

Gibbs Sampling for Bayesian Mixture

Carl Edward Rasmussen

November 26th, 2017

Key concepts

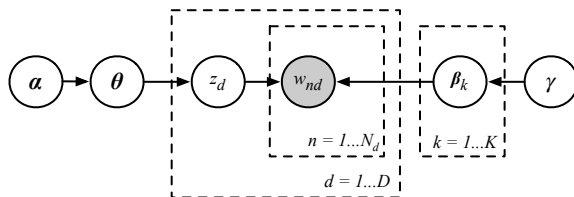
- General Bayesian mixture model
- We derive the Gibbs sampler
- Marginalize out mixing proportions: collapsed Gibbs sampler

Bayesian document mixture model

Our mixture model has observations \mathbf{w}_d the words in document $d = 1, \dots, D$. The parameters are β_k and θ , and latent variables \mathbf{z} .

The mixture model has K components, so the parameters are $\beta_k, k = 1, \dots, K$. Each β_k is the parameter of a categorical over possible words, with prior $p(\beta)$. The discrete latent variables $z_d, d = 1, \dots, D$ take on values $1, \dots, K$.

Note, that in this model the observations are (the word counts of) entire documents.



$$\begin{aligned}\theta &\sim \text{Dir}(\alpha) \\ \beta_k &\sim \text{Dir}(\gamma) \\ z_d | \theta &\sim \text{Cat}(\theta) \\ w_{nd} | z_d, \beta &\sim \text{Cat}(\beta_{z_d})\end{aligned}$$

Bayesian mixture model

The conditional likelihood is for each observation is

$$p(\mathbf{w}_d | z_d = k, \boldsymbol{\beta}) = p(\mathbf{w}_d | \boldsymbol{\beta}_k) = p(\mathbf{w}_d | \boldsymbol{\beta}_{z_d}),$$

and the prior

$$p(\boldsymbol{\beta}_k).$$

The categorical latent component assignment probability

$$p(z_d = k | \boldsymbol{\theta}) = \theta_k,$$

with a Dirichlet prior

$$p(\boldsymbol{\theta} | \boldsymbol{\alpha}) = \text{Dir}(\boldsymbol{\alpha}).$$

Therefore, the latent conditional posterior is

$$p(z_d = k | \mathbf{w}_d, \boldsymbol{\theta}, \boldsymbol{\beta}) \propto p(z_d = k | \boldsymbol{\theta}) p(\mathbf{w}_d | z_d = k, \boldsymbol{\beta}) \propto \theta_k p(\mathbf{w}_d | \boldsymbol{\beta}_{z_d}),$$

which is just a discrete distribution with K possible outcomes.

Gibbs Sampling

Iteratively, alternately, sample the three types of variables:

Component parameters

$$p(\beta_k | \mathbf{w}, \mathbf{z}) \propto p(\beta_k) \prod_{d: z_d = k} p(\mathbf{w}_d | \beta_k),$$

which is now a categorical model, the mixture aspect having been eliminated.

The posterior latent conditional allocations

$$p(z_d = k | \mathbf{w}_d, \boldsymbol{\theta}, \boldsymbol{\beta}) \propto \theta_k p(\mathbf{w}_d | \beta_{z_d}),$$

are categorical and mixing proportions

$$p(\boldsymbol{\theta} | \mathbf{z}, \boldsymbol{\alpha}) \propto p(\boldsymbol{\theta} | \boldsymbol{\alpha}) p(\mathbf{z} | \boldsymbol{\theta}) = \text{Dir}\left(\frac{c_k + \alpha_k}{\sum_{j=1}^K c_j + \alpha_j}\right).$$

where $c_k = \sum_{d: z_d = k} 1$ are the counts for mixture k .

Collapsed Gibbs Sampler

The parameters are treated in the same way as before.

If we **marginalize** over θ

$$p(z_d = k | \mathbf{z}_{-d}, \alpha) = \frac{\alpha + c_{-d,k}}{\sum_{j=1}^K \alpha + c_{-d,j}},$$

where index $-d$ means *all except* d , and c_k are counts;
we derived this result when discussing pseudo counts.

The **collapsed** Gibbs sampler for the latent assignments

$$p(z_d = k | \mathbf{w}_d, \mathbf{z}_{-d}, \boldsymbol{\beta}, \alpha) \propto p(\mathbf{w}_d | \beta_k) \frac{\alpha + c_{-d,k}}{\sum_{j=1}^K \alpha + c_{-d,j}},$$

where now all the z_d variables have become **dependent** (previously they were conditionally independent given θ).

Notice, that the Gibbs sampler exhibits the *rich get richer* property.