Learning Mixtures of Distributions over Large Discrete Domains

Yuval Rabani[1]

1 **The Rachel and Selim Benin School of Computer Science and Engineering The Hebrew University of Jerusalem Jerusalem 91904, Israel**
yrabani@cs.huji.ac.il

Abstract

We discuss recent results giving algorithms for learning mixtures of unstructured distributions.

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems, I.5.2 Design Methodology

Keywords and phrases machine learning, mixture models, topic models

Digital Object Identifier 10.4230/LIPIcs.FSTTCS.2012.1

Summary

The past decade or so has witnessed tremendous progress in the theory of learning statistical mixture models. The most striking example is that of learning mixtures of high dimensional Gaussians. Starting from Dasgupta's ground-breaking paper [14], a long sequence of improvements [15, 5, 27, 21, 1, 17, 8] culminated in the recent results [20, 7, 23] that essentially resolve the problem in its general form. In this vein, other highly structured mixture models, such as mixtures of discrete product distributions [22, 19, 12, 18, 9, 11] and similar models [12, 6, 24, 21, 13, 10, 16], have been studied intensively.

Here we discuss recent results giving algorithms for learning mixtures of *unstructured* distributions. More specifically, we consider the problem of learning mixtures of k arbitrary distributions over a large discrete domain $[n] = \{1, 2, \ldots, n\}$. This problem arises in various unsupervised learning scenarios, for example in learning *topic models* from a corpus of documents spanning several topics. We discuss the goal of learning the probabilistic model that is hypothesized to generate the observed data, in particular the constituents (each topic distribution) of the mixture. It is information-theoretically impossible to reconstruct the mixture model from single-view samples (e.g., single word documents). Thus, multi-view access is necessary. It is desirable to minimize the *aperture* or number of views in each sample point, as well as the number of sample points needed, as these parameters govern both the applicability of an algorithm and its computational complexity. We will survey some of the results in recent papers [4, 2, 3], as well as our joint work with L.J. Schulman and C. Swamy [25, 26]. In particular, we will discuss some of the tools that contribute to these results, in brief: concentration results for random matrices, SVD and other factorizations, dimension reduction, moment estimations, and sensitivity analysis.

References

1 D. Achlioptas and F. McSherry. On spectral learning of mixtures of distributions. In *Proc. of the 18th Ann. Conf. on Learning Theory*, pages 458–469, June 2005.

32nd Int'l Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2012).
Editors: D. D'Souza, J. Radhakrishnan, and K. Telikepalli; pp. 1–3
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

2 A. Anandkumar, D.P. Foster, D. Hsu, S.M. Kakade, and Y.-K. Liu. Two SVDs suffice: Spectral decompositions for probabilistic topic modeling and latent Dirichlet allocation. *CoRR*, abs/1204.6703, 2012.

3 A. Anandkumar, D. Hsu, and S.M. Kakade. A method of moments for mixture models and hidden Markov models. *CoRR*, abs/1203.0683, 2012.

4 S. Arora, R. Ge, and A. Moitra. Learning topic models — going beyond SVD. *CoRR*, abs/1204.1956, 2012.

5 S. Arora and R. Kannan. Learning mixtures of separated nonspherical Gaussians. *Ann. Appl. Probab.*, 15(1A):69–92, 2005.

6 T. Batu, S. Guha, and S. Kannan. Inferring mixtures of Markov chains. In *Proc. of the 17th Ann. Conf. on Learning Theory*, pages 186–199, July 2004.

7 M. Belkin and K. Sinha. Polynomial learning of distribution families. In *Proc. of the 51st Ann. IEEE Symp. on Foundations of Computer Science*, pages 103–112, October 2010.

8 S.C. Brubaker and S. Vempala. Isotropic PCA and affine-invariant clustering. In *Proc. of the 49th Ann. IEEE Symp. on Foundations of Computer Science*, pages 551–560, October 2008.

9 K. Chaudhuri, E. Halperin, S. Rao, and S. Zhou. A rigorous analysis of population stratification with limited data. In *Proc. of the 18th Ann. ACM-SIAM Symp. on Discrete Algorithms*, pages 1046–1055, January 2007.

10 K. Chaudhuri and S. Rao. Beyond Gaussians: Spectral methods for learning mixtures of heavy-tailed distributions. In *Proc. of the 21st Ann. Conf. on Learning Theory*, pages 21–32, July 2008.

11 K. Chaudhuri and S. Rao. Learning mixtures of product distributions using correlations and independence. In *Proc. of the 21st Ann. Conf. on Learning Theory*, pages 9–20, July 2008.

12 M. Cryan, L. Goldberg, and P. Goldberg. Evolutionary trees can be learned in polynomial time in the two state general Markov model. *SIAM J. Comput.*, 31(2):375–397, 2002.

13 A. Dasgupta, J. Hopcroft, J. Kleinberg, and M. Sandler. On learning mixtures of heavy-tailed distributions. In *Proc. of the 46th Ann. IEEE Symp. on Foundations of Computer Science*, pages 491–500, October 2005.

14 S. Dasgupta. Learning mixtures of Gaussians. In *Proc. of the 40th Ann. Symp. on Foundations of Computer Science*, pages 634–644, October 1999.

15 S. Dasgupta and L.J. Schulman. A probabilistic analysis of EM for mixtures of separated, spherical Gaussians. *Journal of Machine Learning Research*, 8:203–226, 2007.

16 C. Daskalakis, I. Diakonikolas, and R.A. Servedio. Learning k-modal distributions via testing. In *Proc. of the 23rd Ann. ACM-SIAM Symp. on Discrete Algorithms*, pages 1371–1385, January 2012.

17 J. Feldman, R. O'Donnell, and R.A. Servedio. PAC learning mixtures of axis-aligned Gaussians with no separation assumption. In *Proc. of the 19th Ann. Conf. on Learning Theory*, pages 20–34, June 2006.

18 J. Feldman, R. O'Donnell, and R.A. Servedio. Learning mixtures of product distributions over discrete domains. *SIAM J. Comput.*, 37(5):1536–1564, 2008.

19 Y. Freund and Y. Mansour. Estimating a mixture of two product distributions. In *Proc. of the 12th Ann. Conf. on Computational Learning Theory*, pages 183–192, July 1999.

20 A.T. Kalai, A. Moitra, and G. Valiant. Efficiently learning mixtures of two Gaussians. In *Proc. of the 42nd Ann. ACM Symp. on Theory of Computing*, pages 553–562, June 2010.

21 R. Kannan, H. Salmasian, and S. Vempala. The spectral method for general mixture models. *SIAM J. Comput.*, 38(3):1141–1156, 2008.

22 M. Kearns, Y. Mansour, D. Ron, R. Rubinfeld, R. Schapire, and L. Sellie. On the learnability of discrete distributions. In *Proc. of the 26th Ann. ACM Symp. on Theory of Computing*, pages 273–282, May 1994.

23 A. Moitra and G. Valiant. Settling the polynomial learnability of mixtures of Gaussians. In *Proc. of the 51st Ann. IEEE Symp. on Foundations of Computer Science*, pages 93–102, 2010.

24 E. Mossel and S. Roch. Learning nonsingular phylogenies and hidden Markov models. In *Proc. of the 37th Ann. ACM Symp. on Theory of Computing*, pages 366–375, May 2005.

25 Y. Rabani, L.J. Schulman, and C. Swamy. Inference from sparse sampling. http://www.cs.technion.ac.il/~rabani/Papers/RabaniSS-manuscript.pdf, 2008.

26 Y. Rabani, L.J. Schulman, and C. Swamy. Learning mixtures of arbitrary distributions over large discrete domains. Unpublished, 2012.

27 S. Vempala and G. Wang. A spectral algorithm for learning mixtures of distributions. *J. Computer and System Sciences*, 68(4):841–860, 2004.