Akash Kumar

Technology Review

Expectation- Maximization (EM) Algorithm has been more prominently used for parameter estimation in data- driven process. EM is an iterative algorithm to find the maximum likelihood when data is missing, and model contains unobserved latent variables. This algorithm was explained and generalized by Arthur Dempster, Nan Laird, and Donald Rubin in 1977. The EM iteration alternates between two steps. E-step (Expectation) uses the observed available data of data set to estimate values of missing data and M-Step (Maximation) computes and updates the parameters maximizing the expected log-likelihood found in E-step. EM algorithms has many applications and frequently used in computer vision, data mining applications, machine, learning, and Bayesian statistics.

EM Algorithm

Let us say a complete dataset consists of $\mathcal{Z} = (\mathcal{A}, \mathcal{B})$ however only $\mathcal{A}$ is observed. Complete data log likely hood will be denoted by $l(\theta; \mathcal{A}, \mathcal{B})$ where θ is an unknown parameter for which we would like to find values of missing data.

E-Step: E-Step of EM algorithm uses observed data A to compute the expected valued of $l(\theta; \mathcal{A}, \mathcal{B})$. We can define our current parameter for example as θ_{init}. Therefore, we define E-Step as:

$$Q(\theta; \theta_{int}) := E\, l(\theta; \mathcal{A}, \mathcal{B}) \mid \mathcal{A}, \theta_{init}]$$

$$= \int l(\theta; \mathcal{A}, b)\, p(b \mid \mathcal{A}, \theta_{init})\, dy$$

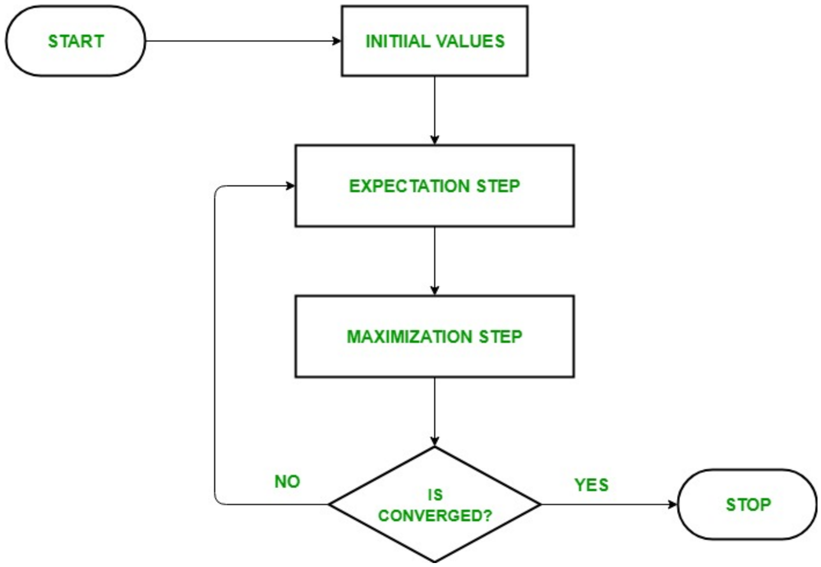
where $p(\cdot \mid \mathcal{A}, \theta_{init})$ is the conditional density of b given the observed data, $\mathcal{A}$, and assuming $\theta = \theta_{init}$.

M-Step: M-Step of EM algorithm computes the parameters maximizing the expected log-likelihood found in E-Step

$$\theta_{final} := \max Q(\theta; \theta_{init})$$

Both E-Step and M-Step are repeated until convergence. Below is a flow chart that describes this process.

Flow chart of EM algorithm



Filtering and Smoothing EM algorithms

Kalman Filter and Smoothing is used to filtering and smoothing EM algorithms.

E-step

Operate a Kalman filter or a minimum-variance smoother designed with current parameter estimates to obtain updated state estimates.

M-step

Use the filtered or smoothed state estimates within maximum-likelihood calculations to obtain updated parameter estimates.

EM algorithm has both advantages and disadvantages. Advantages of EM always guarantee that likelihood will increase after each iteration. Both E-step and M-step are easy to implement for most of the problems. One of the biggest disadvantages of EM algorithm is having a slow convergence. However, when the missing data is small, the EM converges quickly. Another disadvantage is that it requires both forward and backward probabilities.

EM algorithm is an alternating algorithm that iterates and improves by finding the lower bound to the log-likelihood and maximizing the bound. It converges to a local maximum and stops when likelihood does not change. E-Step of EM algorithm predicts the hidden values and M-step improves the estimate of parameters. E-step computes the lower bound whereas M-Step maximizes the lower bound. This algorithm has been useful in many applications in Machine learning and data mining.

References:

http://www.columbia.edu/~mh2078/MachineLearningORFE/EM_Algorithm.pdf

https://people.eecs.berkeley.edu/~pabbeel/cs287-fa13/slides/Likelihood_EM_HMM_Kalman.pdf

https://www.ics.uci.edu/~smyth/courses/cs274/notes/EMnotes.pdf

https://jonathan-hui.medium.com/machine-learning-expectation-maximization-algorithm-em-2e954cb76959