DATA130004: Homework 8

Due in class on December 15, 2016

1. EM algorithm can be seen as a Minorize-Maximization algorithm. Let $\ell(\theta)$ be a concave function to be maximized. A constructed function $Q(\theta|\theta_m)$ is called the minorized version of the objective function at θ_m if

   (a) $Q(\theta|\theta_m) \leq \ell(\theta)$ for all θ

   (b) $Q(\theta_m|\theta_m) = \ell(\theta_m)$

   Then, maximizing $\ell(\theta)$ reduces to maximizing $Q(\theta|\theta_m)$, i.e.,

   $$\theta_{m+1} = \arg\max_{\theta} Q(\theta|\theta_m).$$

   Now show that the above iterative method will guarantee that $\ell(\theta_m)$ converges to its local maximum or a saddle point.

2. Refer to the lecture notes about mixture models. Now suppose Y follows p-component Gaussian mixture density,

   $$f(y|\theta) = \sum_{r=1}^{p} \pi_r f(y|\mu_r, \sigma_r^2)$$

   where $\theta = \{\pi_r, \mu_r, \sigma_r^2 : 1 \leq r \leq p\}$. Follow the steps to derive the EM algorithm.

   (a) Let U be the latent variable which indicates which subpopulation Y comes from. Derive the log-likelihood function for complete data $\log f(y, u|\theta)$.

   (b) The conditional distribution of $U|Y, \theta$ is a discrete distribution. Given another θ' which is different from θ, calculate $\mathrm{P}(U = r|Y = y, \theta')$ and denote it by $w_r(y, \theta')$.

   (c) (E-step) When the complete data $(y_1, u_1), \ldots, (y_n, u_n)$ are assumed to be i.i.d. samples, show that

   $$\begin{aligned} Q(\theta, \theta') &= \mathrm{E}\{\log f(U, Y|\theta)|Y = y, \theta'\} \\ &= \sum_{r=1}^{p} \left\{ \sum_{j=1}^{n} w_r(y_i, \theta') \right\} \log \pi_r + \sum_{r=1}^{p} \sum_{j=1}^{n} w_r(y_i, \theta') \log f(y_j|\mu_r, \sigma_r^2) \end{aligned}$$

(d) (M-step) Maximize $Q(\theta, \theta')$ with respect to θ when fixing θ'. Use † to indicate the maximizer. Show that

$$\pi_r^\dagger = \frac{1}{n}\sum_{j=1}^{n} w_r(y_i, \theta')$$

$$\mu_r^\dagger = \frac{\sum_{j=1}^{n} w_r(y_i, \theta') y_j}{\sum_{j=1}^{n} w_r(y_i, \theta')}$$

$$\sigma_r^{2\dagger} = \frac{\sum_{j=1}^{n} w_r(y_i, \theta')(y_j - \mu_r^\dagger)^2}{\sum_{j=1}^{n} w_r(y_i, \theta')}$$