Eötvös Loránd University  
Faculty of Informatics  
Computer Science Msc (Data Science specialization)

Description of the courses

**Name of the course: Advanced Machine Learning**

**Description**

We shall deal with practical problems in this course via deep learning approaches. Network architectures, image processing, speech processing, motion and control, anomaly detection make the core of the course. The course enables the student for researching and tracking the literature, helps to deepen her/his knowledge in diverse ways that includes the deepening of the mathematics, physics and control related knowledge among others.

**Literature**

**Compulsory**

**Recommended**

**Name of the course: Data models and databases**

**Description**

The course is about the following topics: E/R modeling, the relational data model, normal forms, SQL, referential integrity, security, transactions and data warehousing.

**Literature**

**Compulsory**
### Name of the course: Foundations of data science

**Description**

Basic univariate and multivariate statistics: location (e.g. mode, median, quantiles), dispersion (e.g. range, deviation, variance), covariance, correlation (e.g. Pearson, Spearman), contingency tables, data visualization.

Basic concepts of probability theory: the Bayes rule, continuous and discrete distributions (e.g. Gaussian, Bernoulli).

Basic concepts from geometry: vectors and points, hyperplanes, distances and metrics.

Basic concepts from linear algebra: vectors and matrices, eigenvectors and eigenvalues, matrix decompositions, transformations (e.g. PCA, SVD).

Basic concepts from information theory (e.g. entropy).

**Literature**

**Compulsory**


---

### Name of the course: Interactive media design and development

**Description**

The course introduces Human–Computer Interaction (HCI) involving the study, planning, and design of the interaction between people (users) and computers.

Its aim is to understand the theoretical basics of Perception, Multimedia design, Information Visualization, Interaction Design, the Virtual Continuum, Serious Games, Tangible, Collaborative, Location-based, and Gesture-based technologies, etc.) and recent innovations in these areas.

Activities involve the exploration of emerging interactive technologies designed for demonstration, education, entertainment, navigation, narrative, support …etc. purposes and their variety of creative applications in different disciplines and user interest groups.

Students from different disciplines form groups to design and implement a specified innovative project that could well serve the basis of an industrial entrepreneurship.
Literature

Compulsory

- http://intmedia.elte.hu/ (ELTE IK 2013)


Recommended


- Journal of Virtual World Research: http://jvwresearch.org/

Name of the course: Introduction to data science

Description

clustering;
- frequent pattern mining;
- linear classification and regression model: model parameters and hyper-parameters, validation, overfitting-underfitting and the bias-variance trade-off;
- introduction to prediction techniques (as black-box functions);
- data quality and pre-processing: noise, missing values, data transformation, normalization;
- the CRISP-DM process;
- recommendation techniques;

Literature

Compulsory


- Jiawei Han, Micheline Kamber, Jian Pei (2011). Data Mining: Concepts and Techniques. Morgan Kaufmann.

### Name of the course: Machine learning

**Description**

- decision trees;
- support vector machines and kernel methods;
- graphical and probabilistic models;
- neural networks;
- factorization techniques;
- semi-supervised learning;
- ensemble techniques, bagging, boosting;
- time-series mining;
- text-mining;

**Literature**

**Compulsory**


### Name of the course: Network science

**Description**

- Introduction
- Networks and complex systems
- Random graphs
- Small world networks
- Scale free networks
- Growing network models
- Analysis and visualization of networks
- Robustness of networks
- Walking and searching on networks
- Spreading and traffic on networks
- Examples of networks: internet, social networks

**Literature**

**Compulsory**

- A. Barrat, M. Barthélémy and A. Vespignani: Dynamical Processes on Complex Networks (Cambridge UP, 2008)
<table>
<thead>
<tr>
<th>Name of the course: Open-source technologies for real-time data analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
</tr>
<tr>
<td>data analysis in Unix-based operating systems;</td>
</tr>
<tr>
<td>data acquisition and storage;</td>
</tr>
<tr>
<td>data preparation: transformation, validation &amp; cleaning;</td>
</tr>
<tr>
<td>data analysis in Spark, TensorFlow, Keras, scikit-learn;</td>
</tr>
<tr>
<td>graph analytics in GraphX;</td>
</tr>
<tr>
<td>data visualization: Tableau, Kibana, seaborn;</td>
</tr>
<tr>
<td><strong>Literature</strong></td>
</tr>
<tr>
<td><strong>Compulsory</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name of the course: Optimization for data science</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
</tr>
<tr>
<td>basic concepts from optimization: stochastic gradient descent, linear programming, lagrange multiplier;</td>
</tr>
<tr>
<td>basic concepts from graph theory: skeleton, shortest paths, trees, structural properties of graphs;</td>
</tr>
<tr>
<td>stochastic processes: gibbs sampling, markov chain monte carlo, multi-armed bandits, nature-inspired optimization;</td>
</tr>
<tr>
<td><strong>Literature</strong></td>
</tr>
<tr>
<td><strong>Compulsory</strong></td>
</tr>
<tr>
<td>• Robert M. Gover - Alexandre Gamfort, Optimization for Data Science, Master 2 Data Science, Univ. Paris Saclay</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name of the course: Sensor data analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
</tr>
<tr>
<td>3D sensing: Stereo vision, photometric stereo, structured light, Time-of-Flight (ToF) depth cameras, LIDAR devices, principles of operation, sensor calibration;</td>
</tr>
<tr>
<td>Comparison of 3D sensors: Advantages and drawbacks, conditions of use and possible failures, areas of application, processing speed, data flow;</td>
</tr>
<tr>
<td>Basic algorithms behind 3D sensing: Active and passive stereo, 3D data processing, segmentation and</td>
</tr>
</tbody>
</table>
grouping, multiple 3D data registration, fast implementation on GPU;
3D sensor fusion: Sensor combinations in different tasks, ToF-stereo fusion, temporal and spatial depth upsampling, image and video based depth upsampling, fusing optical images to 3D shapes and LIDAR data, image based LIDAR data upsampling

**Literature**

**Compulsory**

**Name of the course: Software engineering / Data science lab**

**Description**
During the lab, students will work in teams on data science tasks on real data gathered from industrial as well as academic partners of the Faculty of Informatics. The tasks will concern both basic and applied research, under the supervision of experienced data scientists, necessary for delivering the results in a desired quality. The projects will follow suitable industrial data science methodologies such that, for example, the CRISP-DM process. Emphasis will be given on delivering prototype solutions to the determined tasks concerning, but not limited to, data pre-processing, data transformation, data visualization, modeling (model selection, hyperparameter tuning, model combination, etc.), model evaluation as well as deployment, real-time data analytics for descriptive and predictive mining, anomaly detection, just to name a few. The concrete tasks will be determined by the industrial partners and they will play an important role also in the evaluation of the delivered solutions/prototypes. If applicable, teams will participate in data mining challenges (e.g. Kaggle).

**Literature**

**Compulsory**

**Recommended**

**Name of the course: Software for Advanced Machine Learning**

**Description**
During the semester, students learn about different deep learning techniques and deep neural network architectures. The course focuses on the most popular software tools used in the implementation of deep
learning methods.

Topics of the course in short:
- Static and dynamic computational graphs in machine learning frameworks
- Implementation of simple supervised neural network models (dense networks, convolutional networks, recurrent networks)
- Implementation of complex supervised neural network models (residual networks)
- Implementation of unsupervised neural network models (autoencoder networks, generative networks)
- Implementation of custom neural network layers
- Implementation of custom training procedures

**Literature**

**Compulsory**

---

**Name of the course: Software technology**

**Description**

**Purpose:**
The course gives a broad overview of the process and methodologies of software development and its execution. We cover all phases of development from requirements to maintenance and quality assurance with emphasize on architectural design. The course tries to deliver a balanced mixture of theoretical knowledge and practical skills with currently used technologies.

**Competencies delivered:**
Students completing the class will understand software development process, its different strategies and methodologies. They will be able to make sensible architectural decisions and plans well in advance using the acquired mixture of theoretical and hands-on skills.

**Prerequisites:**
- advanced knowledge of at least one object oriented programming language
- understanding of web technologies full stack (client, database, server...)
- (optional) project experience

**Literature**

**Compulsory**
### Name of the course: Stream mining

**Description**

- Challenges: one-pass constraint, change diagnosis, concept drift, concept evolution, feature evolution, load shedding;
- The sliding window computation model;
-Synopsis construction;
-Clustering data streams;
-Data stream classification;
-Frequent pattern mining in data streams;
-Dimensionality reduction and forecasting in data streams;
-Indexing and querying data streams;
-Distributed mining of data streams;

**Literature**

**Compulsory**


### Name of the course: Theory of programming

**Description**


**Literature**

**Compulsory**

Name of the course: Web engineering

Description

This curriculum introduces the students with the modern, state-of-the-art client and server side web technologies, methodologies of web engineering, the programming and design patterns, especially with the web service oriented architectures. By the end of the course the student has a global overview of the up-to-date web trends and technologies, and, with the help of them, is able to develop a web application and web information systems.


Literature

Compulsory