An Examination of the Expected Maximization Algorithm Over an Eye

2012-04-28

Abstract

1 Specification and Organization

For the exercise reported here, there is a digital picture of a human retina as shown in figure **??**. The objective of the work is to

1. Take the gradient of the monochrome image
2. Use the Expected-Maximization algorithm to extract the edge points from the blurred optical disc boundary.

2 Background

2.1 Origins of Expected Maximization

The expected maximization algorithm is credited as first published in [1]. Various explanations have appeared in publications like [2, 3, 4].

EM extends properties of Maximum Likelihood Estimation (MLE) methods by determining via iterative convergence. [1] showed that algorithm is applied effectively to two datasets. One data set is observed and the other dataset is realized from the observed dataset. This application implies that there is a mapping between the two datasets. The dataset realized from the observed data set is considered a complete dataset.

In general, the equation for the terminating condition of the EM algorithm [2] is stated in equation 1.

$$\vec{\Theta}^{*} = \arg\max_{\Theta} \sum_{\vec{x} \in \mathcal{X}^{n}} E[\ln p(\vec{x}|\vec{\theta}, \vec{y})] \tag{1}$$

The meaning of the terms in equation 1 are as follows:

- $\mathbf{X}$ is the true data set, $\vec{x}$ is the true data set sample,
- $\vec{\Theta}$ is the set of parameters for the statistic, and
- $\vec{y}$ is the data sample in the given data set.

Equation 1 allows a characteristic to be derived in equation 2 that identifies a convergence point, $Q(\vec{\theta}; \vec{\theta^{i}})$ for the parameters identified by $\vec{\theta}$ and $\vec{\Theta}$,.

$$Q(\vec{\theta}; \vec{\theta^{i}}) = E_{D_b}[\ln p(D_g, D_b; \vec{\theta})|D_g; \vec{\theta^{i}}] \tag{2}$$

where :

- $\vec{\theta^{i}}$ is the current (best) estimate for the full distribution.
- $\vec{\theta}$ is a candidate vector for an improved estimate.
- $Q(\vec{\theta}; \theta^{i})$ is a function of $\vec{\theta}$ and $\vec{\theta^{i}}$.
- D_b is the actual data set.
- D_g is the unknown and uncorrupted data set.

- $E_{D_b}[\ln p(D_g, D_b; \vec{\theta})|D_g; \vec{\theta}^i]$ is the expected value over the missing features. The expected value hinges on $\vec{\theta}^i$ which are the estimated true parameters.

In [5], the complete dataset is assumed to be a collection of Gaussian datasets. A Gaussian dataset is defined by its mean and variance (or its multivariate equivalent.)

In the Gaussian case, each class has a proportion that defines how much each class contributes as determined by their mean and covariance (denoted $\alpha, \vec{\mu}, \mathbf{\Sigma}$ respectively). If one acquires a sample as an initial set, is there an EM algorithm that will refine these sufficient statistics? In [5], there is an example using a matrix as the result of the expectation step. From this expectation, the sufficient statistics for the next step are computed. When an expected equation forms a matrix $\mathbf{A}$, as in equation 3, it is called an expected matrix.

$$a_{ij}^{(k)} = \frac{\alpha_j p(\vec{y_i}^{(k)}|\vec{\mu_j}^{(k)}, \Sigma_j^{(k)})}{\sum_{j=1}^{M} \alpha_j p(\vec{y_i}^{(k)}|\vec{\mu_j}^{(k)}, \Sigma_j^{(k)})} \tag{3}$$

The three sufficient statistics are computed in the maximization step. They are computed by the following equations.

$$\vec{\mu}_j^{(k+1)} = \frac{\sum_{i=1}^{N} a_{ij}^{(p)} \vec{y}_i}{\sum_{i=1}^{N} a_{ij}} \tag{4}$$

$$\mathbf{\Sigma}_j^{(k+1)} = \frac{\sum_{i=1}^{N} (\vec{y}_i \vec{\mu}_j^{(p)})^T (\vec{y}_i \vec{\mu}_j^{(p)})}{\sum_{i=1}^{N} a_{ij}^{(k)}} \tag{5}$$

$$\alpha_j^{(k+1)} = \frac{1}{N} \sum_{i=1}^{N} a_{ij}^{(k)} \tag{6}$$

The obvious question is, what reduction on $\mathbf{A}$ is used to assess the convergence of the estimations? The following equation answers this question:

$$Q(\theta^*; \theta) = \log(\prod_{i=1}^{N} a_{ii}) \tag{7}$$

Also, the initial guess for $\mathbf{A}$ can not be the zero matrix. As such, the sufficient statistics would be rendered zero, and no convergence would occur. Typically, the guesses are for μ to be scattered for each class and for the expected matrix to be

$$\mathbf{A} = \frac{1}{N}\mathbf{I}.$$

2.2 Demonstrations of the Expected Maximization Algorithm

3 Approach

3.1 Univariate

3.2 Multivariate

3.3 Image Arrangement

4 Results

4.1 Expected Maximization Images

4.2 Distinguishing the Eye

4.3 Performance Under the CPU

4.4 Performance Under the Graphics Processing Unit

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):1–38, 1977.

[2] Richard O. Duda, Peter E. Hart, and David E. Stork. *Pattern Classification*. Wiley and Sons, 2nd edition, 2000.

[3] Todd K Moon. The expectation-maximization algorithm. In *IEEE Signal Processing Magazine*, pages 47–60, 1996.

[4] Seong wook Joo. Expectation maximization for mixture models.

[5] T. Yamazaki. Introduction of em algorithm into color image segmentation, 1998.