Midterm exam BAYa, group A - 16. 11. 2022,

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

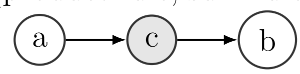
1. Let π_c be the probabilities of the categories $c = 1..C$. The Maximum Likelihood estimates of these probabilities as parameters of the categorical distribution are: $\pi_c^{ML} = N_c/N$, where N_c is the number of training observations from category c and N is the total number of training observations. 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.

2. Draw the Bayesian Networks for the Gaussian Mixture Model (GMM). Write the corresponding factorization for the joint distribution of the observed and hidden random variables x and z. Explain what distributions the individual factors correspond to. From this factorization, derive the equation for evaluating the GMM probability density function (i.e. the marginal probability density $p(x)$).

3. 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? How is $q(\mathbf{Z})$ used in the M-step? Which term from the equation is optimized in the M-step and how?

4. 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: 

5. Draw the Markov Random Field graphical model for the distribution defined as

$$P(x_1, x_2, x_3, x_4, x_5) = \frac{1}{Z}\psi_a(x_1, x_2)\psi_b(x_1, x_3)\psi_c(x_2, x_4)\psi_d(x_3, x_4, x_5),$$

where ψ are some positive (potential) functions and Z is an appropriate normalizing constant.

---

6. For each of the following statements, say whether the statement holds true for the Markov Random Field from the previous questions and explain why.

(a) $p(x_1, x_5|x_3) = p(x_1|x_3)p(x_5|x_3)$

(b) $p(x_1, x_5|x_3, x_4) = p(x_1|x_3, x_4)p(x_5|x_3, x_4)$

(c) $p(x_3, x_4|x_1, x_5) = p(x_3|x_1, x_5)p(x_4|x_1, x_5)$

(d) $p(x_2, x_3|x_1, x_5) = p(x_2|x_1, x_5)p(x_3|x_1, x_5)$

7. What is sum-product message passing algorithm (or Belief Propagation)? What problems does it solve? What limitations does it have? What is the fundamental idea behind the sum-product message passing algorithm that makes it more efficient than the brute-force approach?

---

8. Consider the factorization $p(x_1, x_2, x_3, x_4, x_5) = p(x_1)p(x_2)p(x_3|x_1)p(x_4|x_1)p(x_5|x_4, x_2)$. 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 the following probabilities.

   (a) $p(x_3)$

   (b) $p(x_5)$

   (c) $p(x_1|x_2)$

   (d) $p(x_1|x_3)$

   (e) $p(x_3, x_4)$

   (f) $p(x_3, x_4|x_1)$