T-61.5140 Machine Learning: Advanced Probabilistic Methods
Marttinen, Remes
Examination, April 9th, 2015 at 9:00 o'clock.

In order to pass the course, you must also pass the term project. Results of the examination are valid for one year after the examination date. In the exam, you are allowed to have with you 1) printed out lecture slides with notes written by you, 2) printed out PDFs for the 'simple examples' (simple_vb_example, simple_em_example), 3) calculator with memory erased. Note that the distributions required in the exam are provided at the end of the exam sheet.

1) Belief (Bayesian) networks

A) True or false? Justify your answer briefly, in max 2 sentences. (correct answer and justification: 1p per question).

1. C and E are d-separated by $\{A, F\}$ in Fig. 1.
2. A and G are d-separated by $\{C, E\}$ in Fig. 1.
3. The Markov equivalence class to which the graph in Fig. 2 belongs has three members.

B) Answer the questions (1.5p each)

1. Factorize the probability distribution $p_{A,B,C,D}(a, b, c, d)$ according to the graph in Fig. 2.
2. Write down the formula to compute $p_{C|A,B,D}(c|a, b, d)$ for the distribution represented by the graph in Fig. 2.

Fig. 1

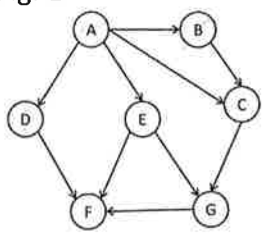
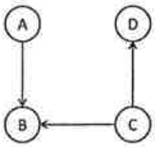


Fig. 2



2) EM algorithm

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

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

Represent the model using latent variables and derive the E and M steps of the expectation maximization algorithm to learn the maximum likelihood estimates of the parameters $(\theta, \lambda_1, \lambda_2)$.

3) Variational approximation

A) Compute the Kullback–Leibler divergence $KL(q, p)$ between $q(x) = \text{Uniform}(x|a, b)$ and $p(x) = N(x|0, 1)$.
B) Approximate the Gaussian $N(0, 1)$ distribution using a variational approximation with approximating distribution $q(x) = \text{Uniform}(x|a, b)$. The distribution q has two parameters, a and b, which you have to optimize.

4) Gibbs sampling

Consider the factor analysis model

$$\begin{aligned}
\mathbf{x}_n &\sim \mathcal{N}_D\left(\mathbf{W}\mathbf{z}_n, \mathrm{diag}(\psi)^{-1}\right), \quad n = 1, \dots, N \\
\psi_d &\sim \mathrm{Gamma}(a, b), \quad d = 1, \dots, D \\
\mathbf{W}_k &\sim \mathcal{N}_D(\mathbf{0}, \alpha \mathbf{I}), \quad k = 1, \dots, K \\
\mathbf{z}_n &\sim \mathcal{N}_K(\mathbf{0}, \mathbf{I}), \quad n = 1, \dots, N,
\end{aligned}$$

where $\mathbf{W}_k$ denotes the loadings for the kth factor and ψ_d^{-1} is the specific noise variance of the dth observed variable. Furthermore let D denote the number of observed variables (i.e. $\mathbf{x}_n \in R^D$), N the number of data points, and K the number of factors in the model. $\mathrm{diag}(\psi)$ is a diagonal matrix with elements $\psi = (\psi_1, \dots, \psi_D)^T$ on the diagonal.

A) Write down pseudo-code for the Gibbs sampler to generate samples from the posterior distribution $p(\psi, \mathbf{W}, \mathbf{z}|\mathbf{x})$, where we have denoted $\mathbf{z} = (z_1, \dots, z_n)$ and $\mathbf{x} = (x_1, \dots, x_n)$. The exact forms of the distributions are *not* required here.

B) Derive the conditional distribution $p(\psi_d|\psi_{-d}, \mathbf{W}, \mathbf{z}, \mathbf{x})$ required in the Gibbs sampler. Here ψ_{-d} denotes vector ψ from which the dth element has been removed. *Hints*: start by writing the likelihood proportionally s.t. all terms not dependent on ψ_d have been discarded. Note that a multivariate Gaussian with a diagonal covariance matrix can be expressed as a product of univariate Gaussian distributions.

5) Miscellaneous

Briefly (max. 4 sentences each) explain the terms/concepts and their usage/relevance in the context of the course (1.5p each).

1. proposal distribution
2. rotation invariance in the factor analysis models
3. marginal likelihood
4. Markov chain Monte Carlo (MCMC)

Distribution reference

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

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

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

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

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