Assignment #4

Due: 23:59pm November 8, 2013

Problem 1 (10 pts)

For target distribution $\pi(x)$ and proposal distribution $q(x' \leftarrow x)$, the Metropolis–Hastings transition operator is given by

$$T(x' \leftarrow x) = q(x' \leftarrow x) \, \min \left\{ 1, \frac{\pi(x') \, q(x \leftarrow x')}{\pi(x) \, q(x' \leftarrow x)} \right\}.$$

Show that this transition operator satisfies detailed balance.

Problem 2 (20 pts)

Here is a hierarchical model that looks like a ten-dimensional "funnel":

$$\begin{aligned} v &\sim \mathcal{N}(v \,|\, 0, 3^2) \\ x_k \,|\, v &\sim \mathcal{N}(x_k \,|\, 0, e^v) \text{ for } k = 1, 2, \ldots, 9. \end{aligned}$$

This model leads to a joint distribution

$$p(v, x_1, x_2, \ldots, x_9) = \mathcal{N}(v \,|\, 0, 3^2) \prod_{k=1}^{9} \mathcal{N}(x_k \,|\, 0, e^v). \tag{1}$$

1. Draw a graphical model describing this joint distribution.
2. Out of 50,000 samples from this joint distribution, how many of them will have $v < -5$?
3. Exploiting the directed structure, generate 50,000 samples from this joint distribution. Plot a histogram of the marginal samples for v. How many samples had $v < -5$?
4. Pretend you don't know about the directed structure and that you just had the joint distribution in Equation 1. Use Metropolis–Hastings to generate 50,000 samples. Use a ten-dimensional Gaussian proposal distribution with identity covariance, centered at the current state. That is, if $\boldsymbol{w} = [v, x_1, \ldots, x_9]$, then

   $$q(\boldsymbol{w}' \leftarrow \boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}' \,|\, \boldsymbol{w}, \mathbb{I}_9).$$

   Start from $v = 0$ and $x_k = 1$. Plot a histogram of v. How many samples had $v < -5$?
5. Implement slice sampling on the same joint distribution. Use any variant of slice sampling you wish (e.g., coordinate-wise, random directions, hyperrectangle, exponential step-out, etc.). Specify which you use. Plot a histogram of v. How many samples had $v < -5$?

In the following two questions, you will implement latent Dirichlet allocation using MCMC and variational inference. You can apply LDA to any data you wish, including the text of the Jester data we have previously studied. Other possible corpora are:

- NIPS papers: `http://cs.nyu.edu/~roweis/data.html`
- Newsgroups: `http://people.csail.mit.edu/jrennie/20Newsgroups/`
- Associated Press: `http://www.cs.princeton.edu/~blei/lda-c/index.html`

Remember, you'll need to do some preprocessing to turn things into word counts. You'll probably want to remove punctuation, tokenize and lowercase. You may also want to remove stop words and very rare words.

In each case, provide some analysis and discussion of your implementation and results. For example: What were the most common topics? What were the most common words across each topic? If you had metadata, how did this data relate to the topics? How many topics did you use and how did you arrive at this choice? What preprocessing did you carry out? How many iterations did you run and how long did this take? Did the implementations find interestingly different solutions? How did you set the hyperparameters? Did you try alternative values? Did you do any quantitative analysis of your model?

Problem 3 (35 pts)

One approach to performing inference in latent Dirichlet allocation is to use Markov chain Monte Carlo, as described in Griffiths and Steyvers (2004). This is an example of Gibbs sampling. Implement this approach and describe what you find.

Problem 4 (35 pts)

Another approach to LDA inference is to use a variational approximation, as described in Blei, Ng and Jordan (2003). This is an example of mean field variational inference. Feeling ambitious? Check out recent results on *stochastic* and *online* variational inference in LDA.