Johns Hopkins Engineering

625.464 Computational Statistics

The Expectation Maximization Algorithm

Module 3 Lecture 3B

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Expectation Maximization Method

Assume we have observed data from r.v. X
along with missing (latent) data from r.v. Z.
And we wish to envision the complete data
from $Y = (X, Z)$.

Given obs data X, we want to maximize
a likelihood $L(\theta|X) \rightarrow$ but not directly
instead we want to work with $L(\theta|Y)$
and the densities $Y|\theta$; $Z|(X,\theta)$

EM Intuitively

observed X complete Y

missing Z

want $L(\theta \mid Y)$ based on $Y \mid \theta$ & $Z \mid (X, \theta)$

① Fills in Z based on X & θ

② Reestimates θ based on $Y = (X, Z)$

Notation

X observed data

Y complete data

Z missing data

$f_X(x|\theta)$ density of obs data

$f_Y(y|\theta)$ density of complete data

M is the many to fewer mapping $X = M(Y)$

then the missing data amounts to a marginalization model in which we observe x having density

$$f_X(x|\theta) = \int_{\{y: M(y)=x\}} f_Y(y|\theta)$$

and the cond. density of missing Z given X is

$$f_{Z|X} = \frac{f_Y(y|\theta)}{f_X(x|\theta)}$$

Also can view $L(\theta|X)$ as margin. of $L(\theta|Y) = L(\theta|X, Z)$

The EM Algorithm (to max $L(\theta|x)$)

Let $\theta^{(t)}$ be our estimate at iteration $i = 0, 1, 2, \ldots$

Define $Q(\theta|\theta^{(t)})$ to be the expectation for the joint log likelihood function of Y cond. on $X = x$.

$$Q(\theta|\theta^{(t)}) = E\left[\log L(\theta|Y) \,\middle|\, x, \theta^{(t)}\right]$$

$$= E\left[\log f_Y(Y|\theta) \,\middle|\, x, \theta^{(t)}\right]$$

$$= \int \left(\log f_Y(y|\theta)\right] f_{Z|X}(z|x, \theta^{(t)})\,dz$$

Z is only random part of Y once $X = x$

The EM Algorithm

Starting with $\theta^{(0)}$

1. E step: Compute $Q(\theta|\theta^{(t)})$
2. M step: Maximize $Q(\theta|\theta^{(t)})$ w.r.t. θ and set $\theta^{(t+1)}$ equal to this maximizer
3. Return to E step unless stopping criteria has been met.

Stopping criteria usually built upon

$$|Q(\theta^{(t+1)}|\theta^{(t)}) - Q(\theta^{(t)}|\theta^{(t)})| \quad \text{or} \quad (\theta^{(t+1)}-\theta^{(t)})^T(\theta^{(t+1)}-\theta^{(t)})$$

A Very Simple EM Example

$y_1, y_2 \sim$ i.i.d. $\text{Exp}(\theta)$ so $f(y_i) = \theta e^{-\theta y_i}$ & $E[y_i] = \frac{1}{\theta}$

suppose we know $y_1 = 5$ and y_2 is missing.

$\underbrace{y_1}_{x}$ $\underbrace{y_2}_{z}$

$y = (x, z) = (y_1, y_2) = (5, y_2)$

- Write down the complete density of

$$f_y(y|\theta) = \theta e^{-\theta y_1} \theta e^{-\theta y_2}$$

$$f_{z|x}(z|x,\theta) = f_z(z|\theta) = \theta e^{-y_2 \theta}$$

A Very Simple EM Example

- write down the log likelihood function of the complete data y

$$\log L(\theta|y) = \log f_y(y|\theta) = 2\log\theta - \theta y_1 - \theta y_2$$

- Find $Q(\theta|\theta^{(t)}) = E[\log L(\theta|y) \mid x, \theta^{(t)}]$

$$= E[2\log\theta - \theta y_1 - \theta y_2 \mid y_1, \theta^{(t)}]$$

$$= 2\log\theta - \theta \cdot 5 - \theta E[y_2 \mid y_1, \theta^{(t)}]$$

$$\boxed{= 2\log\theta - 5\theta - \theta/\theta^{(t)} = Q(\theta|\theta^{(t)})}$$

E step

A Very Simple EM Example

- M step: maximize $Q(\theta|\theta^{(t)}) = 2\log\theta - 5\theta - \theta/\theta^{(t)}$

Set $Q'(\theta|\theta^{(t)}) = 0$

$$0 = \frac{2}{\theta} - 5 - \frac{1}{\theta^{(t)}}$$

$$\theta^{(t+1)} = \theta = \frac{2(\theta^{(t)})}{5(\theta^{(t)}+1)}$$

Repeat

Comments on EM Algorithm

① The sequence $\{\theta^{(k)}\}$ converges at least to a local maximum. (and for well behaved problems to a global max)

② We move uphill at each step.

③ Rate of Convergence: Linear and linked to the proportion of the data that is missing.

④ Starting Points → sensitive to initial guesses.