CS-E4820 Machine Learning: Advanced Probabilistic Methods, Exam
Reponsible teacher: Pekka Marttinen.

Date: April 13th, 2021, from 13:00 to 17:00 o'clock.

You have 3.5 hours for the exam and 0.5 hours for submitting it. All answers must be written on pen and paper and converted into a PDF. The submission must be done as a single PDF in MyCourses and **the deadline is at 17:00 pm**. You may use a scientific calculator and all materials provided on the course: lecture slides, videos, assignments, and model solutions. Use of other materials and communicating with other students by any means during the exam is not allowed. For more information, see the exam information announcement in MyCourses.

Details about grading can be found in the slides of the first lecture. If you have done some exercises last year, and wish those to be taken into account, mention this on the first page of your exam. This exam consists of two sheets. **Required distributions are given in the end of the 2nd sheet.**

Q1) Bayes' rule

Consider three binary variables x_1, x_2, and x_3. Their joint distribution factorizes as $p(x_1, x_2, x_3) = p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_2)$, where $p(x_1 = 0) = 0.4$, $p(x_2 = 0 \mid x_1 = 0) = 0.7$, $p(x_2 = 0 \mid x_1 = 1) = 0.5$, $p(x_3 = 0 \mid x_2 = 0) = 0.9$, and $p(x_3 = 0 \mid x_2 = 1) = 0.2$.
A) Draw the DAG corresponding to the model. (1p)
B) Compute $p(x_2 = 1 \mid x_3 = 1)$. (5p)

Q2) Bayesian networks

List all pairs of variables that are d-separated in the DAG in Fig. 1; for each pair of d-separated variables, give one set that d-separates those variables. (4p)

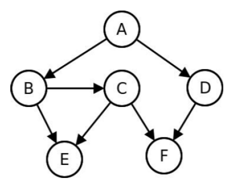

Figure 1

Q3) Variational inference

Assume that N observations x_n, $n = 1, \dots, N$ have been generated from the following mixture model:

$$p(x_n|\tau, \lambda_1, \lambda_2) = \tau N(x_n|0, \lambda_1^{-1}) + (1 - \tau)N(0, \lambda_2^{-1}),$$

where λ_1 and λ_2 are the unknown precisions (inverse variances) of the two components, and τ is the mixing coefficient. Assume the following prior distributions:

$$\tau \sim Beta(\alpha_0, \alpha_0), \quad \lambda_1 \sim Gamma(a_0, b_0), \quad \lambda_2 \sim Gamma(c_0, d_0).$$

A) Define the model using latent variables $\mathbf{z} = \{z_i\}_{i=1}^N$. (1p)
B) Derive the variational update for λ_2. You can assume the mean-field approximation:

$$q(\mathbf{z}, \tau, \lambda_1, \lambda_2) = q(\lambda_1)q(\lambda_2)q(\tau)\prod_n q(z_n)$$

and assume the other factors are given by

$$q(\tau) = Beta(\tau|\alpha_n, \beta_n), \quad q(z_{n1}) = Bernoulli(z_{n1}|r_{n1}), \quad q(\lambda_1) = Gamma(\lambda_1|a_n, b_n).$$

(5p)

Q4) EM algorithm

Consider N observations x_n, $n = 1, \ldots, N$, from a two-component mixture of binomial distributions

$$p(x_n \mid \theta, q_1, q_2) = \theta Bin(x_n \mid q_1) + (1-\theta) Bin(x_n \mid q_2).$$

A) Represent the model using latent variables and derive the E step of the expectation maximization. In the end, simplify the Q-function, $Q(\theta, q_1, q_2 | \theta^0, q_1^0, q_2^0)$, where θ^0, q_1^0, q_2^0 are the current values of the parameters. (4p)
B) Derive the M-step for the θ parameter. (2p)
The binomial distribution has a probability mass function of the form

$$f(k|m, q) = p(x_n = k) = \binom{m}{k} q^k (1-q)^{m-k},$$

where $0 \leq k \leq m$ is an integer. You can treat m as a known constant.

Q5) Stochastic variational inference

Explain in your own words, using examples and formulas when needed, the following concepts.
A) The difference between variational parameters, model parameters, and prior parameters. (3p)
B) Reparametrization trick. (3p)

Distribution reference

$$N(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad \text{(Gaussian)}$$

$$\text{Gamma}(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \quad a > 0, b > 0, x > 0$$

$$\text{Bernoulli}(k|p) = \begin{cases} p, \text{if } k = 1 \\ 1-p, \text{ if } k = 0 \end{cases}, \quad k \in \{0, 1\},\ 0 \leq p \leq 1.$$

$$\text{Beta}(x|a, b) = \frac{\Gamma(a+b)}{\Gamma(a) + \Gamma(b)} x^{a-1} (1-x)^{b-1}, \quad x \in [0, 1], a > 0, b > 0, \Gamma \text{ is the Gamma function.}$$