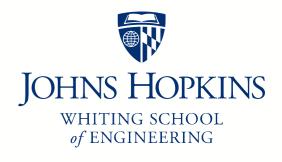
Johns Hopkins Engineering 625.464 Computational Statistics

The Expectation Maximization Algorithm

Module 3 Lecture 3B



Expectation Maximization Method

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

biven obsidate X, we want to maximize a likelihood L(O(X) = but not directly instead we want to work with L(O(Y) and the densities y10 = 2/(x, b)

EM Intuitively

complete Y absence X missina Z wort L(614) based on Y10 F2/(X,0) (1) Fills in 2 based on X ? O 2) Reestimates 0 based on $\sqrt{-(X, Z)}$

Notation fx (x)0) density obsta X observed data y complete data fy (yl D) density of complete Z missing data M is the many to tewer mapping X = M(Y)then the missing data amounts to a mary inalization model in which we doserve x having density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing?

[x|x|th = (fx(y|th)) and the cond. density of missing.

[x|x|th = (fx(y|th)) and the cond. density of missing.

[x|x|th = (fx(y|th)) and the cond. density of missing.

[x|x|th = (fx(y|th)) and the cond. density of missing.

[x|x|th = (fx(y|th)) and the cond. density of missing.

[x|x|th = (fx(y|th)) and the cond. density of missing.

[x|x|th = (fx(y|th)) and the cond. density of missing.

[x|x|th = (fx(y|th)) and the cond. density of missing.

[x|x|th = (fx(y|th)) and the cond. density of missing.

[x|x|th = (fx(y|th)) and the cond. density of missing.

[x|x|th = (fx(y|th)) and the cond.

[x|x|th = (fx(y|th)) a

The EM Algorithm (to Max L(A))

Let $\Theta^{(+)}$ be our estimate at iteration i=0,52,...Define DLO (b) to be the expectation for the joint log likelihood function of I cond. on X=x-Q(blb(t))= E[log L(b))x, D(t) = E [log fy(1/0)] x (9) = \(\log \f_{\formall} \log \f_{\formall} \log \f_{\formall} \log \f_{\formall} \log \f_{\formall} \log \formall Zis only random part of I me X=X

The EM Algorithm

Starting with D10)

- (i) E step: Compute Q(6/0")
- (2) M step: Maximize Q(DD) w.r.t. (2) and set D(tt) equal to this maximizer

 (3) Return to Estep unless stopping criteria has been met.

Stopping criteria usually built upon $|Q(\theta^{(t+1)}|\theta^{(t)}-Q(\theta^{(t+1)}|\theta^{(t)})| \text{ or } (\theta^{(t+1)}-\theta^{(t)})|$

A Very Simple EM Example

$$\sqrt{1},\sqrt{2} \sim i.i.d.$$
 Exp(b) so figure De Bristyring of Suppose we know y=5 and yz is missing. $\sqrt{1} = (X_1Z_1) = (X_1X_2) = (5, 3/2)$

• Write down the complete density of $f_1(Y|B) = \theta e^{-\theta Y_1} \theta e^{-\theta Y_2}$ $f_2(x,\theta) = f_2(z|\theta) = \theta e^{-y_2\theta}$

A Very Simple EM Example

write down the log likelihood function of the complete data y

Find Q(b|btb) = E[log L(bly) | x, b(t)

= 2 log D - 6.5 - D E [y2 | y, D E Ster

$$=2\log \Theta -5\Theta -\Theta_{6t}) = Q(\Theta | \Theta^{(t)})$$

A Very Simple EM Example

of Step: maximize
$$O(6/0^{(t)}) = 2 \log \theta - 50 - O(t)$$
Sut $O'(0)(0^{(t)}) = O(t)$

$$0 = \frac{2}{5} - 5 - \frac{1}{6}$$

$$O^{(t+1)} = O = \frac{Z(e^{(t)})}{5(e^{(t)})}$$

Ropeat

Comments on EM Algorithm

- The sequence EO(K) converges at least to a local maximum (and for well behaved problems to a global max)
- Due more uphill at each step.
- (3) Rate of Convergence: Linear and linked to the proportion of the data that is missing.
 - (4) Starting Points -> sonsitive to initial quesses.