Latent Dirithlet Allocation

Saki

August 29, 2018

Abstract

1 Reference

Blei, D.M., Ng, A.Y. and Jordan, M.I., 2003. Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), pp.993-1022.

2 Model

Generative Process for each word

$$\begin{aligned} \theta &\sim Dir(\alpha) \\ z_n &\sim Multinomial(\theta) \\ w_n &\sim p(w_n|z_n, \beta) \end{aligned}$$

Joint Probbailities of θ, z_n, w_n given α, β

$$p(\theta, Z, W|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta)$$

3 Story

- Imagine a big box, which corresponds to a document. We put small boxes, which corresponds to topics, into the big box.

- The size of the small boxes are determined by θ (e.g. Sports box takes 50% of the big box, Politics box 20%, and Economics box 30%). ($p(\theta|\alpha)$)

- Once the small box are fit, we chose one small box($p(z_n|\theta)$), and throw a ball (i.e. a word) into the chosen small box.($p(w_n|z_n, \beta)$)

- This means that we pick a word which is likely to occur when we write about the chosen topic (e.g. 'baseball' in sport topic), and write down the word on the document.

- Repeat this process for the number of words in the document.

4 Variational Inference

- The unknown parameters are θ, z. Thus we want to get the following posterior probability

$$p(\theta, Z|W, \alpha, \beta) = \frac{p(\theta, Z, W|\alpha, \beta)}{p(W|\alpha, \beta)}$$

- However, the denominator is intractable because of $\theta_i$andβ_{ij}

$$p(W|\alpha, \beta) = \int \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} (\theta^{a_i - 1}))(\prod_{n=1}^{N} \sum_{i=1}^{k} \prod_{j=1}^{V} (\theta_i \beta_i)^{w_m^i})d\theta$$

- Let's approximate this by $q(\theta, \gamma, \phi) = q(\theta|\gamma) \prod_{n=1}^{N} q(z_n|\phi_n)$.
- We want q to be close to p, but how to measure the closeness of p and q?
- KL-divergence (difference of entropy)

$$KL(q||p) = E_q\Big[log \frac{q(z)}{p(z|x)}\Big]$$

where z is unknown parameter and x is known/observed parameter. The smaller the KL divergence is, the "closeer" p and q are.
- How to minimize KL divergence? - Maxmimize ELBO

$$\begin{aligned}
KL(q||p) &= E_q\Big[log \frac{q(z)}{p(z|x)}\Big] \\
&= E_q\Big[logq(z)\Big] - E_q\Big[logp(z|x)\Big] \\
&= E_q\Big[logq(z)\Big] - E_q\Big[log \frac{p(z,x)}{p(x)}\Big] \\
&= E_q\Big[logq(z)\Big] - E_q\Big[logp(z,x)\Big] + E_q\Big[logp(x)\Big] \\
&= -(E_q\Big[logq(z)\Big] + E_q\Big[logp(z,x)\Big]) + E_q\Big[logp(x)\Big]
\end{aligned}$$

- The first two term are ELBO. So Maximizing ELBO $=$ Minimizing KL divergence.