Topics in Probabilistic Modeling and Inference (CS698X)
Homework 4 (Due date: April 20, 2018, 11:59pm)

Instructions

- We will only accept electronic submissions and the main writeup must be as a PDF file. If you are handwriting your solutions, please scan the hard-copy and convert it into PDF. Your name and roll number should be clearly written at the top. In case you are submitting multiple files, all files must to be zipped and **submitted as a single file** (named: your-roll-number.zip). Please do not email us your submissions. Your submissions have to be uploaded at the following link: `https://tinyurl.com/y7pjpnt3`).
- Each late submission will receive a 10% penalty per day for up to 3 days. No submissions will be accepted after the 3rd late day.

Problem 1 (15 marks)

Consider a one-dimensional Gaussian distribution $\mathcal{N}(x|\mu, \tau^{-1})$ that generates N observations $\mathbf{X} = [x_1, \ldots, x_N]$. Assume the mean to have a prior $\mathcal{N}(\mu|\mu_0, s_0)$ where μ_0 and s_0 are the prior's mean and variance, respectively, and the precision to have a prior Gamma$(\tau|a, b)$ where a and b are the shape and rate parameters, respectively.

Derive a Gibbs sampler for this model. In particular, write down the expression for the respective conditional posteriors $p(\mu|\mathbf{X}, \tau)$ and $p(\tau|\mathbf{X}, \mu)$, and sketch the overall Gibbs sampler. Feel free to use Gaussian's properties and results related to conjugate pairs of distributions.

PS - As a side note, you might find this Wikipedia compendium on results related to conjugate priors useful, not just for this problem only but in general: `https://tinyurl.com/on38cle`.

Problem 2 (25 marks)

Consider the Latent Dirichlet Allocation (LDA) model

$$
\begin{aligned}
\beta_k &\sim \text{Dirichlet}(\eta, \ldots, \eta), \quad k = 1, \ldots, K \\
\theta_d &\sim \text{Dirichlet}(\alpha, \ldots, \alpha), \quad d = 1, \ldots, D \\
\boldsymbol{z}_{d,n} &\sim \text{multinoulli}(\theta_d), \quad n = 1, \ldots, N_d \\
\boldsymbol{w}_{d,n} &\sim \text{multinoulli}(\beta_{\boldsymbol{z}_{d,n}})
\end{aligned}
$$

In the above β_k denotes the V dim. topic vector for topic k (assuming vocabulary of V unique words), θ_d denotes the K dime. topic proportion vector for document d, and the number of words in document d is N_d.

Derive a Gibbs sampler to sample the unknowns $\boldsymbol{z}_{d,n}$, θ_d, β_k. In particular, clearly write down their respective conditional posteriors that the Gibbs sampler requires and sketch the overall Gibbs sampler.

Also write down the expressions for the mean of the conditional posteriors of θ_d and β_k. Does it make intuitive sense? Briefly explain your answer.

Problem 3 (25 marks)

Consider modeling an $N \times K$ binary matrix $\mathbf{Z}$ with its binary entries assumed to be generated i.i.d. as follows

$$
\begin{aligned}
\pi_k &\sim \text{Beta}(\alpha/K, 1) \qquad k = 1, \ldots, K \\
Z_{nk}|\pi_k &\sim \text{Bernoulli}(\pi_k) \qquad n = 1, \ldots, N, k = 1, \ldots, K
\end{aligned}
$$

- Derive the expression for $p(\mathbf{Z}|\alpha)$. This expression must not depend on $\{\pi_k\}_{k=1}^K$
- Derive the expression for $p(Z_{nk}|Z_{-nk})$ where Z_{-nk} denotes all the entries of $\mathbf{Z}$, except Z_{nk}. What will $p(Z_{nk}|Z_{-nk})$ be as $K \to \infty$?
- As a function of α, what will be the expected number of ones in each column of $\mathbf{Z}$, and in all of $\mathbf{Z}$?

Problem 4 (15 marks)

Consider the Latent Dirichlet Allocation (LDA) model for documents. Suppose that, in addition to the D documents, you are also provided a document-document binary link matrix $\mathbf{A}$ of size $D \times D$ such that if document d is "linked" to document d' (e.g., if one mentions the other) then $A_{dd'} = 1$, and $A_{dd'} = 0$ otherwise (note that $\mathbf{A}$ is symmetric).

Suggest a generative story for each element (a binary observation) of the $\mathbf{A}$ matrix. Specifically, I am looking for a way to generate $\mathbf{A}$ using the latent variable(s) that are part of the original LDA model we saw in the class. In addition to these latent variables, you may introduce additional parameters if you think it would make the model even better. You only need to give the generative model for A and don't need to give the inference algorithm.