May 9, 2013

Abstract

1 Derivation

Variational Distribution

$$q(\phi,\pi,d,z|\lambda,\gamma,\eta,\varphi) = \prod_{k=1}^{K} q(\phi_k|\lambda) \prod_{m=1}^{M} q(\pi_m|\gamma) \prod_{n=1}^{N} q(d_n|\eta)q(z_n|\varphi) \tag{1}$$

where

$\phi_k \in R^V$ distribution over word of topic k

$\pi_m \in R^K$ distribution over topics of document m

$d_n \in R^M$ indicator vector where if $d_{nm} = 1$ means word n belongs to document m

$z_n \in R^K$ indicator vector where if $z_{nk} = 1$ means word n is assigned to topic k

Lower bound of log likelihood:

$$\begin{aligned}
p(\boldsymbol{w},\boldsymbol{c}|\alpha,\beta,\eta,\sigma) &= \log \int_\phi \int_\pi \sum_d \sum_z p(\phi,\pi,d,z,\boldsymbol{w},\boldsymbol{c}|\alpha,\beta,\eta,\sigma) &(2)\\
&= \log \int_\phi \int_\pi \sum_d \sum_z q(\phi,\pi,d,z|\lambda,\gamma,\eta,\varphi) \frac{p(\phi,\pi,d,z,z,w|\alpha,\beta,\eta,\sigma)}{q(\phi,\pi,d,z|\lambda,\gamma,\eta,\varphi)}\, \mathrm{d}\phi\, \mathrm{d}\pi\, \mathrm{d}d\, \mathrm{d}z &(3)\\
&= \log E_q \frac{p(\phi,\pi,d,z,\boldsymbol{c},\boldsymbol{w}|\alpha,\beta,\eta,\sigma)}{q(\phi,\pi,d,z|\lambda,\gamma,\eta,\varphi)} &(4)\\
&\geq E_q \log p(\phi,\pi,d,z,\boldsymbol{c},\boldsymbol{w}|\alpha,\beta,\eta,\sigma) - E_q \log q(\phi,\pi,d,z|\lambda,\gamma,\eta,\varphi) = L &(5)
\end{aligned}$$

where $E_q \equiv E_{q(\phi,\pi,d,z|\lambda,\gamma,\eta,\varphi)}$

Instead of maximizing p, we estimate it by maximizing the lower bound L

Consider the first term in L

$$\begin{aligned}
E_q[\log p(\phi,\pi,d,z,\boldsymbol{c},\boldsymbol{w}|\alpha,\beta,\eta,\sigma)] &= E_q[\log p(\phi|\beta)\, p(\pi|\alpha)\, p(d|\eta)\, p(\boldsymbol{c}|c_d^{\mathrm{d}},\sigma)\, p(z|\pi_d)\, p(\boldsymbol{w}|\phi_d) &(6)\\
&= E_q[\log p(\phi|\beta)] + E_q[p(\pi|\alpha)] + E_q[\log p(d|\eta)] &(7)\\
&+ E_q[\log p(\boldsymbol{c}|c_d^{\mathrm{d}},\sigma)] + E_q[\log p(z|\pi_d)] + E_q[\log p(\boldsymbol{w}|\phi_z)] &(8)\\
& &(9)
\end{aligned}$$

$$\begin{aligned}
E_q[\log p(\phi|\beta)] &= E_q \log \prod_{k=1}^{K} p(\phi_k|\beta) &(10)\\
&= \sum_{k=1}^{K} E_q[\log\Gamma(\sum_{v=1}^{V}\beta_v) - \sum_{v=1}^{V}\log\Gamma(\beta_v) + \sum_{v=1}^{V}(\beta_v - 1)\log(\phi_{kv})] &(11)\\
&= \sum_{k=1}^{K} \log\Gamma(\sum_{v=1}^{V}\beta_v) - \sum_{v=1}^{V}\log\Gamma(\beta_v) + \sum_{v=1}^{V}(\beta_v - 1)(\Psi(\lambda_{kv}) - \Psi(\sum_{v'=1}^{V}\lambda_{kv'})) &(12)
\end{aligned}$$

$$\begin{aligned}
E_q[\log p(\pi|\alpha)] &= E_q[\log \prod_{m=1}^{M} p(\pi_m|\alpha)] && (13)\\
&= \sum_{m=1}^{M} E_q \log(p(\pi_m|\alpha)) && (14)\\
&= \sum_{m=1}^{M} E_q[\log(\Gamma(\sum_{k=1}^{K}\alpha_k)) - \sum_{k=1}^{K}\log(\Gamma(\alpha_k)) + \sum_{k=1}^{K}(\alpha_k - 1)\log(\pi_m k)] && (15)\\
&= \sum_{m=1}^{M} \log(\Gamma(\sum_{k=1}^{K}\alpha_k)) - \sum_{k=1}^{K}\log(\Gamma(\alpha_k)) + \sum_{k=1}^{K}(\alpha_k - 1)(\Psi(\gamma_{mk}) - \Psi(\sum_{k'=1}^{K}\gamma_{mk'})) && (16)
\end{aligned}$$

Speed up things a little bit:

$$\begin{aligned}
E_q[\log p(d|\eta)] &= \sum_{n=1}^{N} E_q \log(p(d_n|\eta)) && (17)\\
&= \sum_{n=1}^{N} E_q \sum_{m=1}^{M} d_n m \log(\eta) && (18)\\
&= \sum_{n=1}^{N}\sum_{m=1}^{M} \eta \log(\eta) && (19)\\
&= (N+M)\eta\log(\eta) && (20)
\end{aligned}$$

$$\begin{aligned}
E_q[\log p(\boldsymbol{c}|c_d^{\mathrm{d}}, \sigma)] &= \sum_{n=1}^{N} E_q \log p(c_n|c_{d_n}^{\mathrm{d}}) && (21)\\
&= \sum_{n=1}^{N}\sum_{m=1}^{M} E_q[d_{mn}\log p(c_n|c_{d_{mn}}^{\mathrm{d}})] && (22)\\
&= \sum_{n=1}^{N}\sum_{m=1}^{M} \eta(\delta_{g_{d_{mn}}^{\mathrm{d}}} - \frac{(x_{d_{mn}}^{\mathrm{d}} - x_n)^2 + (y_{d_{mn}}^{\mathrm{d}} - y_n)^2}{\sigma^2}) && (23)\\
& && (24)
\end{aligned}$$

$$\begin{aligned}
E_q[\log p(z|\pi_d)] &= \sum_{n=1}^{N}\sum_{m=1}^{M} E_q[d_{mm}\log(\pi_{mn})] && (25)\\
&= \sum_{n=1}^{N}\sum_{m=1}^{M} \eta(\Psi(\gamma_{mn}) - \Psi(\sum_{m'=1}^{M}\gamma_{m'n}) && (26)
\end{aligned}$$

$$\begin{aligned}
E_q[\log p(\boldsymbol{w}|\phi_z)] &= \sum_{n}^{N}\sum_{m}^{M}\sum_{k}^{K}\sum_{v}^{V} E_q[d_{mn} z_{kn} \boldsymbol{w}_{\boldsymbol{n}}^{\boldsymbol{v}} log(\phi_k v)] && (27)\\
&= \sum_{n}^{N}\sum_{m}^{M}\sum_{k}^{K} \eta\phi_{kn}\boldsymbol{w}_{\boldsymbol{n}}^{\boldsymbol{v}}(\Psi(\lambda_{kv}) - \Psi(\sum_{k'}^{K}\lambda_{k'v}) && (28)\\
& && (29)
\end{aligned}$$