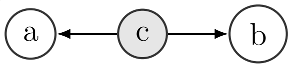
Semestral exam BAYa - 1. 11. 2024 - A,

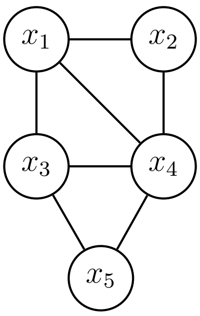
Login: .... **Signature: ....**

---

1. Write an equation expressing that variables a and b are conditionally independent given a variable c. Using basic rules of probability (product rule, sum rule, Bayes rule), provide proof that this equation holds for the Bayesian Network: 

---

2. For the Markov Random Field below, express the joint probability $P(x_1, x_2, x_3, x_4, x_5)$ in terms of potential functions and in terms of energy functions. What are the potential and energy functions? What are their properties? How are they related?



3. What is a "message" in the Belief propagation algorithm? What form does it have for distributions with discrete variables (i.e. how would you represent it when implementing this algorithm)?

4. What does it mean that a prior is conjugate?
Provide mathematical proof that distribution $\text{Gamma}(\lambda|a,b) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}$ is the conjugate prior for the precision parameter λ of the Gaussian distribution with known mean $\mu = 0$, where the likelihood is calculated as

$$\mathcal{L}(\lambda) = P(x_1, x_2, \ldots, x_N|\lambda) = \prod_{n=1}^{N} \mathcal{N}(x_n; 0, \lambda^{-1}) = \prod_{n=1}^{N} \sqrt{\frac{\lambda}{2\pi}} e^{-\frac{\lambda}{2}(x_n-0)^2} = (2\pi)^{-\frac{N}{2}} \lambda^{\frac{N}{2}} e^{-\frac{1}{2}\lambda \sum_{n=1}^{N} x_n^2}.$$

Explain what and why you have derived.

5. What is the Stick breaking process? It describes certain distribution. How does a single sample from this distribution look like? How is it generated? What role does this distribution play in the definition of the Infinite Gaussian Mixture Model?

6. Let us have some observed data $\mathbf{x} = [x_1, x_2, \ldots, x_n]$, which we assume to be samples from a Gaussian distribution with unknown mean μ and variance σ^2. Given the observed data and a prior $p(\mu, \sigma^2)$, we estimate the posterior distribution $p(\mu, \sigma^2|\mathbf{x})$. Now, we generate several samples from the posterior distribution. What does each such sample represent? How are these samples related to the observed data?

7. Variational Bayes (VB) inference makes use of the following equality:

$$\ln p(\mathbf{X}) = \int q(\mathbf{Y}) \ln p(\mathbf{X}, \mathbf{Y}) \, \mathrm{d}\mathbf{Y} - \int q(\mathbf{Y}) \ln q(\mathbf{Y}) \, \mathrm{d}\mathbf{Y} - \int q(\mathbf{Y}) \ln \frac{p(\mathbf{Y}|\mathbf{X})}{q(\mathbf{Y})} \, \mathrm{d}\mathbf{Y}$$

What do the symbols $\mathbf{X}$ and $\mathbf{Y}$ correspond to (in general, not just for GMM training)? What does the term $q(\mathbf{Y})$ represent? What is the purpose of the VB inference? What are we trying to infer and how? How is the above equation used do derive the VB updates? What is estimated in each VB update? Which term from the equation is optimized in each VB update and how?

8. Describe Latent Dirichlet Allocation Model. What is this model good for? Formally describe the generative process for this model (at least for the not fully Bayesian version).

---

9. Consider a PLDA model, as often used to model speaker embeddings in the speaker verification context. We assume that S speaker means were generated from a Gaussian distribution $\mathcal{N}(\mathbf{z}_s; \boldsymbol{\mu}, \boldsymbol{\Sigma}_{ac})$, and then N_s embeddings were generated for each of such speakers from the Gaussian distribution $\mathcal{N}(\mathbf{x}_{sn}; \mathbf{z}_s, \boldsymbol{\Sigma}_{wc})$. Consider the case with only S=3 speakers, where $N_1 = 2$, $N_2 = 3$ and $N_3 = 1$. Draw the Bayesian network that corresponds to this generative process.

   Given such Bayesian network, write the expression for the joint probability $P(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{x}_{11}, \mathbf{x}_{12}, \mathbf{x}_{21}, \mathbf{x}_{22}, \mathbf{x}_{23}, \mathbf{x}_{31})$.

10. Describe the inference for the Bayesian Gaussian Mixture Model based on (non-collapsed) Gibbs sampling. What is this inference used for? Each iteration consists of several steps. Samples from what conditional distributions (use mathematical notation) are drawn in each of these steps? What do these samples represent? What is the problem with this procedure that motivated the introduction of collapsed Gibbs sampling?