CS-E4820 Machine Learning: Advanced Probabilistic Methods
Marttinen, Afrabandpey, Blomstedt, Järvenpää, Kangas, Niinimäki, Tran Quang
Examination, at 13:00 o'clock, April 3rd, 2018.

In order to pass the course, you must also complete 1/3 of the exercises. More 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 in grading, mention this on the first page of your exam. Results of this examination are valid for one year.

Allowed equipment: 1) A laptop, ipad, or similar, with which you can read PDFs. The device **must be disconnected** for the duration of the whole exam, i.e., turn off wifi, Blutooth etc. 2) Any documents you can find under 'Materials' and 'Assignments' in myCourses, excluding 'Additional reading'. These must have been **downloaded before the exam**. Alternatively, it is possible to take the same material as printed on paper. 3) Calculator with memory erased. Calculators with ability for 'symbolic calculation' (i.e. ability to simplify formulas, integrals, etc.) are not allowed. 4) other conventional equipment: pencil, eraser,...

This exam consists of two sheets, both of which must be returned at the end of the exam session. **The required distributions are given in the end of the 2nd sheet.**

1) Bayesian networks

A) Are the conditional independence statements below always 'true' for a Bayesian network with structure shown in Figure 1? Justify your answer by specifying paths between the variables and the blocking variables (if any). (correct answer and justification: 1.5p per question).

1. $x_2 \perp\!\!\!\perp x_6 \mid x_5, x_1$
2. $x_2 \perp\!\!\!\perp x_6 \mid x_5, x_3, x_7$

B) In this question you must model a problem with 4 binary variables: G ('gray'), V ('Vancouver'), R ('rain') and S ('sad'). Consider a Bayesian network for these variables with structure and conditional distributions as shown in Figure 2. Write down an expression for $P(S = 1|V = 1)$ in terms of $\alpha, \beta, \gamma, \delta$. (3p).

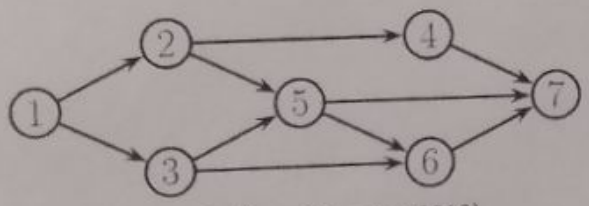

Figure 1 (from Murphy, 2012)

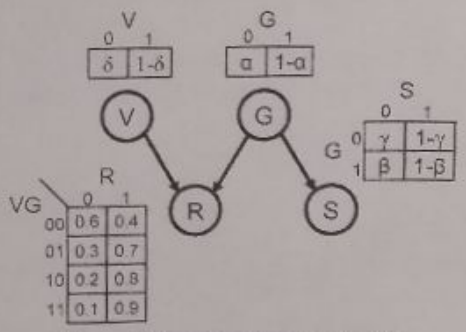

Figure 2 (from Murphy, 2012)

2) EM algorithm

Consider N observations x_n, $n = 1, \ldots, N$, assumed to be i.i.d. from a mixture of Poisson distributions:

$$p(x_n|\pi, \lambda) = \sum_{k=1}^{K} \pi_k \, \text{Poisson}(x_n|\lambda_k).$$

Represent the model using latent variables and derive the E step of the expectation maximization algorithm, which could be used to learn the maximum likelihood estimates for the parameters $\pi = (\pi_1, \ldots, \pi_K)^T$ and $\lambda = (\lambda_1, \ldots, \lambda_K)^T$. (6p)

3) Laplace approximation

Approximate the Beta distribution with parameters a and b, $\text{Beta}(x|a,b)$, 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, and you can assume that $a > 1$, and $b > 1$, such that the Beta distribution has a mode in the interval $(0,1)$. Hint: use $E(x) = -\log Beta(x|a,b)$ as the starting point. (6p)

4) Variational Bayes

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

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

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)

5) Edward

A) Write Edward code for Model and Inference descriptions for the regression model used in Question 4. (3p)
B) Briefly explain the idea of black-box variational inference and how it differs from the 'traditional' variational inference. What are the strengths and weaknesses of the two approaches? (3p)

Distribution reference

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

$$N_k(x|\mu,\Sigma) = (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)} \quad \text{(Multivariate Gaussian)}$$

$$\text{Uniform}(x|a,b) = \begin{cases} 1/(b-a), & \text{if } x \in [a,b] \\ 0, & \text{otherwise} \end{cases}$$

$$\text{Exp}(x|\lambda) = \lambda e^{-\lambda x}, \quad x \in [0,\infty), \quad \lambda > 0 \quad \text{(Exponential)}$$

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

$$\text{Poisson}(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0,1,2,\ldots, \quad \lambda > 0$$

$$\text{Beta}(x|\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, \quad x \in (0,1), \alpha > 0, \beta > 0, B(\alpha,\beta) \text{ is the 'beta function'}.$$