Skoltech

Course Syllabus

Course Title Bayesian Methods of Machine Learning

Course Title (in Russian) Байесовские методы машинного обучения

Lead Instructor Burnaev, Evgeny

Co-Instructor Burnaev, Evgeny; Zaytsev, Alexey

Co-Instructor

First Name	Last Name
Alexander	Kolesov
Sergey	Kholkin
Grigory	Shutov

1. Annotation

Course Description

The course addresses Bayesian approach for solving various machine learning and data analysis problems. It offers the framework for solving inverse problems, model selection and continual learning. The most attention is payed to the fusion between the deep learning techniques and the elements of bayesian approach. This fusion is possible due to the development of a range of approximate inference techniques in the last 20 years.

Modern machine learning papers frequently use machinery of the approximate inference techniques. The main learning outcome of the course is the ability to read and reproduce related papers, and to apply corresponding methods of approximate inference for the development of Bayesian machine learning approaches. In order to reach this goal, the course contains theoretical and practical assignments and the final project.

The purpose of the theoretical tasks is two-fold. Firstly, we would like to develop skills of equation derivation, that is used routinely in papers and is usually suppressed. Secondly, some theoretical tasks provide the intuition about properties of the methods through toy models. The purpose of the practical tasks is straightforward: to translate discussed methods to the code and observe properties of the methods through examples.

These assignments leads to the course projects: reproducing and discussing a relevant paper from a recent conference, and/or development of some students' research ideas using methods of Bayesian machine learning.

Course Description (in Russian)

Курс посвящен Байесовскому подходу к решению различных задач машинного обучения и анализа данных. Байесовскому подход лежит в основе решения важных прикладных задач, таких как обратные задачи, задачи выбора модели, и задачи непрерывного (онлайн) обучения.

Наибольшее внимание в курсе уделяется слиянию методов глубокого обучения с элементами байесовского подхода. Это слияние стало возможным благодаря развитию за последние 20 лет ряда методов приближенного вывода.

Современные статьи по машинному обучению часто используют методы приближенного вывода. Основным результатом обучения на курсе является умение читать и воспроизводить соответствующие статьи, а также применять соответствующие методы приближенного вывода для разработки алгоритмов байесовского машинного обучения. Для достижения этой цели в курсе предусмотрены теоретические и практические задания, а также итоговый проект.

Цель теоретических заданий двоякая. Во-первых, их выполнение позволит студентам развить навыки получения явных формул, используемых в работах в основе методов приближенного вывода. При этом, в современной литературе фактически отсутствуют источники, по которым можно было бы изучить основные подходы к получению такого рода явных результатов. Выполнение соответствующих заданий позволит восполнить этот пробел. Во-вторых, некоторые теоретические задачи, решенные в рамках упрощенных предположений, позволяют лучше интуитивно понять теоретические свойства соответствующих байесовских методов. Цель практических заданий проста: реализовать обсуждаемые методы в программном виде и изучить их свойства при применении к реальным данным.

Указанные задания подводят учащихся к курсовым проектам: воспроизведение и обсуждение соответствующих статей с передовых конференций по машинному обучению, и/или реализация собственных научно-исследовательских проектов с применением байесовских методов машинного обучения.

2. Basic Information

Course Academic Level

MSc PhD

Number of ECTS credits

6

Course Prerequisites / Recommendations

This is an Elective course in Term 1B/2. It is recommended to take Numerical Linear Algebra and Optimization Methods, as well as course on Machine Learning and Applications before this one. As well we suppose an attendee be fluent with linear algebra, probability and real analysis.

Type of Assessment

Graded

Mapping from grades to percentage:

A:	86
B:	70
C:	56
D:	46
E:	36
F:	0

Term

Term 1B-2

Students of Which Programs do You Recommend to Consider this Course as an Elective?

BSc Programs	Masters Programs	PhD Programs
	Advanced Computational Science Applied Computational Mechanics Data Science Internet of Things and Wireless Technologies Life Sciences	Computational and Data Science and Engineering Life Sciences

Maximum Number of Students

	Maximum Number of Students
Overall:	40
Per Group (for seminars and labs):	40

Course Stream

Science, Technology and Engineering (STE)

3. Course Content

Topic	Summary of Topic	Contact Hours: Lectures	Contact Hours: Seminars	Contact Hours: Labs	Non-contact Hours: Student's Independent Study
Foundations. Exact Inference	- Some canonical problems: classification, regression, clustering and density estimation - Representing beliefs and the Cox axioms. The Dutch Book Theorem. Bayesian reasoning and interpretation of probabilities - Bayesian Occam's Razor and Model Selection - Asymptotic Certainty. Asymptotic Concensus. Asymptotic Normality of the Posterior and Bernstein-von Mises theorem - Exponential Families: properties, sufficient statistics, conjugacy - Generalized Linear Models (GLM), Automatic Relevance Determination. Home assignment	6	6	1	10
Non-exact Inference: Variational Approaches	- Expectation-Maximization. Principal Component Analysis - Mean Field Approximation. Complete conditionals in the Exponential Family. Blind Image Denoising - Stochastic Variational Inference. Latent Direchlet Allocation. - Semi-Implicit Variational Inference Home assignment	6	6	2	20
Non-exact Inference: Deep Bayes	- Variational Auto-Encoding models - Stochastic gradients estimators and variance reduction - Variational inference with normalizing flows. - Variational Dropout. Home assignments	10	10	2	20
Non-Exact Inference: MCMC Approaches	- Metropolis-Hastings. Lengevin Dynamics. Scalable Langevin Dynamics. Hamiltonian Dynamics. Home assignment	4	4	2	10
Non-Exact Inference: Diffusion generative models	- Score matching techniques: denoising and sliced. - Noise conditional score networks - Gaussian diffusion models - Denoising diffusion probabilistic models. Home assignment	10	10	2	21

4. Learning Outcomes

Skoltech Learning Outcomes are indicated as per Skoltech Learning Outcomes Framework.

1. FUNDAMENTAL KNOWLEDGE

1.1. KNOWLEDGE OF MATHEMATICS AND NATURAL SCIENCES

2.1. COGNITION AND MODES OF REASONING

2.1.1. Analytical reasoning and problem solving

2.2. ATTITUDES AND LEARNING PROCESS

2.2.2. Willingness to make decisions in the face of uncertainty

2.2.3. Responsibility, intensity, perseverance, urgency and will to deliver

2.3. ETHICS, EQUITY AND OTHER RESPONSIBILITIES

2.3.1. Ethical action, integrity and courage

2.3.2. Social responsibility

3.1. COMMUNICATIONS IN INTERNATIONAL ENVIRONMENTS

3.1.1. Communications strategy and structure

3.2. TEAMWORK AND LEADERSHIP

3.2.1. Forming effective teams

3.3. COLLABORATION AND CHANGE

3.3.1. Establishing diverse connections and networking

4.1. MAKING SENSE OF GLOBAL SOCIETAL ENVIRONMENTAL AND BUSINESS CONTEXT

4.1.2. Taking responsibility for sustainable development, including social, economic, environmental and work environment aspects

4.2. VISIONING — INVENTING NEW TECHNOLOGIES THROUGH RESEARCH

4.2.1. The research process — hypothesis, evidence and defense

4.3. VISIONING — CONCEIVING AND DESIGNING SUSTAINABLE SYSTEMS

4.3.1. Identifying stakeholders need and wants

4.4. DELIVERING ON THE VISION — IMPLEMENTING AND OPERATING

4.4.1. Designing and optimizing sustainable and safe implementation and operations

4.5. DELIVERING ON THE VISION — ENTREPRENEURSHIP AND ENTERPRISE

4.5.1. New venture conceptualization and creation

5. Assignments and Grading

Physical Attendance Requirement (% of classes) 80

In-person Attendance Requirement Comment

students are allowed to miss the classes only due to viable reason with a justification such as a medical report

Assignment Type	Assignment Summary	% of Final Course Grade
Team Project	The participants of the course should choose one project and complete proposed task.	30
Homework Assignments	The course is composed of 4 homework assignments. Each homework consists of theoretical and coding problems.	40
Midterm Exam	The Midterm exam is one test in the middle of the course. The test is in the form of multiple choice questions and short problems.	30

6. Assessment Criteria

Assignment 1 Type Team Project

Sample of Assignment 1

Examples of the topics:
1. Deep kernels and Gaussian processes
2. Bayesian Active Learning
3. Acceleration of Diffusion models
4. Deep Variational autoencoders
5. Bayesian change-point detection
6. Comparison of approaches for approximation of intractable Bayesian models
7. MCMC for Bayesian inference
8. Diffusion generative models

Final course project (groups up to 3):
- Possible to combine with projects from parallel courses
- Default project topics will be announced on week 3
- Stages: Project proposal (week 4),
- Milestone status checkup (week 6),
- Presentation and Final Report submission (week 8)

Final Project types
- Applied: pick an interesting application and figure out how to apply machine learning algorithms to solve it;
- Algorithmic: propose a new learning algorithm, or a variant of some existing one to solve a general problem or group thereof.

The Final Report is a PDF:
- Introduction: motivation and problem statement
- Related work and brief literature overview

- Dataset Description
- ML Methods and algorithms, proposed algorithm modifications, etc.
- Experiments/Discussion: details about (hyper)parameters and how you picked them, cross-validation metrics and details, discussion of failures and successes, equations, results, visualizations, tables, etc.
- Conclusion, references, acknowledgements and contributions
- Up to 5 pages including figures, tables, appendices (in algorithmicprojects only), and excluding references/contributions
- Source code (scripts, notebooks) in ZIP or on Github

Or Upload Sample of Assignment 1

Sample of Assignment 1

Assessment Criteria for Assignment 1

The main assessment criteria:

- General evaluation criteria for the Report

- significance, novelty: toy/real problem or common/unexplored method
- technical quality: insightful choice of clever reasonable methods, cross-validation and general quality assessment of used tools/methods
- general report quality and structure
- relevance to the topics covered during the course

- The Project presentation

- presentation quality and clarity
- relevant technical content and summary
the knowledge demonstrated by the team

Assignment 2 Type

Midterm Exam

Sample of Assignment 2

There is one test in the middle of the course (midterm exam)

The test is in the form of multiple choice questions and short problems.

The main assessment criteria:
- correct answers to multiple choice questions
- complete and correct solutions to problems

Assignment 3 Type

Homework Assignments

Sample of Assignment 3

Example of problems for a homework:

1) Use the Bernoulli mixture to model handwritten digits from the MNIST handwritten digit database. Here turn the digit into a binary vector by setting all elements whose values exceed 0.5 to 1 and setting the remaining elements to 0. Apply the mixture model for clustering of handwritten digits. Investigate clustering performance on parameters of the algorithm.

2) Extend the variational treatment of Bayesian linear regression to include a gamma hyperprior Gam(\beta|c_0,d_0) (on the noise precision parameter \beta) besides a gamma hyperprior on \alpha (the precision of a Gaussian prior distribution on the linear regression model parameters), by assuming a factorized variational distribution of the form q(w)q(\alpha)q(\beta). Derive the variational equations for the three factors in the variational distribution and also obtain an expression for the lower bound and for

the predictive distribution.

Assessment Criteria for Assignment 3

The solution to a homework is:
- a PDF with a report on the problem solution with complete, correct and clear explanations
- source code (scripts, notebooks) in ZIP or on Github

The main assessment criteria:
- complete and correct solutions to problems

7. Textbooks and Internet Resources

You can request at most two required textbooks. Additionaly, you can suggest up to nine recommended textbooks.

Required Textbooks	ISBN-13 (or ISBN-10)
Bishop, C.M. Pattern Recognition and Machine Learning. Springer, 2007	978-0387310732
Barber, D. Bayesian Reasoning and Machine Learning. Cambridge University Press, 2012	978-0521518147

Recommended Textbooks	ISBN-13 (or ISBN-10)
Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2009.	9780387848570
Peter Muller, Fernando Andres Quintana, Alejandro Jara, Tim Hanson. Bayesian Nonparametric Data analysis. Springer, 2015	9783319189673
A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin. Bayesian Data Analysis. Chapman & Hall/CRC, 2004	9781439840955
J. Wakefield. Bayesian and Frequentist Regression Methods. Springer, 2013	9781441909244
J.K. Ghosh, M. Delampady, T. Samanta. An Introduction to Bayesian Analysis. Theory and Methods. Springer, 2006	9780387400846
J.K. Ghosh, R.V. Ramamoorthi. Bayesian Nonparametrics. Springer, 2003	9780387955377
M.-H. Chen, Q.-M. Shao, J.G. Ibrahim. Monte Carlo Methods in Bayesian Computation. Springer, 2000	9781461270744
E.G. Phadia. Prior Processes and Their Applications. Springer, 2013	9783642392795

Recommended Textbooks	ISBN-13 (or ISBN-10)
C.P. Robert. The Bayesian Choice. From Decision-Theoretic Foundations to Computational Implementation. Springer, 2007	9780387715988

Web-resources (links)	Description
http://gaussianprocess.org/gpml/	Carl Rasmussen and Christopher Williams. Gaussian Processes for Machine Learning. The MIT Press, 2006.
http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage	online version of Barber's "Bayesian Reasoning and Machine Learning"
http://wol.ra.phy.cam.ac.uk/mackay/itila/book.html	D. J. C. MacKay (2003) Information Theory, Inference, and Learning Algorithms.

8. Facilities

Software
Software requirements: Python 3.6 and PyTorch: latest Numpy: latest

Equipment
Access to the Internet through a computer class and Wi-Fi network of the institute. Whiteboard/Flipchart; Projector

Labs for Education

- Artificial Intelligence & Supercomputing Laboratory
- Computational Imaging Laboratory
- Laboratory of Artificial Intelligence for Materials Design
- Laboratory of Computational Intelligence
- Laboratory of Intelligent Signal and Image Processing
- Laboratory of Mathematical Foundations of Artificial Intelligence
- Laboratory of Tensor Networks and Deep Learning for Data Mining
- Laboratory “Multiscale Neurodynamics for Intelligent Systems”
- Mobile Robotics Laboratory

9. Additional Notes

Comments

Materials/textbook, specifically selected for different sections of the course

1. Exact Inference & GLM

• S. AMARI and A. CICHOCKI. Information geometry of divergence functions. url: http://fluid.ippt.gov.pl/bulletin/%2858-1%29183.pdf
• Rina Foygel Barber, Mathias Darton, and Kean Ming Tan. “Laplace Ap- proximation in High-dimensional Bayesian Regression”. In: arXiv:1503.08337 [math, stat] (Mar. 2015). arXiv: 1503.08337. url: http://arxiv.org/ abs/1503.08337
• Bradley Efron. “The Geometry of Exponential Families”. EN. In: The Annals of Statistics 6.2 (Mar. 1978), pp. 362–376. issn: 0090-5364, 2168- 8966. doi: 10.1214/aos/1176344130. url: https://projecteuclid. org/euclid.aos/1176344130
• Eric P. Xing. The Exponential Family and Generalized Linear Models. url: https://www.cs.cmu.edu/~epxing/Class/10708-14/scribe_ notes/scribe_note_lecture6.pdf
• Andrew Gelman and Jennifer Hill. Data analysis using regression and multilevel/hierarchical models. English. OCLC: 646068240. Cambridge; New York: Cambridge University Press, 2007. isbn: 978-0-511-76955-9 978-0- 511-26811-3 978-0-511-79094-2 978-1-282-65283-5. url: http://dx.doi. org/10.1017/CBO9780511790942
• S. Kullback and R. A. Leibler. “On Information and Suciency”. EN. In: The Annals of Mathematical Statistics 22.1 (Mar. 1951), pp. 79–86. issn: 0003-4851, 2168-8990. doi: 10.1214/aoms/1177729694. url: https://projecteuclid.org/euclid.aoms/1177729694
• M. Tipping M. Bishop. Bayesian Regression and Classification. url: https: //www.microsoft.com/en-us/research/wp-content/uploads/2016/02/bishop-nato-bayes.pdf
• Frank Nielsen and Vincent Garcia. “Statistical exponential families: A digest with flash cards”. In: arXiv:0911.4863 [cs] (Nov. 2009). arXiv: 0911.4863. url: http://arxiv.org/abs/0911.4863
• Erlis Ruli, Nicola Sartori, and Laura Ventura. “Improved Laplace Ap- proximation for Marginal Likelihoods”. In: Electronic Journal of Statistics 10.2 (2016). arXiv: 1502.06440, pp. 3986–4009. issn: 1935-7524. doi: 10.1214/16-EJS1218. url: http://arxiv.org/abs/1502.06440
• STA 250: Statistics. Laplace Approximation to the Posterior. url: http: //www2.stat.duke.edu/~st118/sta250/laplace.pdf
• Michael E. Tipping. Bayesian Inference: An introduction to Principles and Practice in Machine Learning. url: https://www.ics.uci.edu/~smyth/courses/cs274/readings/bayesian_regression_overview.pdf

2. Non Exact Inference: Variational Methods

• Shun-ichi Amari. Natural Gradient Works Efficiently in Learning.
• David M. Blei, Alp Kucukelbir, and Jon D. McAulie. “Variational Inference: A Review for Statisticians”. In: Journal of the American Statistical Association 112.518 (Apr. 2017). arXiv: 1601.00670, pp. 859–877. issn: 0162-1459, 1537-274X. doi: 10.1080/01621459.2017.1285773. url: http://arxiv.org/abs/1601.00670
• EM and MM Algorithms. url: https://www.samsi.info/wp-content/ uploads/2016/08/Zhou_samsi-opt-summerschool-20160810-1.pdf
• Matt Hoffman et al. “Stochastic Variational Inference”. In: arXiv:1206.7051 [cs, stat] (June 2012). arXiv: 1206.7051. url: http://arxiv.org/abs/ 1206.7051
• M. Tipping M. Bishop. Mixtures of Probabilistic PCA. url: http://www. miketipping.com/papers/met-mppca.pdf
• Razvan Pascanu and Yoshua Bengio. “Revisiting Natural Gradient for Deep Networks”. In: arXiv:1301.3584 [cs] (Jan. 2013). arXiv: 1301.3584. url: http://arxiv.org/abs/1301.3584
• Michael E. Tipping. Probabilistic PCA. url: https://www.microsoft. com/en-us/research/wp-content/uploads/2016/02/bishop-ppca-jrss.pdf
• Tong Tong Wu and Kenneth Lange. “The MM Alternative to EM”. In: Statistical Science 25.4 (Nov.

2010). arXiv: 1104.2203, pp. 492–505. issn: 0883-4237. doi: 10.1214/08-STS264. url: http://arxiv.org/abs/1104. 2203
• Anderson Y. Zhang. Theoretical and computational guarantees of mean field variational inference for community detection. url: http://www.stat.yale.edu/~hz68/variational.pdf

3. Non Exact Inference: Deep Learning and Bayesian Methods

• Martin Arjovsky, Soumith Chintala, and Leon Bottou. “Wasserstein GAN”. In: arXiv:1701.07875 [cs, stat] (Jan. 2017). arXiv: 1701.07875. url: http: //arxiv.org/abs/1701.07875
• Yuri Burda, Roger Grosse, and Ruslan Salakhutdinov. “Importance Weighted Autoencoders”. In: arXiv:1509.00519 [cs, stat] (Sept. 2015). arXiv: 1509.00519. url: http://arxiv.org/abs/1509.00519
• Chris Cremer, Quaid Morris, and David Duvenaud. “Reinterpreting Importance- Weighted Autoencoders”. In: arXiv:1704.02916 [stat] (Apr. 2017). arXiv: 1704.02916. url: http://arxiv.org/abs/1704.02916
• Alex Graves. “Stochastic Backpropagation through Mixture Density Distributions”. In: arXiv:1607.05690 [cs] (July 2016). arXiv: 1607.05690. url: http://arxiv.org/abs/1607.05690
• Shixiang Gu et al. “MuProp: Unbiased Backpropagation for Stochastic Neural Networks”. en. In: (Nov. 2015). url: https://arxiv.org/abs/ 1511.05176
• Jiri Hron, Alexander G. de G. Matthews, and Zoubin Ghahramani. “Variational Gaussian Dropout is not Bayesian”. In: arXiv:1711.02989 [stat] (Nov. 2017). arXiv: 1711.02989. url: http://arxiv.org/abs/1711. 02989
• Eric Jang, Shixiang Gu, and Ben Poole. “Categorical Reparameterization with Gumbel-Softmax”. In: arXiv:1611.01144 [cs, stat] (Nov. 2016). arXiv: 1611.01144. url: http://arxiv.org/abs/1611.01144
• Diederik P. Kingma, Tim Salimans, and Max Welling. “Variational Dropout and the Local Reparameterization Trick”. In: arXiv:1506.02557 [cs, stat] (June 2015). arXiv: 1506.02557. url: http://arxiv.org/abs/1506. 02557
• Diederik P. Kingma and Max Welling. “Auto-Encoding Variational Bayes”. In: arXiv:1312.6114 [cs, stat] (Dec. 2013). arXiv: 1312.6114. url: http://arxiv.org/abs/1312.6114
• Lars Maaløe et al. “Auxiliary Deep Generative Models”. In: arXiv:1602.05473 [cs, stat] (Feb. 2016). arXiv: 1602.05473. url: http://arxiv.org/abs/ 1602.05473
• Andriy Mnih and Karol Gregor. “Neural Variational Inference and Learn- ing in Belief Networks”. en. In: (Jan. 2014). url: https://arxiv.org/ abs/1402.0030
• Dmitry Molchanov, Arsenii Ashukha, and Dmitry Vetrov. “Variational Dropout Sparsifies Deep Neural Networks”. In: arXiv:1701.05369 [cs, stat] (Jan. 2017). arXiv: 1701.05369. url: http://arxiv.org/abs/1701. 05369
• Sebastian Nowozin, Botond Cseke, and Ryota Tomioka. “f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization”. In: arXiv:1606.00709 [cs, stat] (June 2016). arXiv: 1606.00709. url: http: //arxiv.org/abs/1606.00709
• Mihaela Rosca et al. “Variational Approaches for Auto-Encoding Gener- ative Adversarial Networks”. In: arXiv:1706.04987 [cs, stat] (June 2017). arXiv: 1706.04987. url: http://arxiv.org/abs/1706.04987
• Francisco J. R. Ruiz, Michalis K. Titsias, and David M. Blei. “The Generalized Reparameterization Gradient”. In: arXiv:1610.02287 [stat] (Oct. 2016). arXiv: 1610.02287. url: http://arxiv.org/abs/1610.02287
• John Schulman et al. “Gradient Estimation Using Stochastic Computation Graphs”. In: arXiv:1506.05254 [cs] (June 2015). arXiv: 1506.05254. url: http://arxiv.org/abs/1506.05254
• Jakub M. Tomczak and Max Welling. “VAE with a VampPrior”. In: arXiv:1705.07120 [cs, stat] (May 2017). arXiv: 1705.07120. url: http: //arxiv.org/abs/1705.07120
• George Tucker et al. “REBAR: Low-variance, unbiased gradient estimates for discrete latent variable models”. In: (Mar. 2017).

4. Non-exact inference: MCMC methods.

• Handbook of Markov Chain Monte Carlo. url: http://www.mcmchandbook. net/
• Yi-An Ma, Tianqi Chen, and Emily B. Fox. “A Complete Recipe for Stochas- tic Gradient MCMC”. In: arXiv:1506.04696 [math, stat] (June 2015). arXiv: 1506.04696. url: http://arxiv.org/abs/1506.04696
• Monte Carlo theory, methods and examples. url: http://statweb.stanford. edu/~owen/mc/
• Andriy Mnih Ruslan Salakhutdinov. Bayesian Probabilistic Matrix Factor- ization using Markov Chain Monte Carlo. url: https://www.cs.toronto. edu/~amnih/papers/bpmf.pdf
• Yee Whye Teh Welling. Bayesian Learning via Stochastic Gradient Langevin Dynamics. url: https://www.ics.uci.edu/~welling/publications/ papers/stoclangevin_v6.pdf

5. Gaussian Process and Variational Optimization

• Ryan P. Adams A Tutorial on Bayesian Optimization for Machine Learning. url: https://www.iro.umontreal.ca/~bengioy/cifar/NCAP2014-summerschool/slides/Ryan_adams_140814_bayesopt_ncap.pdf
• Mauricio A. Alvarez, Lorenzo Rosasco, and Neil D. Lawrence. “Kernels for Vector-Valued Functions: a Review”. In: arXiv:1106.6251 [cs, math, stat] (June 2011). arXiv: 1106.6251. url: http://arxiv.org/abs/1106.6251
• Eduard Gabriel B az avan. Fourier Kernel Learning. url: http://www. maths.lth.se/matematiklth/personal/sminchis/papers/eccv12bls. Pdf
• [David Barber. Optimization by Variational Bounding. url: https:// www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2013-65.pdf
• Gaussian Processes for Machine Learning: Book webpage. url: http:// www.gaussianprocess.org/gpml/
• Gaussian Process and Uncertainty Quantification Summer School 2017. original-date: 2016-12-22T22:57:23Z. July 2018. url: https:// github.com/gpschool/gpss17
• James Hensman, Nicolas Durrande, and Arno Solin. “Variational Fourier features for Gaussian processes”. In: arXiv:1611.06740 [stat] (Nov. 2016). arXiv: 1611.06740. url: http://arxiv.org/abs/1611.06740
• Mehryar Mohri. Optimistic Bandit Convex Optimization. url: https: //cs.nyu.edu/~mohri/pub/bcop.pdf
• Tim Salimans et al. “Evolution Strategies as a Scalable Alternative to Reinforcement Learning”. In: arXiv:1703.03864 [cs, stat] (Mar. 2017). arXiv: 1703.03864. url: http://arxiv.org/abs/1703.03864
• Bobak Shahriari. Taking the Human Out of the Loop: A Review of Bayesian Optimization. url: https://www.cs.ox.ac.uk/people/nando.defreitas/publications/BayesOptLoop.pdf
• Joe Staines and David Barber. “Variational Optimization”. In: arXiv:1212.4507 [cs, stat] (Dec. 2012). arXiv: 1212.4507. url: http://arxiv.org/abs/ 1212.4507
• Titsias. Variational Learning of Inducing Variables in Sparse Gaussian Processes. url: http://proceedings.mlr.press/v5/titsias09a/ titsias09a.pdf
• Tianbao Yang. Nystro m Method vs Random Fourier Features: A Theoretical and Empirical Comparison. url: https://papers.nips.cc/paper/4588-nystrom-method-vs-random-fourier-features-a-theoretical-and-empirical-comparison.pdf

6. Sequential Data
• Blei. Exact Inference: Elimination and Sum Product (and hidden Markov models). url: http://www.cs.columbia.edu/~blei/fogm/2015F/notes/ inference.pdf
• Henrik I. Christensen. Graphical Models & HMMs. url: https://www.cc. gatech.edu/~hic/CS7616/pdf/lecture6.pdf
• Eric P. Xing. Kalman Filtering. url: http://www.cs.cmu.edu/~epxing/ Class/10708-16/note/10708_scribe_lecture12.pdf
• Jonathan S. Yedidia. Understanding Belief Propagation and its Generalizations. url: http://www.merl.com/publications/docs/TR2001-22.pdf
• Thomas Minka. “Divergence Measures and Message Passing.” Accessed August 9, 2018. https://pdfs.semanticscholar.org/40b0/02847cecce2e3cfaeae83e85393e57179a0c.pdf

7. Non-Exact inference: Diffusion generative models
• Yang Song. Generative Modeling by Estimating Gradients of the Data Distribution. url:https://proceedings.neurips.cc/paper/2019/file/3001ef257407d5a371a96dcd947c7d93-Paper.pdf
• Jonathan Ho. Denoising Diffusion Probabilistic Models. url:https://deepsense.ai/wp-content/uploads/2023/03/2006.11239.pdf
• Yang Song. Score-Based Generative Modeling through Stochastic Differential Equations url:https://openreview.net/forum?id=PxTIG12RRHS

The proposed course 1) has explicit academic content and requirements for receiving credits, 2) is in alignment with the program's learning outcomes, 3) adheres to policies and Skoltech regulations.

Lead Instructor confirms