Gaussian Mixture Models (GMM)

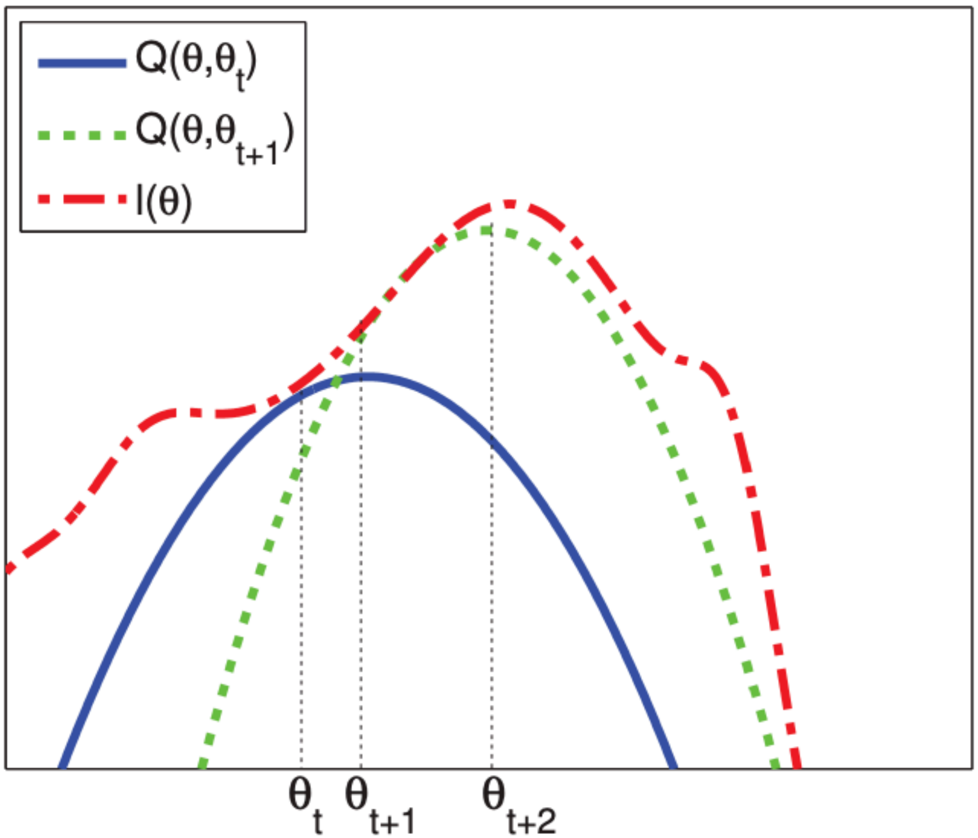
ECE57000: Artificial Intelligence, Fall 2019

David I. Inouye

The Expectation-Maximization (EM) algorithm can be seen as a generalization of k-means

- The EM algorithm for GMM alternates between
  - Probabilistic/soft assignment of points
  - Estimation of Gaussian for each component
- Similar to k-means which alternates between
  - Hard assignment of points
  - Estimation of mean of points in each cluster

EM algorithm is **guaranteed** to increase *observed* likelihood, i.e., $\prod_i p_{mixture}(x_i)$



Observation: If we knew z_i, then optimizing the complete log likelihood is easy

- Observed/marginal log likelihood (if z_i is **unknown**)

$$\ell(\theta) = \log \prod_i \left(\sum_j p(z_i) p(x_i | z_i) \right)$$

- Complete log likelihood (if z_i is **known**)

$$\ell_c(\theta) = \log \prod_i p(x_i, z_i; \theta) = \log \prod_i \pi_{z_i} p_{\mathcal{N}}\left(x_i; \mu_{z_i}, \Sigma_{z_i}\right)$$

  - For GMMs, this is convex and easy to solve

Derivation of EM iteration for GMM

- Complete log-likelihood
$$\ell_c(\theta) = \sum_i \log p(x_i, z_i | \theta)$$
- Expected complete log likelihood
$$Q(\theta; \theta^{t-1}) = Q_{\theta^{t-1}}(\theta) = \mathbb{E}_{\boldsymbol{z}_{\dots}|\boldsymbol{x}_{\dots}, \theta^{t-1}}[\ell_c(\theta)]$$
  - **NOTE:** Q is a function of θ **given** the previous parameter value θ^{t-1}
- Let's write the joint density of x and z as:
$$p(x_i, z_i | \theta) = \prod_j \left(\pi_j p(x_i | \theta_j)\right)^{I(z_i = j)}$$
  - $I(z_i = j)$ is an indicator function that is 1 if the inside expression is true or 0 otherwise
- See 11.22-11.26 pp. 351 of [ML] for derivation

Proof that it monotonically increases likelihood

- See 11.4.7 in [ML] for full derivation of proof
- Show that $Q(\theta;\ q^t)$ is lower bound observed likelihood $\ell(\theta)$, i.e., $\ell(\theta) \geq Q(\theta; q^t), \forall\theta$
- Choose $q^t(z_i) = p(z_i|x_i, \theta^t)$, which corresponds to $Q(\theta; \theta^t)$
- Show that lower bound is tight at θ_t
- Combine three concepts
  1. Lower bound inequality
  2. Maximization inequality
  3. Tightness of lower bound