Semestral exam BAYa - 4. 1. 2023,

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

---

1. Draw a Bayesian Network with 3 nodes A, B, and C, where random variables A and B are marginally independent, but they are conditionally dependent given variable C (i.e. so-called "explaining away"). Explain intuitively (e.g. on an example) what and how causes the dependency.

---

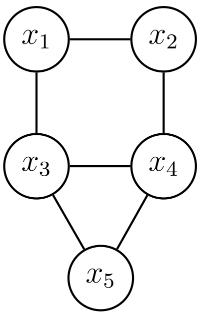
2. Let us have discrete training observations $x_n \in \{0, 1\}$ generated from the Bernoulli distribution $\text{Bern}(x|\mu) = \mu^x(1-\mu)^{1-x}$. Let H and T be the number of observations where $x_n = 1$ and $x_n = 0$, respectively. Given the training data, write the formula for the likelihood function $P(x_1, x_2, \ldots, x_N|\mu)$. Provide mathematical proof that the Beta distribution $\text{Beta}(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}$ is the conjugate prior for the parameters of the Bernoulli distribution. What does it mean that a prior is conjugate? Explain!

3. What is the posterior predictive distribution? How is it evaluated? Provide the basic general formula and interpret it. Now, write again the formula for evaluating the posterior predictive distribution, but this time specifically for a Gaussian distributed variable.

4. How can we see that some probability distribution is from the exponential family? What form do we have to be able to write the distribution in? What do we need for obtaining Maximum likelihood estimates of parameters for such ditribution? What are the advantages of using such distributions in Bayesian inference?

5. What is the max-product message-passing algorithm? How is it different from the sum-product algorithm? What problems does it solve? What limitations does it have? What is the backtracking used for in this algorithm?

6. 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?



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 or LDA 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 to 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 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?

9. Draw the Bayesian Network for the full Bayesian Latent Dirichlet Allocation Model. Write the corresponding factorization for the joint distribution of the observed variables w_{dn} and all hidden variables. What distributions would you use for the individual factors in order to allow for a tractable and efficient inference (e.g. using Variational Bayes or Gibbs sampling)? Describe the assumed process (steps) of generating the observed variables w_{dn}. If you cannot describe the full Bayesian LDA, try to describe at least the non-Bayesian one.

10. Let π_c be the probabilities of the categories $c = 1..C$. The Maximum Likelihood estimates of these probabilities as parameters of the categorical distribution are: $\pi_c^{ML} = N_c/N$, where N_c is the number of training observations from category c and N is the total number of training observations. Outline, how do we derive this formula? What is the objective function that we try to maximize? The full derivation will earn a bonus point.