Convergence of the EM Algorithm in Gaussian Mixture Models

Lorenzo Masoero, Will Stephenston
{lom, wtstephe} @mit.edu
18.408 final project
December 20, 2017

Maximum likelihood estimation is the classical approach to learn the unknown parameters of the mixture which generated observed data. Despite the simplicity of the model, the non-convex nature of the optimization task at hand requires resorting to heuristic methods for estimation. The Expectation-Maximization (EM) algorithm (Dempster et al. [1977]) is one of the most popular techniques for estimating the parameters via maximum likelihood. Despite its widespread use and practical effectiveness, little is known whether and under which conditions EM converges to the true maximum of the likelihood function. Understanding the properties of EM and, more generally, the problem of learning mixtures of Gaussian is an active area of research. In this short report we focus on two recent results with very different flavors and conclusions:

In this short paper, we discuss both positive (section 2) and negative (section 3) recent results, sketching the ideas behind the main theorems and highlighting their implications. While for the case of two components Daskalakis et al. [2017] show that "almost any" initialization is guaranteed to converge to the global solution, Jin et al. [2016] stress the fact that for the case of three or more components, careful initialization is needed to prevent convergence to bad local optima. We are interested in understanding what causes the brittleness of EM when the model has more than two components. Specifically, we would like to understand how specific are the "bad" instances in Jin et al. [2016]: do random instances have similar properties? Does EM converge to the true optimum if we perturb the "bad" instances? We investigate these questions by means of experiments in section 4. We provide some thoughts on the results, and further directions in section 5.

References

C. Daskalakis, C. Tzamos, and M. Zampetakis. Ten steps of EM suffice for mixtures of two Gaussians. In *COLT*, 2017.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 39(1):1–38, 1977.

C. Jin, Y. Zhang, S. Balakrishnan, M. J. Wainwright, and M. I. Jordan. Local Maxima in the Likelihood of Gaussian Mixture Models: Structural Results and Algorithmic Consequences. In *NIPS*, 2016.

S. Mei, Y. Bai, and A. Montanari. The landscape of empirical risk for non-convex losses. 2017.

Y. Wang and D. M. Blei. Frequentist consistency of variational Bayes. 2017.