COMS30035, Machine learning: The EM algorithm

James Cussens
james.cussens@bristol.ac.uk

Department of Computer Science, SCEEM
University of Bristol

October 23, 2020

Agenda

- The EM algorithm for Gaussian mixtures
- The EM algorithm

MLE for a Gaussian mixture

$$\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\} \quad (1)$$

- No closed form for the MLE
- (At least $K!$ solutions)
- So have to resort to an iterative algorithm where we are only guaranteed a local maximum.
- Algorithm is called the *Expectation-Maximization (EM) algorithm*.

Settings derivatives to zero

$$\boldsymbol{\mu}_k = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})\mathbf{x}_n \quad (2)$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^T \quad (3)$$

$$\pi_k = \frac{N_k}{N} \quad (4)$$

where $\gamma(z_{nk}) = p(z_k = 1|\mathbf{x}_n)$ and $N_k = \sum_{n=1}^{N}\gamma(z_{nk})$.

EM for Gaussian mixtures

- The following is just an edited version of the description from Bishop [Bis06, p.438-439]
- To initialise the EM algorithm we choose starting values for $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and π.

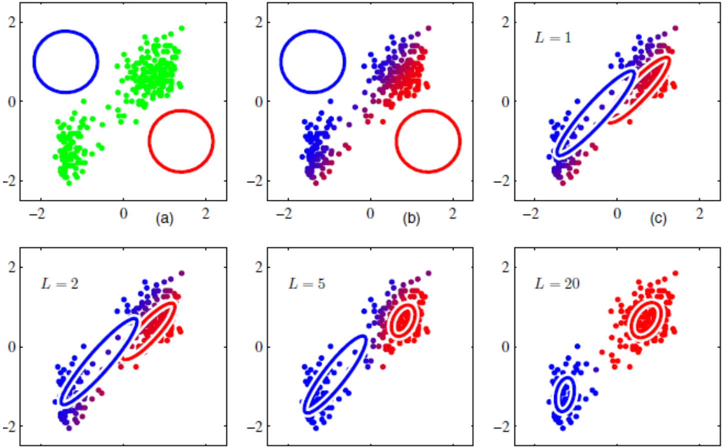
E step Compute the values for the responsibilities $\gamma(z_{nk})$ given the current parameter values:

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$

M step Re-estimate the parameters using the current responsibilities:

$$\begin{aligned}
\boldsymbol{\mu}_k^{\text{new}} &= \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n \\
\boldsymbol{\Sigma}_k^{\text{new}} &= \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}})(\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}})^T \\
\pi_k^{\text{new}} &= \frac{N_k}{N}
\end{aligned}$$

EM in pictures



Towards General EM

- EM is used to (try to) maximise log-likelihood functions of the following form:

$$\ln p(\mathbf{X}|\theta) = \ln \left\{ \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\theta) \right\} \quad (5)$$

- $\{X, Z\}$ is the *complete data*. Assume that if we had the complete data then MLE would be easy.
- $\{X\}$ is the *incomplete data*.

The General EM algorithm

E step Evaluate $p(\mathbf{Z}|\mathbf{X}, \theta^{\text{old}})$ so we have:

$$Q(\theta|\theta^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \theta^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\theta)$$

M step

$$\theta^{\text{new}} = \arg\max_{\theta} Q(\theta|\theta^{\text{old}})$$

Now do the quiz!

Yes, please do the quiz for this lecture on Blackboard!

Christopher M. Bishop.
Pattern Recognition and Machine Learning.
Springer, 2006.