1. Annotation

Course Description

The course addresses Bayesian methods for solving various machine learning and data analysis problems (classification, regression, dimension reduction, topic modeling, etc.).

The course starts with an overview of canonical machine learning (ML) applications and problems, learning scenarios, etc. and then introduces foundations of Bayesian approach to solve these problems. Bayesian approach allows one to take into account subject domain knowledge and/or user’s preferences through a prior distribution when constructing the model. Besides, it offers an efficient framework for model selection. We discuss which prior distributions types are usually used, limit properties of a posterior distribution, and provide some illustrations of the Bayesian approach.

The practical applicability of Bayesian methods in the last 20 years has been greatly enhanced through the development of a range of approximate inference algorithms such as variational Bayes and expectation propagation, as well as posterior simulation methods based on the Markov chain Monte Carlo approach. As a result Bayesian methods have grown from a specialist niche to become mainstream. Therefore, we devoted a second part of the course to approximation tools, vitally important for Bayesian inference, and provide examples how to use Bayesian approaches to automatically select features, tune the regularization parameter in regression and classification, etc.

The last part of the course is devoted to advanced Bayesian methods, namely, Gaussian Processes and deep Bayesian neural networks, which have become widespread in the last 5-8 years. We discuss deep Bayesian framework and then illustrate its applications through construction of deep variational autoencoders, approaches to variational dropout, Wasserstein Generative Adversarial Networks, deep Kalman filter, etc. Home assignments include solution of applied problems, development of modifications of Bayesian ML algorithms, and some theoretical exercises.
This is an Elective course in Term 5. Course prerequisites among the Skoltech courses are Numerical Linear Algebra and Optimization Methods, as well as course on Machine Learning and Applications. As well we suppose an attendee be fluent with linear algebra, probability and real analysis.

2. Structure and Content

<table>
<thead>
<tr>
<th>Course Academic Level</th>
<th>Master-level course suitable for PhD students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of ECTS credits</td>
<td>6</td>
</tr>
<tr>
<td>Topic</td>
<td>Summary of Topic</td>
</tr>
<tr>
<td>-------</td>
<td>------------------</td>
</tr>
</tbody>
</table>
| 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), Laplace Approximation, Automatic Relevance Determination  
Theory: individual home assignments | 6 | 2 | 0 |
| Non-exact Inference: Variational Approaches | - Expectation-Maximization. Principal Component Analysis  
- Stochastic Variational Inference. Latent Dirichlet Allocation  
Practice: home assignment 1 | 6 | 3 | 0 |
| Non-exact Inference: Deep Bayes | - Variational Auto-Encoding models  
- Stochastic gradients estimators and variance reduction  
- Variational Dropout  
- Generative Adversarial Models (GAN). Alpha-GAN, Wasserstein GAN  
Theory: individual home assignments | 8 | 5 | 0 |
| Non-Exact Inference: MCMC Approaches | - Metropolis-Hastings. Langevin Dynamics. Scalable Langevin Dynamics  
- Gibbs Sampling. Applications to Matrix Completion (NetFlix problem)  
Practice: home assignment 2. Project announcement | 5 | 3 | 0 |
| Gaussian Processes for Bayesian Machine Learning | - Gaussian Process model (GP). Kernels and RKHS. Multi-output GP. GLM GP  
- Scalability issues. Induced points, Fourier features  
- Bayesian optimization. Variational optimization  
Theory: individual home assignments | 6 | 4 | 0 |
| Bayesian Models for Sequential Data | - Hidden Markov Models (HMM). Belief propogation for HMM  
- Kalman Filter  
Practice: home assignment 3. Presentation of project intermediate results | 5 | 5 | 0 |

3. Assignments

<table>
<thead>
<tr>
<th>Assignment Type</th>
<th>Assignment Summary</th>
</tr>
</thead>
</table>
### Homework

Example of problems:

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 \( \text{Gam}(\beta|c_0,d_0) \) (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.
Examples of the topics:
1. Deep kernels and Gaussian processes
2. Bayesian Active Learning
3. Bayesian black-box optimization
4. Multi-Fidelity Gaussian Process regression
5. Bayesian change-point detection
6. Comparison of approaches for approximation of intractable Bayesian models
7. MCMC for Bayesian inference

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 algorithmic projects only), and excluding references/contributions
• Source code (scripts, notebooks) in ZIP or on Github

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

There are two tests:
- one is in the middle of the course (midterm exam)
- the final test at the end of the course (final exam)

The tests are in the form of multiple choice questions and short problems.

E.g.
- Does Variational Inference denote a set of methods to estimate variance of the model?
  Answer: Yes/No. Provide short comments
- Approximate the evidence for the Bayesian linear regression. Provide brief calculations
4. Grading

<table>
<thead>
<tr>
<th>Type of Assessment</th>
<th>Graded</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Grade Structure</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity Type</td>
<td>Activity weight, %</td>
</tr>
<tr>
<td>Homework Assignments</td>
<td>30</td>
</tr>
<tr>
<td>Midterm Exam</td>
<td>15</td>
</tr>
<tr>
<td>Projects</td>
<td>35</td>
</tr>
<tr>
<td>Final Exam</td>
<td>20</td>
</tr>
</tbody>
</table>

Grading Scale

<table>
<thead>
<tr>
<th>Grade</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>86</td>
</tr>
<tr>
<td>B</td>
<td>76</td>
</tr>
<tr>
<td>C</td>
<td>66</td>
</tr>
<tr>
<td>D</td>
<td>56</td>
</tr>
<tr>
<td>E</td>
<td>40</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
</tr>
</tbody>
</table>

Attendance Requirements: Mandatory

5. Basic Information

Maximum Number of Students

<table>
<thead>
<tr>
<th></th>
<th>Maximum Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall:</td>
<td>40</td>
</tr>
<tr>
<td>Per Group (for seminars and labs):</td>
<td>40</td>
</tr>
</tbody>
</table>

Course Stream: Science, Technology and Engineering (STE)

Course Term (in context of Academic Year): Term 1

Course Delivery Frequency: Every year

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

<table>
<thead>
<tr>
<th>Course Tags</th>
<th>Masters Programs</th>
<th>PhD Programs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computational Science and Engineering</td>
<td></td>
<td>Computational and Data Science and Engineering</td>
</tr>
<tr>
<td>Data Science</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Course Tags

| Math Programming |

6. Textbooks and Internet Resources

<table>
<thead>
<tr>
<th>Required Textbooks</th>
<th>ISBN-13 (or ISBN-10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bishop, C.M. Pattern Recognition and Machine Learning. Springer, 2007</td>
<td>978-0387310732</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Recommended Textbooks</th>
<th>ISBN-13 (or ISBN-10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peter Muller, Fernando Andres Quintana, Alejandro Jara, Tim Hanson. Bayesian Nonparametric Data analysis. Springer, 2015</td>
<td>9783319189673</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Web-resources (links)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage">http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage</a></td>
<td>online version of Barber’s &quot;Bayesian Reasoning and Machine Learning&quot;</td>
</tr>
</tbody>
</table>

7. Facilities
Access to the Internet through a computer class and Wi-Fi network of the institute.

Whiteboard/Flipchart; Projector

- Python 3.4+ (preferable) or 2.7+
- Jupyter
- Open source python libraries including machine learning ones
  - Scikit learn
  - BayesPy
  - XGboost
  - Pandas
  - Matplotlib
  - Seaborn
  - pytorch and pyro

### 8. Learning Outcomes

#### Knowledge

Obtain a big picture of practical problems exploiting Bayesian ML methods; applications include automatic feature selection and predictive modeling applications, Bayesian optimization, topics modeling, etc.

- Know main ML problem statements;
- Know available standard Bayesian ML methods and areas of their applications;
- Know functionality and constraints of existing ML algorithmic software libraries, containing realizations of Bayesian ML methods (Scikit-learn, BayesPy, etc.);
- Know the theoretical basis and conceptual tools needed for the investigation and justification of Bayesian algorithms;

#### Skill

- Be able to formulate in mathematical terms a real-world problem, built a corresponding probabilistic model, select an appropriate Bayesian inference method;
- Be able to apply Bayesian ML methods from existing ML algorithmic software libraries (Scikit-learn, BayesPy, etc.) and interpret obtained results in subject domain terms;
- Be able to develop either exact or approximate Bayesian inference algorithms for probabilistic models and implement them into efficient programming code;
- Be able to formulate the domain-specific knowledge in terms of a prior distribution;
- Be able to read and discuss research papers on probabilistic framework, Bayesian ML methods and their applications;

#### Experience

Obtain a sufficient experience during practical exercises and project activities to become a qualified user of Bayesian ML methods.

### 9. Assessment Criteria
### Input or Upload Example(s) of Assignment 1:

<table>
<thead>
<tr>
<th>Select Assignment 1 Type</th>
<th>Team Project</th>
</tr>
</thead>
</table>

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 algorithmic projects only), and excluding references/contributions  
- Source code (scripts, notebooks) in ZIP or on Github

### Or Upload Example(s) of Assignment 1

https://ucarecdn.com/035792cb-4ba4-41c9-8e2f-6c8e6763c8d6/

### 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

### Input or Upload Example(s) of Assignment 2:

<table>
<thead>
<tr>
<th>Select Assignment 2 Type</th>
<th>Test/Quiz</th>
</tr>
</thead>
</table>

---

The information is organized in a table format, with headers and subheaders for each section. The text is readable and well-structured, making it easy to understand the course requirements and assessment criteria.
There are two tests:
- one is in the middle of the course (midterm exam)
- the final test at the end of the course (final exam)

The tests are in the form of multiple choice questions and short problems.

E.g.
- Does Variational Inference denotes a set of methods to estimate variance of the model? Answer: Yes/No. Provide short comments
- Approximate the evidence for the Bayesian linear regression. Provide brief calculations

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

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 $\text{Gam}(\beta|c_0,d_0)$ (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.

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

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

1. Exact Inference & GLM
   - S. AMARI and A. CICHOCKI. Information geometry of divergence

2. Non Exact Inference: Variational Methods
• Shun-ichi Amari. Natural Gradient Works Efficiently in Learning.
• Anderson Y. Zhang. Theoretical and computational guarantees of mean field variational inference for community detection. url:
3. Non Exact Inference: Deep Learning and Bayesian Methods

- Handbook of Markov Chain Monte Carlo. url: http://www.mcmchandbook.net/
5. Gaussian Process and Variational Optimization


6. Sequential Data