Gibbs Sampling for Bayesian Mixture

Carl Edward Rasmussen

November 25th, 2016

Key concepts

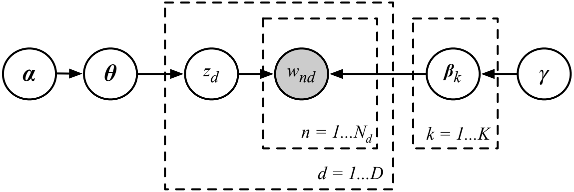
- General Bayesian mixture model
- We derive the Gibbs sampler
- Marginalize out mixing proportions: collapsed Gibbs sampler

Bayesian mixture model

A mixture model has observations $\mathbf{y}$, parameters $\boldsymbol{\beta}$, and latent variables $\mathbf{z}$.

There are N observations, $y_n, n = 1, \ldots N$. The mixture model has K components, so the parameters are $\beta_k, k = 1, \ldots K$ with prior $p(\boldsymbol{\beta})$ and the discrete latent variables $z_n, n = 1, \ldots N$ take on values $1, \ldots K$.

The Bayesian mixture of categoricals is an example (although in this case, the observations are the D documents).



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

Bayesian mixture model

The conditional likelihood is for each observation is

$$p(y_n|z_n = k, \boldsymbol{\beta}) \;=\; p(y_n|\beta_k) \;=\; p(y_n|\beta_{z_n}),$$

and the prior

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

The categorical latent component assignment probability

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

with a Dirichlet prior

$$p(\theta|\alpha) \;=\; \mathrm{Dir}(\alpha).$$

Therefore, the latent posterior is

$$p(z_n = k|y_n, \boldsymbol{\theta}, \boldsymbol{\beta}) \;\propto p(z_n = k|\boldsymbol{\theta})p(y_n|z_n = k, \boldsymbol{\beta}) \;\propto\; \theta_k p(y_n|\beta_{z_n}),$$

which is just a discrete distribution with K possible outcomes.

Gibbs Sampling

Iteratively, alternately, sample the three types of variables:

Component parameters

$$p(\beta_k|\mathbf{y},\mathbf{z}) \;\propto\; p(\beta_k)\prod_{n:z_n=k} p(y_n|\beta_k),$$

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

The latent allocations

$$p(z_n = k|y_n, \theta, \boldsymbol{\beta}) \;\propto \theta_k p(y_n|\beta_{z_n}),$$

and mixing proportions

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

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

Collapsed Gibbs Sampler

The parameters are treated in the same way as before.

If we marginalize over θ

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

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

The collapsed Gibbs sampler for the latent assignements

$$p(z_n = k | y_n, z_{-n}, \boldsymbol{\beta}, \alpha) \;\propto p(y_n | \beta_k) \frac{\alpha + c_{-n,k}}{\sum_{j=1}^{K} \alpha + c_{-n,j}},$$

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

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