# A new understanding of the Expectation Maximization algorithm

Daniel Hsu

The Expectation Maximization (EM) algorithm is one of the primary workhorses for fitting statistical models to large and complex data sets. Modern statistical models, called *latent variable models*, reveal hidden patterns that explain complex observations found in data. These types of models have been used, for example, to understand community structure in social networks through observed links between people, and to uncover human ancestry through genetic markers in present-day individuals. The EM algorithm—which was formalized by Dempster, Laird, and Rubin in a 1977 paper (although special cases of the algorithm had appeared in several earlier papers)—is a computational procedure for fitting latent variable models to data based on the maximum likelihood principle. But despite its popularity and widespread use, relatively little is known about when the EM algorithm actually fulfills its intended purpose. Indeed, all results about EM until recently had come with extra "fine print" that can leave users of the algorithm with some doubt about its effectiveness.

In a 2016 paper presented at the Conference on Neural Information Processing Systems, Ji Xu, Arian Maleki, and I proved for the first time that EM can work as intended without the extra fine print, at least for one non-trivial latent variable model: a mixture of two multivariate normal distributions. This result is important because it settles decades-old questions of convergence and consistency for EM in a natural latent variable model, and it affirms empirical observations made by practitioners using EM.

The EM algorithm begins with an initial "guess" for the model that fits the data, and then iteratively revises this guess to better fit the data. The hope is that this process eventually yields the best fit model, called the maximum likelihood solution. Maximum likelihood has been studied in statistics for over a century, and is optimal in many senses. However, whether EM actually returns the maximum likelihood solution depends on the initial guess. In general, it is possible that with certain initial guesses, EM converges to a solution that is very far from a good solution. Thus, practitioners often resort to running the algorithm repeatedly from many different initial guesses, with the hope of chancing upon a guess that does lead to a good solution.

Our main theorem gives users of EM some cause for optimism. For the latent variable model we consider, we proved almost all initial guesses lead to good solutions. So, in cases where the model is indeed a good fit for the data, finding that fit can be easy to do with EM.

There are still many limitations of our theorem. First, our analysis does not consider what happens with EM when the data is not well-fit by the model we consider. For example, if there are outliers in the data, our theorem is no longer applicable. Second, it is known that an analogous theorem can not hold for models that are even slightly more complicated than the one we consider. Indeed, at the same conference where our paper was presented, another group of researchers presented a paper that proves EM almost always fails to return a good solution for mixtures of *three* multivariate normal distributions. Thus, there is still much work to be done to fully understand the behavior of this important algorithm.