Advanced Course in Machine Learning: Exam

1 Passing the course

Passing the course requires a combination of exam and home exercises. If you have completed the weekly exercises (receiving at lesat 50% of the points), you can take part in the regular course exam, or any of the separate exams until Spring 2018 – they are treated as renewal exams for the course.

If you have not completed the weekly exercises but want to get the grade, you need to perform a separate set of small project exercises. The project details are available in Moodle.

2 What to bring to the exam?

As with all the exams at the department, you should bring writing materials (pencil etc. but not your own paper) and some means of identification (student card, passport etc.).

Additionally, for the exams of this course (including both the course exam and the separate exams), you may bring a "cheat sheet" which is one hand-written A4 sheet, to which you can write whatever information you think might be useful in the exam, using both sides if you wish. Even if you don't think you'll really need a cheat sheet in the exam, you may wish to create one just to help clarify to yourself what you think the important things are. You are not allowed to bring any other written material.

You will not need, and should not bring, a calculator. Use of any electronic devices, including mobile phones, is prohibited.

3 What will be asked in the exam?

In the exam, you may be asked to

- Briefly define and explain key concepts and terms
- Explain algorithms, techniques and other broader topics in more detail; you should explain all possible aspects of the topic, but exact mathematical details are typically not critical unless explicitly asked for (however, presenting them often clarifies the answer)
- Manual execution of an algorithm on a small data set
- Derivation of the mathematical details for some algorithm or model
- Problem solving: given a task you need to come up with a practical solution using the techniques presented on the course

4 Key topics and concepts

The most important things you should understand:

- Risk minimization as the goal of typical machine learning solutions, estimating the risk
- Probabilistic generative models as a way of deriving practical loss functions
- Preventing overfitting: regularization, priors, sparsity and early stopping
- Basics of probability, differential calculus and linear algebra, so that you can perform simple derivations
- Gradient-based optimization, with particular focus on stochastic gradients
- Typical machine learning tasks

You should also familiarize yourself with the practical models and algorithms covered on the course, focusing on the ones we covered also in the exercises at some level:

- Clustering: Mixture models (including the EM algorithm) and spectral clustering
- Linear latent variable models and matrix factorization: PCA, ICA and NMF
- Non-linear dimensionality reduction: SNE
- Linear supervised models: least squares, logistic regression, Lasso (including coordinate descent for l_1 regularization)
- Kernel methods: Non-linearity by preprocessing, support vector machines, max-margin principle
- Adaptive basis functions: Decision trees and random forests
- Boosting, bagging and ensembles
- Neural network concepts, backpropagation for MLPs
- Different deep learning architectures (CNN, RNN etc) and their uses

5 Exam material

To prepare for the exam you should

1. Read through the lecture slides
2. Check the exercises and their model solutions
3. Read selected parts of the course book "Machine Learning: A Probabilistic Perspective"

We did not cover the whole book, and hence detailed list of sections covering the course exam is provided below. Phrases like "3.1 until 3.1.3" mean that you should start from the beginning on Section 3.1 and read still Section 3.1.3. Note that the page numbers below refer to the 5th printing of the book; the author says the page numbers can vary a bit from print to print. However, the Section numbers should be stable – use those as the definitive guide and think of the page ranges as an additional hint.

Some topics are not covered by the book. This includes multidimensional scaling, stochastic neighbor embedding, Isomap and many of the latest deep learning concepts. In addition, NMF is covered very briefly. For these topics the level of detail presented on the lecture slides and exercises is sufficient.

Sections	Pages	Notes
Chapter 1: Basics		
All	1-25	
Chapter 2: Probability		
2.1-2.6, 2.8	27-53, 56-61	No need to remember the density function except for Bernoulli and multivariate normal.
Chapter 3: Generative models for discrete data		
3.1, 3.2, 3.5.3	67-74, 88	We did not cover 3.2, but it might help understanding what the prior, likelihood and posterior mean.
Chapter 4: Gaussian models		
4.1, 4.2 except 4.2.8, 4.3 until 4.3.1, 4.3.4.2, 4.4.1	99-110, 112-113, 121-122	The rest of 4.4 might help understanding the latent linear models (Section 12)
Chapter 5: Bayesian statistics		
5.1-5.2, 5.3 until 5.3.1, 5.7 until 5.7.1.5	151-159	
Chapter 6: Frequentist statistics		
6.1-6.2, 6.3 until 6.3.2, 6.4-6.5, 6.6.4	193-199, 202-214, 217-218	
Chapter 7: Linear regression		
7.1-7.5 except 7.5.3	219-230, 232	
Chapter 7: Logistic regression		
8.1-8.3, 8.5 until 8.5.3, 8.6 until 8.6.1	247-257, 264-268, 270-271	One table is on page 273
Chapter 9: Generalized linear models		
9.5	298-300	
Chapter 10: Directed graphical models		
10.3-10.4	321-325	
Chapter 11: Mixture models and the EM algorithm		
11.1-11.4 until 11.4.2.8, 11.4.7 (and subsections), 11.5-11.6	339-359, 365-367, 372-376	The theoretical foundation of EM is not critical, but might help understanding the algorithm. Being able to apply EM for simple models is enough.
Chapter 12: Latent linear models		
12.1-12.3 until 12.3.2, 12.6	383-403, 409-418	Cursory reading of the SVD part is enough
Chapter 13: Sparse linear models		
13.1-13.3 until 13.3.4, 13.4 until 13.4.2, 13.8 until 13.8.3	423-440, 443-444, 470-474	Cursory reading of 13.2 is enough, treating it as motivation for l_0 regularization
Chapter 14: Kernels		
14.1-14.2 until 14.2.7, 14.3-14.4 until 14.4.3.3, 14.5	481-486, 488-495, 498-507	Treat the kernels not covered in lectures (and K-medoids) as further motivation
Chapter 16: Adaptive basis function models		
16.1-16.2, 16.4 until 16.4.4, 16.4.8, 16.5 until 16.5.6.4, 16.6 except 16.6.2, 16.7	545-554, 556-562, 564-578, 582-587	
Chapter 25: Clustering		
25.1, 25.4, 25.5 until 25.5.1.3	877-881, 892-900	No need to remember the details for the evaluation methods. Note that Chapter 11 also covers clustering models
Chapter 27: Latent variable models for discrete data		
27.6.2	983-987	
Chapter 28: Deep learning		
28.1, 28.3-28.5	999, 1003-1011	