Module Leader

Prof Z Ghahramani [1]

Lecturers

Prof Z Ghahramani & Dr M Hernandez-Lobato [2]

Timing and Structure

Michaelmas term. 14 lectures + 2 examples classes. Assessment: 100% coursework

Prerequisites

3F3 useful

Aims

The aims of the course are to:

- introduce students to basic concepts in machine learning, focusing on statistical methods for supervised and unsupervised learning.

Objectives

As specific objectives, by the end of the course students should be able to:

- demonstrate a good understanding of basic concepts in statistical machine learning.
- apply basic ML methods to practical problems.

Content

Machine learning (ML) is an interdisciplinary field focusing on both the mathematical foundations and practical applications of systems that learn, reason and act. The goal of machine learning is to automatically extract knowledge from observed data for the purposes of making predictions, decisions and understanding the world.

The aim of this module is to introduce students to basic concepts in machine learning, focusing on statistical methods for supervised and unsupervised learning. The module will be structured around three recent illustrative successful applications: Gaussian processes for regression and classification, Latent Dirichlet Allocation models for unsupervised text modelling and the TrueSkill probabilistic ranking model.

- Linear models, maximum likelihood and Bayesian inference
- Gaussian distribution and Gaussian process
- Model selection
- The Expectation Propagation (EP) algorithm
- Latent variable models
- The Expectation Maximization (EM) algorithm
- Dirichlet Distribution and Dirichlet Process
- Variational inference
- Generative models, graphical models: Factor graphs

Lectures will be supported by Octave/MATLAB demonstrations.

A detailed syllabus and information about the coursework is available on the moodle website: [https://www.vle.cam.ac.uk/course/view.php?id=69021](https://www.vle.cam.ac.uk/course/view.php?id=69021)

## Coursework

<table>
<thead>
<tr>
<th>Coursework activity #1 Gaussian Processes</th>
<th>Format</th>
<th>Due date &amp; marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coursework 1 brief description</td>
<td>Individual/group Report / Presentation</td>
<td>Fri week 5</td>
</tr>
<tr>
<td>Learning objective:</td>
<td></td>
<td>Nonanonymously marked for MPHIL/MLSALT &amp; Undergraduates</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[20/60]</td>
</tr>
</tbody>
</table>

- To gain experience in Bayesian Gaussian Process (GP) regression.
- To familiarise yourself with the GPML toolbox.
- To understand properties of covariance functions.
- To perform hyperparameter learning.
- To understand how model selection can be done using the marginal likelihood.

<table>
<thead>
<tr>
<th>Coursework activity #2 Probabilistic Ranking</th>
<th>Format</th>
<th>Due date &amp; marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coursework 2 brief description</td>
<td>Individual Report</td>
<td>Fri week 7</td>
</tr>
<tr>
<td>Learning objective:</td>
<td>Anonymously marked for MPHIL/MLSALT &amp; Undergraduates</td>
<td>[20/60]</td>
</tr>
<tr>
<td></td>
<td>Nonanonymously marked for PhDs</td>
<td></td>
</tr>
</tbody>
</table>

- To understand inference in continuous probabilistic models represented as factor graphs.
- To understand the Gibbs sampling algorithm and gain experience with using Markov chain Monte Carlo (MCMC) for inference.
- To understand message passing on (loopy) factor graphs.
- To learn how to construct approximate messages using Expectation Propagation (EP).

<table>
<thead>
<tr>
<th>Coursework activity #3 Latent Dirichlet Allocation models for documents</th>
<th>Format</th>
<th>Due date &amp; marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coursework 3 brief description</td>
<td>Individual Report</td>
<td>Fri week 9</td>
</tr>
<tr>
<td>Learning objective:</td>
<td>Anonymously marked for MPHIL/MLSALT &amp; Undergraduates</td>
<td>[20/60]</td>
</tr>
<tr>
<td></td>
<td>Nonanonymously marked for PhDs</td>
<td></td>
</tr>
</tbody>
</table>

- To understand unsupervised learning in discrete graphical models for documents.
- To develop an understanding of graphical models with more complex latent structure.
- To understand and apply the Expectation Maximization (EM) and Gibbs sampling algorithms.
Coursework

<table>
<thead>
<tr>
<th>Format</th>
<th>Due date &amp; marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>• To perform unsupervised learning using Latent Dirichlet Allocation model on a collection of documents.</td>
<td></td>
</tr>
</tbody>
</table>

**Booklists**

Please refer to the Booklist for Part IIB Courses for references to this module, this can be found on the associated Moodle course.

**Examination Guidelines**

Please refer to [Form & conduct of the examinations](http://teaching.eng.cam.ac.uk/content/form-conduct-examinations) [4].

**UK-SPEC**

The [UK Standard for Professional Engineering Competence (UK-SPEC)](http://www.engc.org.uk/ukspec.aspx) [5] describes the requirements that have to be met in order to become a Chartered Engineer, and gives examples of ways of doing this.

UK-SPEC is published by the Engineering Council on behalf of the UK engineering profession. The standard has been developed, and is regularly updated, by panels representing professional engineering institutions, employers and engineering educators. Of particular relevance here is the 'Accreditation of Higher Education Programmes' (AHEP) document [6] which sets out the standard for degree accreditation.

The [Output Standards Matrices](http://teaching.eng.cam.ac.uk/content/output-standards-matrices) [7] indicate where each of the Output Criteria as specified in the AHEP 3rd edition document is addressed within the Engineering and Manufacturing Engineering Triposes.

Last modified: 07/10/2020 16:34