ECE 7251: Signal Detection and Estimation

Spring 2002

Prof. Aaron Lanterman

Georgia Institute of Technology

Lecture 14, 2/6/02:

"The" Expectation-Maximization Algorithm (Theory)

Convergence of the EM Algorithm

- We'd like to prove that the likelihood goes up with each iteration:

$$L_{id}(\hat{q}^{new}) \geq L_{id}(\hat{q}^{old})$$

- Recall from last lecture,

$$p_{Z|Y}(z \mid y;q) = \frac{p_Z(z;q)}{p_Y(y;q)}$$

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

- Take logarithms of both sides and rearrange:

$$\ln p_Y(y;q) = \ln p_Z(z;q) - \ln p_{Z|Y}(z \mid y;q)$$

Deriving the Smiley Equality

- Multiply both sides by $p_{Z|Y}(z \mid y;q)$ and integrate with respect to z:

$$\int p_{Z|Y}(z \mid y;\hat{q}^{old}) \ln p_Y(y;q)dz$$

$$= \int p_{Z|Y}(z \mid y;\hat{q}^{old}) \ln p_Z(z;q)dz$$

$$- \int p_{Z|Y}(z \mid y;\hat{q}^{old}) \ln p_{Z|Y}(z \mid y;q)dz$$

- Simplifies to:

$$L_{id}(q) = Q(q;\hat{q}^{old}) - \int p_{Z|Y}(z \mid y;\hat{q}^{old}) \ln p_{Z|Y}(z \mid y;q)dz$$

Call this equality

Using the Smiley Equality

- Evaluate at $q = \hat{q}^{new}$ and $q = \hat{q}^{old}$

$$L_{id}(\hat{q}^{new}) = Q(\hat{q}^{new};\hat{q}^{old}) - \int p_{Z|Y}(z \mid y;\hat{q}^{old}) \ln p_{Z|Y}(z \mid y;\hat{q}^{new})dz$$

$$L_{id}(\hat{q}^{old}) = Q(\hat{q}^{old};\hat{q}^{old}) - \int p_{Z|Y}(z \mid y;\hat{q}^{old}) \ln p_{Z|Y}(z \mid y;\hat{q}^{old})dz$$

- Subtract from

$$L_{id}(\hat{q}^{new}) - L_{id}(\hat{q}^{old}) = Q(\hat{q}^{new};\hat{q}^{old}) - Q(\hat{q}^{old};\hat{q}^{old})$$

$$- \int p_{Z|Y}(z \mid y;\hat{q}^{old}) \ln \frac{p_{Z|Y}(z \mid y;\hat{q}^{new})}{p_{Z|Y}(z \mid y;\hat{q}^{old})} dz$$

Use A Really Helpful Inequality

$$L_{id}(\hat{q}^{new}) - L_{id}(\hat{q}^{old}) = Q(\hat{q}^{new};\hat{q}^{old}) - Q(\hat{q}^{old};\hat{q}^{old})$$

$$- \int p_{Z|Y}(z \mid y;\hat{q}^{old}) \ln \frac{p_{Z|Y}(z \mid y;\hat{q}^{new})}{p_{Z|Y}(z \mid y;\hat{q}^{old})} dz$$

- A really helpful inequality: $\ln(x) \leq x - 1$

$$L_{id}(\hat{q}^{new}) - L_{id}(\hat{q}^{old}) \geq Q(\hat{q}^{new};\hat{q}^{old}) - Q(\hat{q}^{old};\hat{q}^{old})$$

$$\underbrace{- \int p_{Z|Y}(z \mid y;\hat{q}^{old}) \left[\frac{p_{Z|Y}(z \mid y;\hat{q}^{new})}{p_{Z|Y}(z \mid y;\hat{q}^{old})} - 1 \right] dz}$$

Let's focus on this final term for a little bit

Zeroness of the Last Term

$$\int p_{Z|Y}(z \mid y;\hat{q}^{old}) \left[\frac{p_{Z|Y}(z \mid y;\hat{q}^{new})}{p_{Z|Y}(z \mid y;\hat{q}^{old})} - 1 \right] dz$$

$$= \int p_{Z|Y}(z \mid y;\hat{q}^{new}) - p_{Z|Y}(z \mid y;\hat{q}^{old})dz$$

$$= \int p_{Z|Y}(z \mid y;\hat{q}^{new})dz - \int p_{Z|Y}(z \mid y;\hat{q}^{old})dz$$

$$= 1 - 1 = 0$$

I Think We Got It!

- Now we have:

$$L_{id}(\hat{q}^{new}) - L_{id}(\hat{q}^{old}) \geq Q(\hat{q}^{new};\hat{q}^{old}) - Q(\hat{q}^{old};\hat{q}^{old})$$

- Recall the definition of the M-step:

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

- So, by definition $Q(\hat{q}^{new};\hat{q}^{old}) \geq Q(\hat{q}^{old};\hat{q}^{old})$
- Hence $L_{id}(\hat{q}^{new}) \geq L_{id}(\hat{q}^{old})$

Some Words of Warning

- Notice we showed that the likelihood was nondecreasing; that doesn't automatically imply that the parameter estimates converge
- Parameter estimate could slide along a contour of constant loglikelihood
- Can prove some things about parameter convergence in special cases
  - Ex: EM Algorithm for Imaging from Poisson Data (i.e. Emission Tomography)

Generalized EM Algorithms

- Recall this line:

$$L_{id}(\hat{q}^{new}) - L_{id}(\hat{q}^{old}) \geq Q(\hat{q}^{new};\hat{q}^{old}) - Q(\hat{q}^{old};\hat{q}^{old})$$

- What if the M-step is too hard? Try a "generalized" EM algorithm:

$$\hat{q}^{new} = \text{some easy to compute } q$$

$$\text{such that } Q(q;\hat{q}^{old}) \geq Q(\hat{q}^{old};\hat{q}^{old})$$

- Note convergence proof still works!

The SAGE Algorithm

- Problem: EM algorithms tend to be slow
- Observation: "Bigger" complete data spaces result in slower algorithms than "smaller" complete data spaces
- SAGE (Space-Alternating Generalized Expectation-Maximization), Hero & Fessler
  - Split big complete data space into several smaller "hidden" data spaces
  - Designed to yield faster convergence
- Generalization of "ordered subsets" EM algorithm