ECSE 509 - *Probability and Random Signals 2 - Fall 2020*
Project

1 Background

1.1 EM algorithm

The Expectation-Maximization (EM) algorithm [1] [2] is a method to obtain the Maximum Likelihood (ML) estimator in many difficult parameter estimation problems when direct access to data necessary to estimate the parameters is impossible, or some of the data are missing. EM is an iterative method that has the desirable property of increasing the likelihood at each iteration and, therefore, it is guaranteed (under certain mild conditions) to converge to at least a local maximum.

Recall that in the ML estimation method we would like to maximize the likelihood $f(X;\theta)$ of an observation of a random sample $\underline{X} = [\underline{x}_1, \ldots, \underline{x}_n]$ of size n drawn from a distribution parameterized by θ:

$$\hat{\theta} = \arg\max_{\theta} f(X;\theta). \tag{1}$$

The EM method tries to maximize another likelihood function, which may be considerably simpler by assuming that a set of “missing" or “hidden" data is known. More specifically, we assume that there is a complete to incomplete data transformation given by

$$X = g(Y) \tag{2}$$

where the function $g(\cdot)$ is a many-to-one transformation. Then, we try to maximize $\ln f(Y;\theta)]$, i.e., we estimate θ by

$$\hat{\theta} = \arg\max_{\theta} f(Y;\theta). \tag{3}$$

In summary, the EM algorithm can be expressed in two steps:

Expectation (E): Determine the average likelihood of the complete data using the jth guess of the MLE of θ, $\theta^{(j)}$:

$$Q(\theta|\theta^{(j)}) = \mathrm{E}_{\underline{Y}}[\ln f(\underline{Y};\theta)|X, \theta^{(j)}]. \tag{4}$$

In other words, $Q(\theta|\theta^{(j)})$ is the conditional expectation of $f(\ln\underline{Y};\theta)$ w.r.t. the pdf $f(Y|X;\theta^{(j)})$.

Maximization (M): Maximize the average likelihood function of the complete data to determine the next guess of θ, $\theta^{(j+1)}$:

$$\theta^{(j+1)} = \arg\max_{\theta} Q(\theta|\theta^{(j)}). \tag{5}$$

Refer to [2], [4, Section 8.5], and [3, Section 9.13.4] for a more detailed description of the EM algorithm.

1.2 EM and Mixture models

Mixtures are weighted sums of probability density functions. A mixture of K pdfs $f_k(\cdot;\theta_k)$ (each parameterized with respect to the vector θ_k) is a density of the form

$$f(X;\theta) = \sum_{k=1}^{K} \pi_k f_k(X;\theta_k) \tag{6}$$

where the parameter vector θ is $\theta = [\theta_1^T, \theta_2^T, \ldots, \theta_k^T]^T$ and π_k is the mixing coefficient of the kth component. The weights are all positive and their sum is 1:

$$\pi_k \geq 0 \quad \text{and} \quad \sum_{k=1}^{K} \pi_k = 1 \quad \text{for} \quad k \in \{1,\ldots,K\}. \tag{7}$$

The Expectation-Maximization (EM) algorithm is the most popular algorithm to learn mixture models. The idea behind the EM algorithm for the specific problem of estimating the parameter vector of a mixture is that the problem would be much easier if we had the set of observations plus the labels (the pdf $f_k(\cdot;\theta_k)$ that each of the observations came from). This is the complete data set. Unfortunately, we only have the observations which form an incomplete data set. For a given finite data set X of n observations and an initial mixture value $\theta^{(0)}$, the algorithm provides a means to generate a sequence of estimates $\{\hat{\theta}^{(i)}\}$ with non-decreasing log-likelihood on X. As we mentioned, the EM algorithm is known to converge to a locally optimal solution. However, convergence to a globally optimal solution is not guaranteed. The log-likelihood of the given data set under the found mixture distribution is highly dependent on the initial mixture value $\theta^{(0)}$.

2 Project Description and Deliverables

Suppose that you are given a mixture of two exponential pdfs such that

$$f(x;\theta) = \pi_1 f_1(x;\lambda_1) + (1-\pi_1) f_2(x;\lambda_2) \tag{8}$$

where f_1 and f_2 are exponential distributions with parameters λ_1 and λ_2, respectively. In this assignment we want to estimate the parameter vector of this model $\theta = [\lambda_1, \lambda_2, \pi_1, \pi_2]^T$. Assume that $\lambda_1 = 1$, $\lambda_2 = 3$, and $\pi_1 = 0.25$.

(a) Plot the pdf of the exponential mixture.

(b) Suppose that we observe n independent realizations of the above mixture. Derive the E and M update equations of the EM-algorithm for this problem. (Please include the derivation in your project report).

(c) Write a program which applies the EM algorithm you derived. You must include your source code (with comments so that it is easy to follow how it works) in your project report. *If the source code is not included, you will receive a mark of zero (0) for this assignment.* Assume that $n = 20$.

(d) Compute the log-likelihood of the incomplete set at each iteration and plot it to verify that it increased monotonically.

(e) For some iterations plot the estimate of the exponential mixture pdf:

$$\hat{f}(x) = \hat{\pi}_1 f_1(x;\hat{\lambda}_1) + \hat{\pi}_2 f_2(x;\hat{\lambda}_2). \tag{9}$$

References

[1] 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*, 39:1-38, 1977.

[2] T. K. Moon, "The expectation-maximization algorithm," *Signal Processing Magazine*, Vol. 13 , Issue: 6, pp. 47 - 60, Nov. 1996.

[3] L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference*, Springer, 2004.

[4] T. Hastie *et al.*, *The Elements of Statistical Learning*, 2nd ed., Springer 2017. Available at: `https://web.stanford.edu/~hastie/ElemStatLearn/`