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

Date: June 3rd, 2021, from 16:00 to 20:00 o'clock.

You have 3.5 hours for the exam and 0.5 hours for submitting it. **The exam must be done with pen and paper** (or with pen on a device with a touch-screen) **and converted into a PDF. Answers that are not hand-written will not be graded**. The submission must be done as a single PDF in MyCourses and **the deadline is at 20:00 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) Bayesian networks

A) Are the following conditional independence statements true or false for variables in Fig. 1 (on last page)? Justify your answer by specifying paths between the variables and the blocking variables (if any). (correct answer and justification: 1p per question).

1. tuberculosis $\perp\!\!\!\perp$ smoking | shortness of breath
2. lung cancer $\perp\!\!\!\perp$ bronchitis | smoking
3. visit to Asia $\perp\!\!\!\perp$ smoking | lung cancer, shortness of breath

B) Compute the probability $p(F = 0|S = 0)$, where the structure of the model and the required conditional probabilities are specified in Fig. 2. All variables are assumed to have two possible states, 0 and 1.(3p)

Q2) EM algorithm

Consider N i.i.d. observations x_n, $n = 1, \dots, N$, from a two-component mixture model of exponential distributions

$$p(x_n|\theta, \lambda_1, \lambda_2) = \theta \operatorname{Exp}(x_n|\lambda_1) + (1-\theta) \operatorname{Exp}(x_n|\lambda_2)$$

with parameters $(\theta, \lambda_1, \lambda_2)$.
A) Represent the model using latent variables and derive the Q-function of the EM algorithm. (4.5p)
B) Derive the M step update for the λ_1 parameter. (1.5p)

Q3) Laplace approximation

Approximate the Gamma$(x|a, b)$ distribution with parameters a and b using the Laplace approximation, i.e., the approximating distribution is a Gaussian centered at the mode of the original distribution. Parameters a and b are known constants. Hint: use $E(x) = -\log Gamma(x|a, b)$ as the starting point. (6p)

Q4) Variational Bayes

Suppose you are given data $(y_n, \mathbf{x}_n)$, where $y_n \in \mathbb{R}$ and $x_n \in \mathbb{R}^2$ for all $n = 1, \dots, N$. We model this using a linear regression model

$$y_n = ax_{n1} + bx_{n2} + \epsilon_n, \quad n = 1, \dots, N,$$

where

$$\epsilon_n \overset{i.i.d}{\sim} N(0, 1).$$

Prior distributions for the parameters are

$$a \sim N(0,1), \text{ and}$$
$$b \sim N(0,1).$$

Assume a variational distribution $q(a,b) = q(a)q(b)$ for the parameters of the model, where the factors are assumed to be of the form

$$q(a) = N(a|\mu_a, \sigma_a^2)$$
$$q(b) = N(b|\mu_b, \sigma_b^2).$$

Derive the variational update for factor $q(a)$. (6p)

Q5) Miscellaneous

Briefly (max. 4 sentences each) explain the terms/concepts and their usage/relevance in the context of the course (2p each).

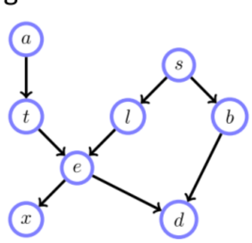
1. Kullback-Leibler divergence
2. ML-II
3. Bayes factor

Distribution reference

$$N(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad \text{(Gaussian)}$$
$$\text{Gamma}(x|a,b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \quad a>0, b>0, x>0 \quad \text{(Gamma)}$$
$$\text{Exp}(x|\lambda) = \lambda e^{-\lambda x}, \quad x \in [0,\infty), \quad \lambda > 0 \quad \text{(Exponential)}$$

Figures

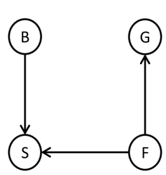
Fig. 1



x = Positive X-ray
d = Dyspnea (Shortness of breath)
e = Either Tuberculosis or Lung Cancer
t = Tuberculosis
l = Lung Cancer
b = Bronchitis
a = Visited Asia
s = Smoker

(Barber, Fig. 3.15)

Fig. 2



P(B=1)
0.95

P(F=1)
0.8

F	P(G=1|F)
0	0.01
1	0.95

B	F	P(S=1|F,B)
0	0	0.001
1	0	0.01
0	1	0.02
1	1	0.98