CS-E4820 Machine Learning: Advanced Probabilistic Methods
Examination, from 13:00 to 17:00 o'clock, April 9th, 2019. Reponsible teacher: Pekka Marttinen.

To pass the course, 1/3 of exercises must be completed. 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. 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 three sheets, both of which must be returned at the end of the exam session; copying the exam by any means is not allowed. **Figures and required distributions are given on the 3rd sheet.**

Q1) Bayes' rule

A) Suppose a distribution factorizes according to the graph shown in Fig. 1, and the domains of the variables are as follows: $dom(X) = \{x_1, x_2\}$, $dom(Y) = \{y_1, y_2\}$, $dom(Z) = \{z_1, z_2\}$. In addition, the following conditional probabilities are known:

$$p(x_2) = 0.8, \quad p(z_1) = p(z_2) = 0.5, \quad p(y_2|x_2, z_1) = 0.9, \quad p(y_2|x_2, z_2) = 0.2.$$

Compute $p(z_2|y_2, x_2)$. (3p)

B) Briefly explain *conjugate priors* and their relevance and usage in the context of the course. (3p)

Q2) Bayesian networks

A) Are the following statements true or false for the graph in Fig. 2? Justify your answer by specifying the paths between the variables and the blocking variables (if any). (correct answer and justification: 1p per question).

1. C and G are d-separated by $\{B, D\}$.
2. A and C are d-separated by $\emptyset$.
3. A and D are d-separated by B.

B) How many graphs are there which are Markov equivalent to the graph in Figure 2? Justify your answer. (3p)

Q3) Variational Bayes

Suppose our data $\mathbf{x} = (x_1, \ldots, x_N)$ consist of N observations drawn independently from the normal distribution

$$x_i \sim N(\mu, \tau^{-1}), \text{ for } i = 1, \ldots, N.$$

We assume the following prior on the parameters

$$p(\mu) = N(\mu|\mu_0, \lambda_0^{-1})$$
$$p(\tau) = \text{Gamma}(\tau|a_0, b_0).$$

Derive the variational update for factor $q(\tau)$, when we assume that the posterior distribution $p(\mu, \tau|\mathbf{x})$ is approximated using a factorized distribution $q(\mu, \tau) = q(\mu)q(\tau)$. You can assume that the current factor for μ is

$$q(\mu) = N(\mu|\mu_*, \sigma_*^2).$$

(6p)

Hint 1: $Var(X) = E(X^2) - E(X)^2$.
Hint 2: The Gamma prior is conjugate here.

Q4) Black-box variational inference

Assume that N observations $x_n, n = 1, \dots, N$ have been generated from the model in Fig. 3 with some conditional distributions $p(\lambda_1)$, $p(\lambda_2)$, $p(\lambda_3|\lambda_2)$, $p(z_n|\lambda_1)$, $p(x_n|z_n, \lambda_3)$. Assume that the variational approximation is

$$q(\lambda_1, \lambda_2, \lambda_3, z_1, \dots, z_n) = q(\lambda_1|\theta_1)q(\lambda_2|\theta_2)q(\lambda_3|\theta_3)\prod_{n=1}^{N} q(z_n|\eta_n),$$

where $\theta_1, \theta_2, \theta_3, \eta_1, \dots, \eta_N$ are variational parameters.

A) Write and simplify the formula to calculate the ELBO for the model in Figure 3. (2p)

B) Explain the idea of black-box variational inference, and compare it with the 'standard' variational inference. What are the advantages and disadvantages? (2p)

C) Using generic notation, the gradient of the ELBO can be written as:

$$\nabla_\lambda L = E_{q(z|\lambda)}[\nabla_\lambda \log q(z|\lambda)(\log p(x, z) - \log q(z|\lambda))]. \tag{1}$$

Write and simplify the following terms in Equation (1): i) $\log p(x, z)$, ii) $\log q(z|\lambda)$, $\nabla_\lambda \log q(z|\lambda)$ for the model specified in Figure 3. (2p)

Q5) EM algorithm

Consider a simple factor analysis model:

$$\mathbf{x}_n \sim N_2(\mathbf{w}z_n, \sigma^2 I), \quad n = 1, \dots, N,$$
$$z_n \sim N(0, 1), \quad n = 1, \dots, N,$$

where $\mathbf{x}_n \in R^2$ and $z_n \in R$ for all $n = 1, \dots, N$. Parameters of the model are the loading matrix (a vector in this case), $\mathbf{w} \in R^2$, and variance $\sigma^2 \in R$.

A) Derive and simplify the complete data log-likelihood. (2p)

B) It can be shown that the posterior distribution $p(z_n|\mathbf{x}_n, \mathbf{w}_0, \sigma_0^2)$, where $\mathbf{w}_0, \sigma_0^2$ are current estimates of the parameters, is a Gaussian $N(z_n|\mu_n, \sigma_z^2)$ with certain μ_n and σ_z^2. Derive formulas for μ_n and σ_z^2. (2p)

C) Derive the Q function needed in the E step of the EM algorithm. (2p)

Hint 1: You can solve C even if you did not solve B, i.e., the solution to C can be given using μ_n and σ_z^2.

Hint 2: Completing the square.

Figures

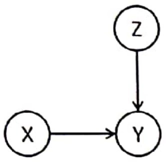

Figure 1

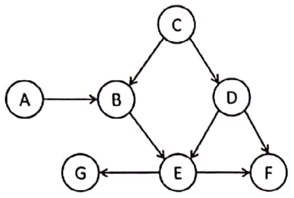

Figure 2

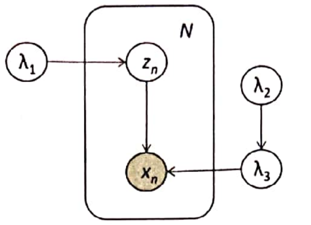

Figure 3

Distribution reference

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

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

$$N_k(x|\mu,\sigma^2 I) = (2\pi)^{-\frac{k}{2}} \sigma^{-k} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^T(x-\mu)\right\} \quad \text{(MVN with diagonal covariance)}$$

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