Latent Dirichlet Allocation

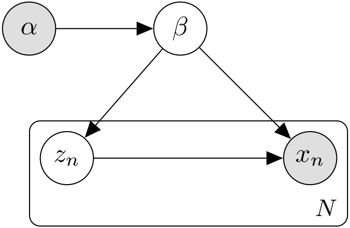
Kangcheng Hou
kangchenghou@gmail.com

Zhejiang University

May 4, 2018

Variational Inference I

Now let's see how to implement do variational inference for the following model.



The joint distribution is

$$p(x, z, \beta|\alpha) = p(\beta|\alpha) \prod_{n=1}^{N} p(x_n, z_n|\beta)$$

Variational Inference II

We make the following assumptions, the conditional distribution belongs to exponential family distribution,

$$p(\beta|x, z, \alpha) = h(\beta) \exp(\eta_g(x, z, \alpha)^\top T(\beta) - A_g(\eta_g(x, z, \alpha)))$$

and

$$p(z_{nj}|x_n, z_{n,\neg j}, \beta) = h(z_{nj}) \exp(\eta_l(x_n, z_{n,\neg j}, \beta)^\top T(z_{nj}) - A_l(\eta_l(x_n, z_{n,\neg j}, \beta)))$$

The evidence lower bound is

$$\mathcal{L}(q) = \mathbb{E}_q[\log p(x, z, \beta)] - \mathbb{E}_q[\log q(z, \beta)]$$

We select the distribution q such that it is both expressive and easier to optimize, (Note that we only restrict the function form for p, but have not specifies the factorization form of p.)

$$q(z, \beta) = q(\beta|\lambda) \prod_{n=1}^{N} \prod_{j=1}^{J} q(z_{nj}|\phi_{nj})$$

Variational Inference III

Other than the factorization of $q(z, \beta)$, we restrict $q(z, \beta)$ and $q(z_{nj}|\phi_{nj})$.

$$q(\beta|\lambda) = h(\beta)\exp(\lambda^\top T(\beta) - A_g(\lambda))$$

$$q(z_{nj}|\phi_{nj}) = h(z_{nj})\exp(\phi_{nj}^\top T(z_{nj}) - A_l(\phi_{nj}))$$

Now let's alter the parameter λ, ϕ to maximize the ELBO. Rather than $\mathcal{L}(q)$, we can write it as $\mathcal{L}(\lambda, \phi)$. It turns out we can optimize λ and ϕ in turn. We first fix ϕ and optimize λ.

Variational Inference IV

$$\begin{aligned}
\mathcal{L}(q) &= \mathbb{E}_q[\log p(x,z,\beta)] - \mathbb{E}_q[\log q(z,\beta)] \\
&= \mathbb{E}_q[\log p(\beta|x,z)] + \underbrace{\mathbb{E}_q[\log p(x,z)]}_{\text{not related to } \beta} - \mathbb{E}_q[\log q(z,\beta)] \\
&= \mathbb{E}_q[\log h(\beta)] + \mathbb{E}_q[\eta_g(x,z)]^\top \mathbb{E}_q[T(\beta)] - \mathbb{E}_q[A_g(\eta_g(x,z))] \\
&\quad - \mathbb{E}_q[\log h(\beta)] - \mathbb{E}_q[T(\beta)^\top \lambda] + \mathbb{E}[A_g(\lambda)] \\
&= \mathbb{E}_q[\eta_g(x,z)]^\top \mathbb{E}_q[T(\beta)] - \mathbb{E}_q[A_g(\eta_g(x,z))] - \mathbb{E}_q[T(\beta)^\top \lambda] \\
&\quad + \mathbb{E}[A_g(\lambda)] \\
&= \mathbb{E}_q[\eta_g(x,z)]^\top \mathbb{E}_q[T(\beta)] - \mathbb{E}_q[T(\beta)]^\top \lambda + A_g(\lambda) \\
&= \nabla_\lambda A_g(\lambda)^\top (\mathbb{E}_{q(z|\phi)}[\eta_g(x,z)] - \lambda) + A_g(\lambda)
\end{aligned}$$

Maximize w.r.t $\mathcal{L}(\lambda)$ a.k.a $\mathcal{L}(q)$, $\frac{d\mathcal{L}(\lambda)}{d\lambda} = 0$, we have

$$\nabla^2_\lambda A_g(\lambda)^\top (\mathbb{E}_{q(z|\phi)}[\eta_g(x,z)] - \lambda) = 0$$

Variational Inference V

Assume that $\nabla^2_\lambda A_g(\lambda) \neq 0$, so we have

$$\mathbb{E}_{q(z|\phi)}[\eta_g(x, z)] - \lambda = 0$$

That's cool! Right? What have we done so far?

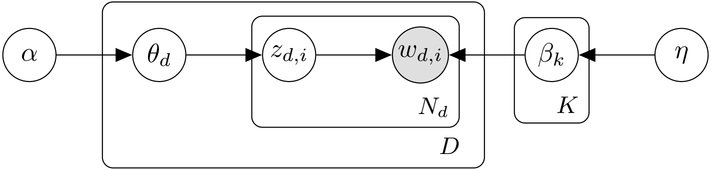
- We first specifies a graphical model, and specifies the latent variables e.g. z, β.
- We define the distribution so that the conditional distribution for latent varaibels given other variables is always exponential family distribution.
- For joint distribution of latent variable e.g. $q(z, \beta)$, we assume that $q(z, \beta)$ factorizes into $q(z|\lambda)q(\beta|\phi)$.

Variational Inference VI

- Then, we have a very nice conclusion from our previous lengthy derivation that if we optimize the parameters e.g. λ, ϕ one by one, we just do it in this way. Repeat between $\lambda = \mathbb{E}_{q(z|\phi)}[\eta_g(x, z)]$ and $\phi = \mathbb{E}_{q(z|\lambda)}[\eta_l(x, \beta)]$ which is **set the parameter to be the expectation of natural prameter conditioning on all other latent variables**.

Variational Inference for LDA I

We will now do variational inference for the following model,



Here is a pipeline for doing variational inference.

- The graphical model encodes the conditional independence information of the model.
- We assign the exponential family distribution to the distribution such that we can get exponential family conditional distribution for future ease.
- Write the q distribution for each full conditional distribution and do optimization.

Variational Inference for LDA II

For LDA, The document-topic distribution is

$$\theta_d \sim \mathrm{Dir}(\alpha)$$

The topic-word distribution is

$$\phi_k \sim \mathrm{Dir}(\beta)$$

The topic assignment of each word is

$$z_{d,i} \sim \mathrm{Mult}(\theta_d)$$

And for each word,

$$w_{d,i} \sim \mathrm{Mult}(\theta_{z_{d,i}})$$

Variational Inference for LDA III

We have three kinds of latent variables, θ_d, $z_{d,i}$ and ϕ_k respectively. We now derive the conditional distribution of these given all other variables. First, we derive the conditional distribution of $z_{d,n}$, it can be shown that the conditional distribution of $z_{d,n}$ is

$$
\begin{aligned}
& p(z_{d,n} = k|\theta_d, w_{d,n}, \phi_k) \\
\propto \quad & p(w_{d,n}|z_{d,n} = k, \phi_k)p(z_{d,n} = k|\theta_d) \\
= \quad & \phi_{k,w_{d,n}}\theta_{d,k} \\
= \quad & \exp(\ln[\phi_{k,w_{d,n}}\theta_{d,k}])
\end{aligned}
$$

Variational Inference for LDA IV

For conditional distribution for the topic distribution of each document θ_d, it can be shown that

$$p(\theta_d | \text{all other variables}) = p(\theta_d | z_d)$$

Variational Inference for LDA V

$$
\begin{aligned}
& p(\theta_d|z_d) \\
\propto \quad & p(\theta_d|\alpha)\prod_{n=1}^{N_d} p(z_{d,n}|\theta_d) \\
= \quad & \prod_{k=1}^{K}(\theta_{d,k}^{\alpha_k-1}\prod_{n=1}^{N_d}\theta_{d,k}^{\delta(z_{d,n},k)}) \\
= \quad & \prod_{k=1}^{K}(\theta_{d,k}^{\alpha_k-1+n_k}) \qquad \text{where } n_k=\sum_{n=1}^{N_d}\delta(z_{d,n},k) \\
= \quad & \exp(\sum_{k=1}^{K}(\alpha_k-1+n_k)\log\theta_{d,k})
\end{aligned}
$$

Variational Inference for LDA VI

For conditional distribution for the word distribution of each topic ϕ_k,

$$
\begin{aligned}
& p(\phi_k|Z,W) \\
\propto \quad & p(\phi_{k,v}|\beta)\prod_{d=1}^{D}\prod_{n=1}^{N_d} p(w_{d,n}|\phi_k)^{\delta(z_{d,n},k)} \\
= \quad & \prod_{v=1}^{V}\phi_{k,v}^{\beta_v-1}\prod_{d=1}^{D}\prod_{n=1}^{N_d} p(w_{d,n}|\phi_k)^{\delta(z_{d,n},k)} \\
= \quad & \prod_{v=1}^{V}\phi_{k,v}^{\beta_v-1}\prod_{d=1}^{D}\prod_{n=1}^{N_d} \phi_{k,w_{d,n}}^{\delta(z_{d,n},k)} \\
= \quad & \exp(\sum_{v=1}^{V}(\beta_v - 1 + n_v)\log\phi_{k,v})
\end{aligned}
$$

Variational Inference for LDA VII

where n_v is the number of occurence of word v which belongs to topic k.
So now we have the three conditional distribution of latent parameters. Now we can perform variational inference for our graphical model.

$$p(z_{d,n} = k|\theta_d, w_{d,n}, \phi_k) = \exp(\ln[\beta_{k,w_{d,n}}\theta_{d,k}])$$

$$p(\theta_d|z_d) = \exp(\sum_{k=1}^{K}(\alpha_k - 1 + n_k)\log\theta_{d,k})$$

$$p(\beta_k|Z,W) = \exp(\sum_{v=1}^{V}(\eta_v - 1 + n_v)\log\beta_{k,v})$$

Variational Inference for LDA VIII

We assume that the variational function is in the form of

$$q(z_{d,n}) = \mathsf{Mult}(\phi_{d,n})$$
$$q(\theta_d) = \mathsf{Dir}(\gamma_d)$$
$$q(\beta_k) = \mathsf{Dir}(\lambda_k)$$

According to the property of exponential family, we have the following conclusion

$$\eta(\phi_{d,n}) = \mathbb{E}_{q(\theta_d)q(\beta_k))}(\ln[\beta_{k,w_{d,n}}\theta_{d,k}])$$
$$\eta(\gamma_d) = \mathbb{E}_{q(z_{d,n})q(\beta_k))}(\left[\alpha_1 - 1 + n_1 \cdots \alpha_K - 1 + n_K\right]^\top)$$
$$\eta(\beta_k) = \mathbb{E}_{q(z_{d,n})q(\theta_d))}(\left[\eta_1 - 1 + n_1 \cdots \eta_V - 1 + n_V\right]^\top)$$

Variational Inference for LDA IX

$$
\begin{aligned}
\eta(\phi_{d,n}) &= \mathbb{E}_{q(\theta_d)q(\beta_k))}(\ln[\beta_{k,w_{d,n}\theta_{d,k}}]) \\
&= \mathbb{E}_{q(\theta_d)}[\ln \theta_{d,k}] + \mathbb{E}_{q(\beta_k)}[\ln \beta_{k,w_{d,n}}] \\
&= \Psi(\gamma_{d,k}) - \Psi(\sum_{k=1}^{K} \gamma_{d,k}) + \Psi(\lambda_{k,w_{d,n}}) - \Psi(\sum_{v=1}^{V} \lambda_{k,v})
\end{aligned}
$$

$$
\eta(\gamma_d) = \left[\alpha_1 - 1 + \phi_{d,n}^1 \cdots \alpha_K 1 + \phi_{d,n}^K\right]^\top
$$

$$
\eta(\beta_k) = \eta - 1 + \sum_{d=1}^{D} \sum_{n=1}^{N} w_{d,n} \phi_{d,n}^k
$$

Huh! That's our updating rules for the variational inference for LDA.

Perplexity