T.61.5140 Machine Learning: Advanced Probabilistic Methods
Hollmén, Raiko (Spring 2008)
Exam requirements for the course, 24th of April, 2008
`http://www.cis.hut.fi/Opinnot/T-61.5140/`

The central topics and reading material will be listed in the following. The course-book [1] forms the core material on the course. It is be complemented with articles covered during the course [2] and lectures, lecture notes, exercises and solutions to the exercises.

1 Graphical models

Graphical models are used to represent complex probability distributions. They have a graphical representation, which can used to visualize the model, also conditional independence properties can be read from this representation. Factorization of the joint probability distribution. Independence, conditional independence, d-separation, Naive Bayes. Inference in graphical models, **Bayes's theorem** (yes, this is absolutely central to everything on the course), sum-product algorithm, junction tree algorithm, tree-width, relation between undirected and directed graphs. Textbook material in [1]: Chapter 8

2 Mixture Models and EM

Relationship with k-means clustering (non-probabilistic) to Naive Bayes or mixture models (probabilistic), mixtures of X (X=any distribution), Gaussian mixtures and mixtures of Bernoulli distributions serve as showcases for continuous and discrete distributions. Likelihood function, incomplete data, complete-data likelihood, E-step (inference), M-step (parameter update), EM algorithm and its derivation. Mixture models are really just simple graphical models, so everything in the Chapter 8 in the textbook [1] applies here: try to see and understand the connections: This course is about inference and learning. Textbook material in [1]: Chapter 9

3 Sequential data

Sequences (such as DNA fragments) and time-series (measurements of outside temperatures). What are the useful independence assumptions when the data in (temporally) dependent? First-order Markov chain and the assumptions made. Limited memory or the conditional independence assumptions that apply. Hidden Markov models (HMM): representation (structure), inference and learning. Forward-backward algorithm (inference) and Baum-Welch (learning)· Extensions to HMM. Viterbi algorithm. Textbook material in [1]: Chapter 13, Sections 13.1 and 13.2, pp. 605-634

4 Approximate Inference

Why is inference sometimes difficult or even not feasible (earlier material), different ways to perform approximate inference. Variational approximations, factorized distributions or (over)simplification of the structure of the underlying graphical model.

Variational mixture models. Textbook material in [1]: Chapter 10, Sections 10.1, 10.2, pp. 461-485

5 Sampling Methods

Sampling as a solution to approximate a posterior distribution or an expectation of a function with regard to the posterior. Random numbers, sampling from a standard distribution, proposal distribution, rejection sampling, importance sampling, MCMC (Markov-Chain Monte-Carlo) methods, Markov chain and the state of the proposal distribution, Metropolis algorithm, Metropolis-Hastings algorithm, Gibbs sampling, a little bit of theory of Markov chains (the ground for convergence of sampling methods). Textbook material in [1]: Chapter 11

6 Related material: PCA and probabilistic PCA

The classic PCA has been taught in earlier courses (prerequisites for this course), probabilistic formulation of it. Textbook material in [1]: Chapter 12, Sections 12.1, 12.2, pp. 559-585

7 Examinations during Spring period 2008

The first opportunity to take the examination during Spring term 2008 is on Thursday, 15th of May, at 9-12 in the lecture hall T1. Registration for the examination is mandatory. Not showing up despite of registration is definitely bad behavior. For possible changes in the examination schedule, see: `http://tieto.tkk.fi/Opinnot/`, this information is from the 24th of April, 2008.

References

[1] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York, 2006.

[2] Michael I. Jordan. Graphical models. *Statistical Science*, 19(1):140–155, 2004.