A. Tsybakov, M. Wegkamp, and A. Barbu. Spades and mixture models. *Annals of Statistics*, 38(4):2525–2558, 2010.
[4] N. Thaper, S. Guha, P. Indyk, and N. Koudas. Dynamic multidimensional histograms. In *ACM SIGMOD International conference on Management of data*, 2002.