Gibbs Sampling for Bayesian Mixture

Carl Edward Rasmussen

November 27th, 2018
Key concepts

- General Bayesian mixture model
- We derive the Gibbs sampler
- Marginalize out mixing proportions: collapsed Gibbs sampler
Our mixture model has observations $w_d$ the words in document $d = 1, \ldots, D$. The parameters are $\beta_k$ and $\theta$, and latent variables $z$.

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

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

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

The conditional likelihood is for each observation is

\[ p(w_d | z_d = k, \beta) = p(w_d | \beta_k) = p(w_d | \beta_{z_d}), \]

and the prior

\[ p(\beta_k). \]

The categorical latent component assignment probability

\[ p(z_d = k | \theta) = \theta_k, \]

with a Dirichlet prior

\[ p(\theta | \alpha) = \text{Dir}(\alpha). \]

Therefore, the latent conditional posterior is

\[ p(z_d = k | w_d, \theta, \beta) \propto p(z_d = k | \theta)p(w_d | z_d = k, \beta) \propto \theta_k p(w_d | \beta_{z_d}), \]

which is just a discrete distribution with \( K \) possible outcomes.
Gibbs Sampling

Iteratively, alternately, sample the three types of variables:

Component parameters

\[ p(\beta_k|w, z) \propto p(\beta_k) \prod_{d:z_d=k} p(w_d|\beta_k), \]

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

The posterior latent conditional allocations

\[ p(z_d = k|w_d, \theta, \beta) \propto \theta_k p(w_d|\beta_{z_d}), \]

are categorical and mixing proportions

\[ p(\theta|z, \alpha) \propto p(\theta|\alpha)p(z|\theta) \propto \text{Dir}(c + \alpha). \]

where \( c_k = \sum_{d:z_d=k} 1 \) are the counts for mixture \( k \).
The parameters are treated in the same way as before.

If we marginalize over $\theta$

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

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

$$p(z_d = k|w_d, z_{-d}, \beta, \alpha) \propto p(w_d|\beta_k) \frac{\alpha + c_{-d,k}}{\sum_{j=1}^{K} \alpha + c_{-d,j}},$$

where now all the $z_d$ variables have become dependent (previously they were conditionally independent given $\theta$).

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