Semestral exam BAYa - 17. 1. 2025,

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

1. For a Bernoulli likelihood $P(X|\mu) = \mu^H(1-\mu)^T$, the conjugate prior is distribution $\mathrm{Beta}(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}$.
   Explain why it is useful to have a conjugate prior.
   Describe how the hyperparameters a and b affect the prior and posterior distributions.

2. Describe (with mathematical notation) the assumed generative process of the PLDA model (i.e. how the training data was generated). Draw the Bayesian network that illustrates such process using plate notation.

3. Give an example of three random variables A, B, and C such that A and B are marginally independent but conditionally dependent given C. Verify this property using probability rules.

4. 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?

5. Compare the Frequentist and Bayesian approaches to parameter estimation. Use the problem of estimating the parameters of a Gaussian distribution as an example, highlighting the role of prior knowledge.

6. What is the Stick Breaking Process (GEM)? How does a single sample from a distribution modeled by GEM look? The GEM has a single parameter called the concentration parameter. How does this parameter affect samples from GEM, and why?

7. Expectation Maximization (EM) algorithm makes use of the following equality:

$$\ln p(\mathbf{X}|\boldsymbol{\eta}) = \sum_{\mathbf{Z}} q(\mathbf{Z}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\eta}) - \sum_{\mathbf{Z}} q(\mathbf{Z}) \ln q(\mathbf{Z}) - \sum_{\mathbf{Z}} q(\mathbf{Z}) \ln \frac{p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\eta})}{q(\mathbf{Z})}$$

How is this equation used in the EM algorithm? What the symbols $\mathbf{X}, \mathbf{Z}$ and $\boldsymbol{\eta}$ correspond to (in general, not just for GMM training)? What does the term $q(\mathbf{Z})$ represent? How is $q(\mathbf{Z})$ set in the E-step and why?? How is $q(\mathbf{Z})$ used in the M-step? Which term from the equation is optimized in the M-step and how?

8. What is Factor Graph? What does it represents and how is it different form or similar to MRF or BN?

9. Draw the Bayesian Network for the full Bayesian Latent Dirichlet Allocation (LDA) Model. Write the corresponding factorization for the joint distribution of the observed variables w_{dn} and all hidden variables. Explain which distributions you would use for the individual factors and why, to enable 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. Consider the factorization $p(x_1, x_2, x_3, x_4, x_5) = p(x_1)p(x_2)p(x_3|x_1, x_2)p(x_4|x_3)p(x_5|x_3, x_4)$. Consider that all the random variables x_i are discrete and that we know all the distributions corresponding to the individual factors (i.e. we have the corresponding tables with probabilities). Let the symbol $\sum_{x_i}$ represent the sum over all possible values of the random variable x_i. Using mathematical notation, express how can the following probabilities be inferred most efficiently (i.e. use the right order of sums and/or brackets). Notice that not all the factors (probability tables) are necessary to evaluate some of the following probabilities.

    (a) $p(x_2|x_1)$
    (b) $p(x_3|x_1)$
    (c) $p(x_1|x_3)$
    (d) $p(x_1|x_2, x_3)$
    (e) $p(x_5)$
    (f) $p(x_4, x_5|x_3)$