Sampling Based EM Algorithm

Ashutosh Garg
Beckman Institute
University of Illinois
Urbana, IL 61801
ashutosh@uiuc.edu

Ira Cohen
Beckman Institute
University of Illinois
Urbana, IL 61801
iracohen@ifp.uiuc.edu

Thomas S. Huang
Beckman Institute
University of Illinois
Urbana, Il 61801
huang@ifp.uiuc.edu

Abstract

A sampling based EM algorithm is proposed. The algorithm tries to add some randomness to the likelihood function thereby increasing the chances of the convergence of the EM algorithm to the true maxima as against getting stuck in the local maximum of likelihood manifold. The experiments for the problem of parameter estimation for a mixture of two gaussians are presented. The improved accuracy in the parameter estimation clearly justifies the importance of the algorithm.

1 Introduction

This paper addresses the problem of parameter estimation for the mixture of gaussian. This is a very common problem and lot of work has been devoted to coming up with learning techniques that provide the estimate of the parameters as close to the true distribution as possible. Expectation-Maximization [1] is one such algorithm, which has gained lot of popularity and is widely accepted as the standard way of learning the parameters. Some of the reasons of its wide popularity can be attributed to its nice convergence properties. However, one important point that needs to be noted is that it is guaranteed to converge to the local maxima (or a saddle point) of the likelihood function, which may be very different from the true global maximum. In this work, we show using a series of examples that the estimate of parameters for a simple problem of mixture of two gaussians can be very different from the true parameters. This is followed by proposing a new algorithm that is based on the sampling idea and is referred to as sampling based EM algorithm. The intuition behind the idea is evident from the likelihood plots given in Figure 1. The three plots show the likelihood for the data that comes from the mixture of two gaussians. The x,y coordinates are the means of two gaussians. The crossbar shows the location of two gaussians. Each plot corresponds to the different mixing parameter for the two gaussians. We observe that in each plot there are two maximums of the likelihood function. One is always at the location of the true likelihood function where as the location of the other changes by changing the mixing parameter. The idea is to change the mixing parameter (by resampling the

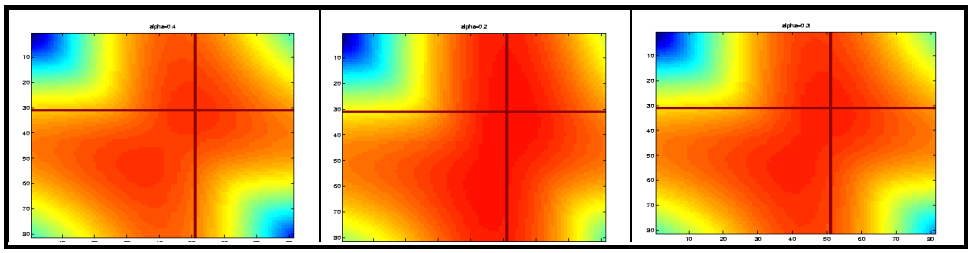

Figure 1: Likelihood plot for the mixture of two gaussians. Different plots reflect different mixing parameters for the two gaussians.

data) at each iteration of the EM algorithm. This will make the likelihood function change at each iteration. Which would imply that at each iteration the location of the maxima of the likelihood function is going to be different from that of previous iteration, however, the location of the true global maxima of the likelihood function will remain the same. This idea is very similar to the simulated annealing however, this provides a more systematic way of achieving the same thing. We also show that this algorithm has some nice properties as-

- The point to which this algorithm converges is same as the global maxima of the likelihood function
- The estimate of the parameters by this algorithm (on convergence) relate to the true parameters.

We observe above that the global maxima of the likelihood function is invariant to the mixing weights, thus if the mixing weights are changed at each iteration of the EM algorithm, it will eventually converge to the true global maxima. A similar idea was used by Freund et.al. [2] where they were estimating a mixture of two product distributions. However, a problem here is to sample the data without actually violating the i.i.d. assumption. This is achieved by introducing a third known Gaussian. By sampling the additional Gaussian and adding the new samples to the samples from the original mixture, we change the mixing weights for each Gaussian. We then run an iteration of EM on the new data, and the estimate of the parameters is obtained For the next iteration obtain a different number of samples from the third Gaussian. This changes the mixing weights at each iteration, and therefore have a different likelihood function at each iteration, but with a common global maxima.

In Figure 2., we compare the performance of this algorithm with the traditional EM algorithm. The plots give the error in the estimate of the means of the two gaussians (as the difference in the means corresponds to the KL divergence between the true distribution and the learned one). The red plot is the error for the estimated distribution (using sampling EM algorithm) whereas blue is the error for the estimated distribution using the traditional EM algorithm. The plots clearly show the improved performance by the use of sampling EM algorithm.

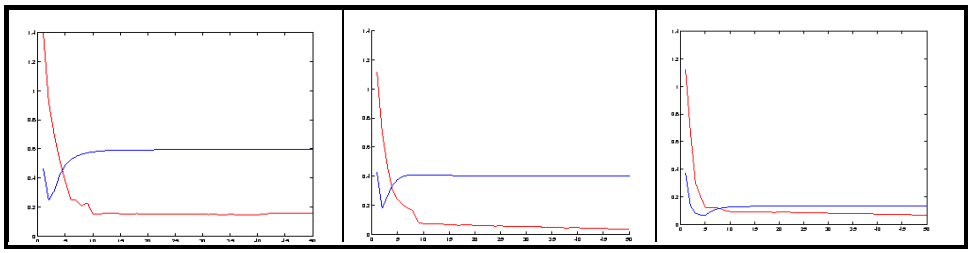

Figure 2: Performance comparison for Sampled EM vs Regular EM. Different plots are for the different mixing proportions of the two gaussians. X-axis gives the iterations. Y-axis gives the absolute error in the estimate of the means of the two gaussiams. Blue(upper) is for the regular EM where as Red(lower) is for sampled EM.

References

[1] Dempster, A.P., Laird, N.N. & Rubin, D.B. (1977) Maximum Likelihood from the incomplete data via the EM Algorithm, *Journal of Royal Statistical Society*, vol.~39, pp.~1-38.

[2] Freund, Y. & Mansour, Y. (1999) Estimating a mixture of two product distributions, *Proceedings of the twenty second Conference on Computational Learining Theory*, pp.~53-59.