Advanced course in machine learning
582744
Lecture 13

Arto Klami

Exam and grading

- Exam on May 11th 9am
- You need 50% of the points for both the exercises and exam to pass (well, 44% for the exercises if you count the bonus exercise set)
- 40% weight for the exercises, 60% for the exam
- The exercises are valid at least until the next year's course – you can take part in any separate exam and count them towards the grade

Not enough exercise points? You can take part in the separate exam and solve another set of small projects, but talk to me first

Exam question types

- Explain concepts briefly
- Explain some model in more detail: Explain all aspects, such as the loss function, optimization algorithm, potential regularization, typical uses, interesting properties, etc. Exact mathematical details are not so important unless explicitly asked for.
- Derivation of some practical algorithm
- Manual execution of an algorithm
- Problem solving
- No coding related problems – we already did that in the exercises

Would you want to have a cheat sheet?

Example solutions

Question: Explain briefly the concept of *Ensemble learning*:

Answer: A strategy for combining multiple supervised models by taking the weighted average of their outputs. The ensemble is $f(y|\mathbf{x}) = \sum_{m \in \mathcal{M}} w_m f_m(y|\mathbf{x})$, where w_m are the weights and $f_m(y\|x)$ are the models.

Example solutions

Question: Explain the NMF model with sufficient detail:

Answer: Non-negative matrix factorization (NMF) is a method that approximates a matrix $X \in \mathbb{R}_+^{N \times D}$ consisting of non-negative entries using a low-rank product WH where both $W \in \mathbb{R}_+^{N \times K}$ and $H \in \mathbb{R}_+^{K \times D}$ are also non-negative. The parameters of this model are hence W and H and they are learnt by minimizing the squared loss $\|X - WH\|^2$ under the non-negativity constraint. One pratical way is to initialize W and H for some non-negative initial values and use product-based updates that converge to a local optimum of the loss function.
The complexity of the model is determined by the rank K of the approximation, such that if K is identical to $min(N, D)$ we can reconstruct the input matrix exactly. No regularization methods for NMF were explained during the course. The model can be used to, for example, to learn more interpretable basis for natural images."

Exam material

- The lecture slides
- The course book
- The exercises and model solutions (but note that model solutions for the 7th exercise or the bonus exercises will not be available before the exam)

Reading the book

- Detailed reading instructions (page ranges) provided on the web page
- In total around 380 pages
- The level of technical detail a bit deeper than on the lecture slides; whever neither the book nor the slides give the mathematical details it is enough to understand the basic concepts

Topics not in the book

- NMF and recurrent neural networks only very briefly – complement with the lecture slides
- Multidimensional scaling, Isomap, stochastic neighbor embedding – for these the lecture slides and exercises are sufficient material
- Self-organizing map – you can skip this
- The recent advances in neural networks (dropout, autodiff etc) – the lecture slides are enough

Alternative reading

Combining the lecture slides with the following should also be enough

- The UML lecture notes: http://www.cs.helsinki.fi/u/ahyvarin/teaching/uml2015/uml2015_lecturenotes.pdf covering the background and linear unsupervised models
- David Barber's book http://web4.cs.ucl.ac.uk/staff/D.Barber/textbook/181115.pdf covers generative models well
- The deep learning book http://www.deeplearningbook.org/ covering both basics of ML and all the details on neural networks
- Elements of statistical learning https://web.stanford.edu/~hastie/local.ftp/Springer/OLD/ESLII_print4.pdf covers most of the topics on the course but has quite different perspective

Important topics

- Basics of probabilities; independence, mean, variance, random variable, Bayes rule, etc
- Gradient-based optimization, including stochastic gradients, second order methods and coordinate descent
- Risk minimization (frequentist and Bayesian formulations), empirical risk, regularization, validation, overfitting
- The concept of generative models, plate diagrams, likelihood, prior and posterior
- Learning tasks: Supervised, unsupervised, multi-task etc

Important topics

- Clustering: Mixture models, spectral clustering, hierarchical clustering
- Linear latent variable models: PCA, factor analysis, ICA
- Matrix factorization and recommender systems, NMF
- Non-linear dimensionality reduction: MDS, SNE, spectral embeddings
- Parameter inference for latent variable models: The EM algorithm
- Probabilistic and classical formulations; know both if applicable

Important topics

- Linear methods for regression and classification: least squares, logistic regression
- Regularization and sparsity for linear models; Lasso
- Generative and discriminative classifiers; Gaussian generative classifiers

Important topics

- Nonlinear methods by nonlinear preprocessing for linear methods
- Kernels and the kernel trick
- Maximum margin methods; SVM and the SMO algorithm
- Adaptive basis functions; CART, random forests
- Boosting (AdaBoost), bagging, ensembles

Important topics

- Terminology and representation: nodes, weights, activation functions, layers etc
- Multi-layer perceptron, the backpropagation algorithm
- Restricted Boltzmann machine
- Convolutional neural network and its applications
- Autoencoder, de-noising autoencoder

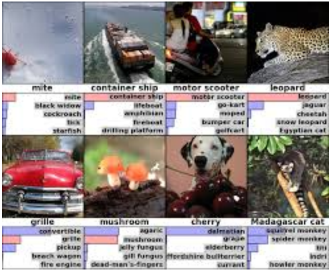
Important topics

- Practical learning of MLPs: regularization, initialization, batch normalization etc
- The concept of recurrent neural networks
- The deep generative networks; basic ideas are enough
- Representation learning; hidden layers of deep models are distributed representations

The really important topics

- Generalization: What does it mean, how to measure it, and how to avoid poor generalization
- Probabilistic formulation of machine learning: Log-probability of generative models as the modeling goal, relationship to classical losses and regularization
- Practical non-linear models: kernels, adaptive basis function models and neural networks
- Important algorithms: EM, backpropagation, gradient-based methods in general, coordinate descent for l_1 regularization
- All practical methods are equally important; you should know how to explain any of them

Which of these could you replicate after the course?



Leaderboard

Showing Test Score. Click here to show quiz score

Display top 20 leaders.

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries !	0.8591	9.81	2009-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace	0.8612	9.59	2009-07-24 17:18:43