Lecture 9b: The EM Algorithm Extras

Sam Roweis

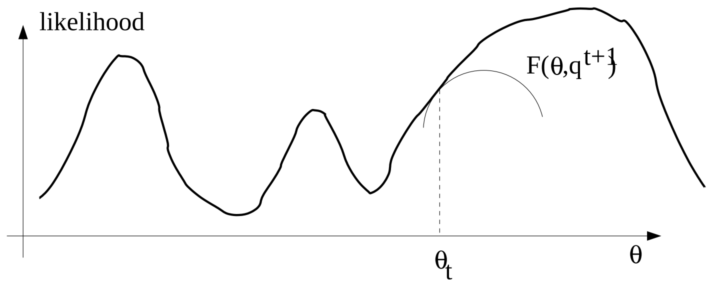
February 13, 2002

Variants

- Sparse EM:
  Do not recompute exactly the posterior probability on each data point under all models, because it is almost zero.
  Instead keep an "active list" which you update every once in a while.
- Generalized (Incomplete) EM: It might be hard to find the ML parameters in the M-step, even given the completed data. We can still make progress by doing an M-step that improves the likelihood a bit (e.g. gradient step).

EM Constructs Sequential Convex Lower Bounds

- Consider the likelihood function and the function $F(q^{t+1}, cdot)$.



Partially Hidden Data

- Of course, we can learn when there are missing (hidden) variables on some cases and not on others.
- In this case the cost function is:

$$\ell(\theta;\mathcal{D}) = \sum_{\text{complete}} \log p(\mathbf{x}^c, \mathbf{z}^c|\theta) + \sum_{\text{hidden}} \log \sum_{\mathbf{z}} \log p(\mathbf{x}^h, \mathbf{z}|\theta)$$

- In the E-step we estimate the hidden variables on the incomplete cases only.
- The M-step uses the sufficient statistics from the complete data plus the expected sufficient statistics from the E-step.