LDA updates

Runjing (Bryan) Liu

November 25, 2016

For a particular document d, the posterior distribution for its topic distribution θ_d and the latent topics $\{z_{dn}\}_{n=1}^{N_d}$ for the N_d words is given by

$$p(\theta_d, \{z_{dn}\}_{n=1}^{N_d} | \{w_{dn}\}_{n=1}^{N_d}, \alpha, \beta, a_d, b_d) = \frac{p(\theta_d|\alpha)\Big(\prod_{n=1}^{N_d} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn}, \beta)\Big) p(a_d, b_d | \{z_{dn}\}_{n=1}^{N_d}, \eta, \sigma^2)}{\int p(\theta_d|\alpha) \sum_{z \cdot d} \Big(\prod_{n=1}^{N_d} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn}, \beta)\Big) p(a_d, b_d | \{z_{dn}\}_{n=1}^{N_d}, \eta, \sigma^2)\ d\theta_d}$$

In particular, the normalizer gives us the likelihood of the observed words $\{w_{dn}\}_{n=1}^{N_d}$ and the response a_d and b_d. This is computationally intractable, so we apply a variational method in which we assume a fully factorized posterior distribution for θ_d and z_{dn} of the form

$$q(\theta_d, \{z_{dn}\}_{n=1}^{N_d} | \gamma, \{\phi_n\}_{n=1}^{N_d}) = q(\theta_d|\gamma) \prod_{n=1}^{N} q(z_{dn}|\phi_n)$$

where ϕ_n is a vector on the symplex that gives the multinoulli probabilities of z_{dn}, and γ is the K dimensional Dirichlet parameter for θ_d. With respect to these two variational parameters, we seek to maximize the evidence lower bound (ELBO) given by

$$\mathcal{L}(\gamma, \phi; \alpha, \beta, \eta, \sigma) = E_q[\log p(\theta_d|\alpha)] + \sum_{n=1}^{N_d} E_q[\log p(z_{dn}|\theta)] + \sum_{n=1}^{N_d} E_q[\log p(w_{dn}|z_{dn}, \beta)] + ...$$
$$+ E[\log p(a_d, b_d | \{z_{dn}\}_{n=1}^{N_d}, \eta, \sigma^2)] - E_q[\log q(\theta_d|\gamma)] - \sum_{i=1}^{N_d} E([\log q(z_{dn}|\phi_n)])$$

The updates for ϕ_j ($j \in \{1, ..., N_d\}$) and γ were derived in (Blei & McAuliffe, 2008) and given by

$$\phi_j^{new} \propto \exp\left\{ E[\log \theta_d|\gamma] + E[\log p(w_{dn}|\beta)] + \Big(\frac{y}{N_d \sigma^2}\Big)\eta - \frac{[2(\eta^T \phi_{-j})\eta + (\eta \circ \eta)]}{2N_d^2 \sigma^2} \right\}$$

where $\phi_{-j} := \sum_{n \neq j} \phi_n$; and $E(\log \theta_i|\gamma] = \Psi(\gamma_i) - \Psi(\sum \gamma_j)$, where Ψ is the digamma function. Now the updates for γ is:

$$\gamma^{new} \to \alpha + \sum_{n=1}^{N_d} \phi_n$$

These give the updates for a single document response pair. Having updated γ_d and ϕ_d for each document, the updates for β are given by examing the entire corpus (Blei et al, 2003):

$$\beta_{k,w}^{new} \propto \sum_{d=1}^{D} \sum_{n=1}^{N_d} 1\{(w_{dn} = w)\} \phi_{d,n}^k$$