EE5111 - Estimation Theory

(Spring 2019)

Mini Project 3

Released: March 31, 2019 **Deadline**: April 9, 2019

Study of Expectation Maximization (EM) algorithm

The objective of this exercise is to understand the Expectation Maximization (EM) algorithm. Assume that there are two biased coins, labelled as A and B. Coin A lands up as head with probability p and coin B lands up as head with probability q. An experiment is performed with n independent trials and in each trial, a coin is chosen at random (coin A with probability π and coin B with probability $1-\pi$) and tossed m times (independently).

Let $z^{(i)} \in \{A, B\}$ be the label of the coin selected and $\boldsymbol{x}^{(i)} \in \{H, T\}^m$ be the observed sequence in the i^{th} trial of the experiment, $i \in 1, 2, \ldots, n$. The labels of the coins $\{z^{(i)}, i \in 1, 2, \ldots, n\}$ remain hidden and the observer could only record the faces $\{\boldsymbol{x}^{(i)}, i \in 1, 2, \ldots, n\}$. Hence, the observations $\boldsymbol{x}^{(i)}$ can be considered as *i.i.d* samples from a *Mixture of two Binomial models*:

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = \pi \times P_{\boldsymbol{X}}(\boldsymbol{x}|z^{(i)} = A) + (1-\pi) \times P_{\boldsymbol{X}}(\boldsymbol{x}|z^{(i)} = B)$$

Assume that we know $\pi = 0.5$. Thus, the vector of parameters is given by $\boldsymbol{\theta} = [p \quad q]^T$. Our aim is to obtain maximum likelihood estimate for $\boldsymbol{\theta}$ using Expectation Maximization (EM) algorithm. Generate the observations $(\boldsymbol{x}^{(i)}, z^{(i)})$ with $\pi = 0.5,\ p = 0.72$ and $q = 0.43$. Choose $m = 1, 10$ and $n = 10, 1000, 10000$. Run the EM algorithm with initial estimates $\hat{\boldsymbol{\theta}}_0 = [0.5 \quad 0.5]^T$, $[0.9 \quad 0.2]^T$, $[0.2 \quad 0.9]^T$ and $[0.7 \quad 0.4]^T$.

Make the following inferences from the algorithm for the aforementioned choices of $n,\ m$ and $\hat{\boldsymbol{\theta}}_0$:

(1) Plot the *learning curve*[1] and show the convergence of the algorithm[2].

(2) Report the number of iterations needed for convergence and final estimate of $\boldsymbol{\theta}$ from the algorithm (call it $\hat{\boldsymbol{\theta}}_{EM}$). Compare $\hat{\boldsymbol{\theta}}_{EM}$ with the true value and ML estimate with known labels $\{z^{(i)}, i \in 1, 2, \ldots, n\}$, $\hat{\boldsymbol{\theta}}_{ML}$.

(3) Observe how the estimate, $\hat{\boldsymbol{\theta}}_{EM}$ and number of iterations needed for convergence change when we increase m and n.

(4) Fix $n = 10000$ and $m = 1$. Perform EM algorithm on $K = 10000$ different sets of data. Plot the histograms[3] of $\hat{\boldsymbol{\theta}}_{EM}$ and $\hat{\boldsymbol{\theta}}_{ML}$ for the above choices of $\hat{\boldsymbol{\theta}}_0$. Using this show how the algorithm is sensitive to initial choice.

---

[1]Plot of the estimate at iteration k, $\hat{\boldsymbol{\theta}}_k$ *vs.* iteration index, k.

[2]You shall consider that the algorithm has converged at iteration k when the update to any of the parameter is not more than $\epsilon = 10^{-6}$ (i.e., $||\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_{k-1}||_\infty = max(|\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_{k-1}|) \leq \epsilon$).

[3]Overlay the histograms of $\hat{\boldsymbol{p}}_{EM}$ and $\hat{\boldsymbol{p}}_{ML}$ in the same graph and those of $\hat{\boldsymbol{q}}_{EM}$ and $\hat{\boldsymbol{q}}_{ML}$ in different graph.