ECE 7251: Signal Detection and Estimation

Spring 2002

Prof. Aaron Lanterman

Georgia Institute of Technology

Lecture 13, 2/4/02:

"The" Expectation-Maximization Algorithm

(Basic Formulation and Simple Example)

Ingredients of an EM Algorithm

- Incomplete data *Y*, that we actually measure
  - Goal: maximize the incomplete data loglikelihood (function of specific collected data)

$$l_{id}(q) = \log p_Y(y;q)$$

- Complete data *Z*, a hypothetical data set
  - Tool: complete data loglikelihood (function of complete data as a random variable)

$$l_{cd}(q) = \log p_Z(z;q)\Big|_{z=Z} = \log p_Z(Z;q)$$

- Complete data space must be "larger" and determine the incomplete data, i.e. there must be a many-to-one mapping $y=h(z)$

The EM Recipe

- Step 1: Decide on a complete data space
- Step 2: The expectation step

$$Q(q \mid \hat{q}^{old}) = E[l_{cd} \mid Y = y; \hat{q}^{old}]$$

- Step 3: The maximization step

$$\hat{q}^{new} = \arg\max_{q \geq 0} Q(q \mid \hat{q}^{old})$$

- Start with a feasible initial guess $\hat{q}^{old}$, then iterate steps 2 and 3 (which can usually be combined)

What is that Expectation?

$$E[l_{cd} \mid Y = y; \hat{q}^{old}] = \int p_{Z|Y}(z \mid y; \hat{q}^{old}) \log p_Z(z;q)\,dz$$

$$p_{Z|Y}(z \mid y; \hat{q}^{old}) = \begin{cases} \dfrac{p_Z(z;\hat{q}^{old})}{\int_{\mathcal{Z}(y)} p_Z(z;\hat{q}^{old})} & z \in \mathcal{Z}(y) \\ 0 & z \notin \mathcal{Z}(y) \end{cases}$$

$$\mathcal{Z}(y) = \{z : h(z) = y\}$$

$$\int_{\mathcal{Z}(y)} p_Z(z;\hat{q}^{old}) = p_Y(y \mid \hat{q}^{old})$$

Tasty Aspects of EM Algorithms

- Incomplete data loglikelihood is guaranteed to increase with each EM iteration
  - Must be careful; might converge to a local maxima which depends on the starting point
- Often, the estimates naturally stay in the feasible space (i.e., nonnegativity constraints)
- In many problems, a candidate complete data space naturally suggests itself

Example: Poisson Signal in Additive Poisson Noise

$$y = s + n$$

$$s \sim Poisson(q), \quad n \sim Poisson(l_n)$$

- Incomplete-data loglikelihood is

$$L_{id}(q) = -(q + l_n) + y\ln(q + l_n)$$

- ML estimator can be found in closed form:

$$\hat{q}(y) = \max(0, y - l_n)$$

- Simple "toy" example to apply EM approach

Step 1: Choose the Complete Data

- Can often choose the complete data in several different ways; try to choose to make remaining steps easy
- Different choices lead to different algorithms; some will converge "faster" than others.
- Here, take complete data to be *z=(s,n)*; suppose we could magically measure the signal and noise counts separately!
- Complete data loglikelihood is:

$$l_{cd}(q) = [-q + S\ln(q)] + [-l_n + N\ln(l_n)]$$

Step 2: The E-Step

$$Q(q;\hat{q}^{old}) = E[l_{cd} \mid Y = y;\hat{q}^{old}]$$

$$= E[-(q + l_n) + S\ln(q) + N\ln(l_n) \mid y;\hat{q}^{old}]$$

$$= -(q + l_n) + E[S \mid y;\hat{q}^{old}]\ln(q)$$

$$+ E[N \mid y;\hat{q}^{old}]\ln(l_n)$$

- Often convenient to leave explicit computation of conditional expectation until the last minute
- As with loglikelihoods, we sometimes drop terms which are constants w.r.t. q

Step 3: The M-Step

$$\hat{q}^{new} = \arg\max_{q \geq 0} Q(q;\hat{q}^{old})$$

- Take derivative as usual

$$\frac{d}{dq}Q(q;\hat{q}^{old}) = -1 + \frac{E[S \mid y;\hat{q}^{old}]}{q}$$

- Setting equal to zero yields

$$\hat{q}^{new} = E[S \mid y;\hat{q}^{old}]$$

- Now we just have to compute that pesky expectation. (That's usually the hardest part.)

That Pesky Conditional Expectation

$$E[S \mid y;q^{old}] = \int s\, p_S(s \mid y;q^{old})\,ds$$

- Let's look at the conditional density

$$p_S(s \mid y;q^{old}) = \frac{p_{Y|S}(y \mid s;q^{old})\, p_S(s;q^{old})}{p_Y(y;q^{old})}$$

$$= \frac{\dfrac{\exp[-l_n]\, l_n^{y-s}}{(y-s)!} I(y \geq s) \dfrac{\exp[-q^{old}](q^{old})^s}{s!}}{\dfrac{\exp[-(q^{old} + l_n)](q^{old} + l_n)^y}{y!}}$$

$$= \frac{y!}{s!(y-s)!} \frac{l_n^{y-s}}{(q^{old} + l_n)^{y-s}} \frac{(q^{old})^s}{(q^{old} + l_n)^s} I(s \leq y)$$

That Pesky Expectation Con't

- Ah! Conditional density is just binomial. For $0 \leq s \leq y$,

$$p_S(s \mid y;q^{old}) = \binom{y}{s} \left(\frac{q^{old}}{q^{old} + l_n}\right)^s \left(\frac{l_n}{q^{old} + l_n}\right)^{y-s}$$

$$E[S \mid y;\hat{q}^{old}] = y\frac{\hat{q}^{old}}{\hat{q}^{old} + l_n}$$

- So this particular EM algorithm is:

$$\hat{q}^{new} = E[S \mid y;q^{old}] = y\frac{\hat{q}^{old}}{\hat{q}^{old} + l_n}$$

A Quick Sanity Check

- Let's see if our analytic formula for the maximizer, $\hat{q} = \max(0, y - l_n)$, is a fixed point for the EM iteration
- For $y > l_n$,

$$\hat{q}^{new} = y\frac{\hat{q}^{old}}{\hat{q}^{old} + l_n}$$

$$y - l_n = y\frac{y - l_n}{y - l_n + l_n}$$

$$y - l_n = y - l_n$$

- For $y < l_n$, immediately get 0=0
- So everything is good

Back in Bayesianland

- EM algorithm also good for MAP estimation; just add the logprior to the *Q*-function

$$Q_P(q;\hat{q}^{old}) = E[l_{cd} \mid Y = y;\hat{q}^{old}] + \log p(q)$$

$$\hat{q}^{new} = \operatorname*{argmax}_{q \geq 0} Q_P(q;\hat{q}^{old})$$

- Consider previous example, with an exponential prior with mean 1/*a*

$$Q_P(q;\hat{q}^{old}) = -q + E[S \mid y;\hat{q}^{old}]\ln(q) - aq$$

$$\frac{Q(q;\hat{q}^{old})}{dq} = -1 + \frac{E[S \mid y;\hat{q}^{old}]}{q} - a$$

$$\hat{q}^{old} = \frac{E[S \mid y;q^{old}]}{1+a} = \left(\frac{\hat{q}^{old}}{\hat{q}^{old} + l_n}\right)\frac{y}{1+a}$$