T-61.5140 Machine Learning: Advanced Probabilistic Methods
Hollmén, Raiko
Examination, 28th of August, 2009 from 9 to 12 o'clock.

In order to pass the course and earn 5 ECTS credit points, you must also pass the term project. Results of this examination are valid for one year after the examination date. Information for Finnish speakers: Voit vastata kysymyksiin myös suomeksi, kysymykset on ainoastaan englannin kielellä. Information for Swedish speakers: Du får också svara på svenska, frågorna finns dock endast på engelska.

1. Define the following terms shortly:
a) Bayesian network
b) first-order Markov chain
c) conditional independence
d) Hammersley-Clifford theorem
e) Restricted Boltzmann machine
f) likelihood function

2. Derive the Bayes's rule from the definition of conditional independence of discrete random variables X and C and name the parts.

3. For the Bayesian network that decomposes the joint probability as in $p(x_1, \ldots, x_5) = p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2, x_3)p(x_5|x_4)$, draw the corresponding graphical representation. Assuming all the variables have discrete values $x_i \in \{0, 1, 2\}$, give the sizes of the tables representing the probabilities for the conditional distributions. Moreover, derive the junction tree representation (and name the steps). Draw the resulting junction tree.

4. Write the probability $p(\mathbf{x})$ for the finite mixture model of exponential distributions, name the parts of the mixture model, and derive the E-step and the M-step of the Expectation-Maximization (EM) algorithm. Hint: The probability for an exponentially distributed random variable can be calculated with the following equation: $p(x \mid \lambda) = \lambda e^{-\lambda x}, x \geq 0$.

5. For a probabilistic PCA model for 2-dimensional data with 1 component $\mathbf{x} = \mathbf{a}s + \mathbf{b} + \boldsymbol{\epsilon}$, $p(\boldsymbol{\epsilon}) = N(\epsilon; \mathbf{0}, v\mathbf{I})$, and $p(s) = N(s; 0, 1)$, parameters are fixed to $a_1 = 2$, $a_2 = 1$, $b_1 = 0$, $b_2 = 1$, and $v = 2$. When doing Metropolis algorithm for inference with an observation $x_1 = x_2 = 2$, with what probability is a proposed jump from $s = 1$ to $s = 0$ accepted?