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 architectres, image processing, speech processing, motion and control, anomaly detection make the core of the couse. 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

- Ian Goodfellow and Yoshua Bengio and Aaron Courville: **Deep Learning**. MIT Press, 2016. Hardcover: ISBN: 9780262035613, eBook: ISBN: 9780262337434

Recommended

- Aurélien Geron: **Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems** O'Reilly Media, Inc. 2017. ISBN: 1491962291
- Josh Patterson and Adam Gibson: **Deep Learning: A Practitioner's Approach**. O'Reilly Media, Inc. 2017. ISBN: 1491914254
- Sayan Pathak, Roland Fernandez, and Jonathan Sanito, **Deep Learning Explained,** MOOC edX, https://www.edx.org/course/deep-learning-explained-microsoft-dat236x

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

- Database Systems The Complete Book (2nd edition) – 2014 by Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom, ISBN-13: 978-0131873254

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

- William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1-2, 3rd Edition, 2008, John Wiley & Sons Inc., ISBN-13: 978-8126518050
- John A, Rice, *Mathematical Statistics and Data Analysis*, 3rd Edition, 2010, Cengage Learning, ISBN-13: 978-8131519547
- K. W. Gruenberg, A. J. Weir, *Linear Geometry*, 2nd Edition, 1977, Springer, ISBN-13: 978-0387902272
- Gilbert Strang, *Introduction to Linear Algebra*, 5th Edition, 2016, Wellesley-Cambridge Press, ISBN: 978-09802327-7-6
- Thomas M. Cover, Joy A. Thomas, *Elements of Information Theory*, 2nd Edition, 2006, Wiley Series in Telecommunications and Signal Processing, ISBN-13: 978-0471241959

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 Continum, 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)
- The Encyclopedia of Human-Computer Interaction, 2nd Ed. At: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed

Recommended

- Ioannis Deliyannis, Interactive Multimedia, ISBN 978-953-51-0224-3, Hard cover, 312 pages, Publisher: InTech, Chapters published March 07, 2012 under CC BY 3.0 license OpenAccess: http://www.intechopen.com/books/interactive-multimedia
- Annetta and S. C. Bronack, (eds.), Serious Educational Game Assessment: Practical Methods and Models for Educational Games, Simulations and Virtual Worlds, 1–18. © 2011 Sense Publishers. ISBN: 978-94-6091-327-3 (paperback)
- The Functional Art: An Introduction to Information Graphics and Visualization (Peachpit/Pearson Education, 2012): http://www.thefunctionalart.com/ ISBN-13: 978-0321834737
- Ed. Xin-Xing Tang, Virtual Reality - Human Computer Interaction, ISBN 978-953-51-0721-7, Hard cover, 306 pages, Publisher: InTech, Chapters published September 05, 2012 under CC BY 3.0 license, OpenAccess: http://www.intechopen.com/books/virtual-reality-human-computer-interaction
- 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

- Peter Flach (2012). Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press.
- Jiawei Han, Micheline Kamber, Jian Pei (2011). Data Mining: Concepts and Techniques. Morgan Kaufmann.
- Pang-Ning Tan, Michael Steinbach, Vipin Kumar (2005). Introduction to Data Mining. Addison Wesley

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

- Ethem Alpaydin (2014). Introduction to Machine Learning. The MIT Press.
- Shai Shalev-Shwartz, Shai Ben-David (2014). Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press.
- Christopher Bishop (2006). Pattern Recognition and Machine Learning. Springer-Verlag.

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

- Romualdo Pastor-Satorras, Alessandro Vespignani : Evolution and Structure of the Internet (Cambridge University Press, 2004)
- M.E.J. Newman: Networks – An Introduction (Oxford UP, 2010)
- A.-L. Barabási: Network Science (Cambridge UP, 2016)
- A. Barrat, M. Barthélemy and A. Vespignani: Dynamical Processes on Complex Networks (Cambridge UP, 2008)

Name of the course: **Open-source technologies for real-time data analytics**

Description

data analysis in Unix-based operating systems;
data acquistion and storage;
data preparation: transformation, validation & cleaning;
data analysis in Spark, TensorFlow, Keras, scikit-learn;
graph analytics in GraphX;
data visualization: Tableau, Kibana, seaborn;

Literature

Compulsory

- Sam R. Alapati (2017). Expert Hadoop Administration – Managing, Tuning, and Securing Spark, YARN, and HDFS. Addison-Wesley.
- Robert Johansson (2015). Numerical Python: A Practical Techniques Approach for Industry. Apress.
- Aurélien Géron (2017). Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media.
- Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher J. Pal (2016). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Series in Data Management Systems – 4th Edition.
- Sebastian Raschka, Vahid Mirjalili (2017). Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow – 2nd Edition. Packt Publishing.

Name of the course: **Optimization for data science**

Description

basic concepts from optimalization: stochastic gradient descent, linear programming, lagrange multiplier;
basic concepts from graph theory: skeleton, shortest paths, trees, structural properties of graphs;
stochastic processes: gibbs sampling, markov chain monte carlo, multi-armed bandits, nature-inspired optimization;

Literature

Compulsory

- Robert M. Gover - Alexandre Gamfort, Optimization for Data Science, Master 2 Data Science, Univ. Paris Saclay

Name of the course: **Sensor data analytics**

Description

3D sensing: Stereo vision, photometric stereo, structured light, Time-of-Flight (ToF) depth cameras, LIDAR devices, principles of operation, sensor calibration;
Comparison of 3D sensors: Advantages and drawbacks, conditions of use and possible failures, areas of application, processing speed, data flow;
Basic algorithms behind 3D sensing: Active and passive stereo, 3D data processing, segmentation and

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

- M.Sonka, V.Hlavac, R.Boyle, "Image Processing, Analysis and Machine Vision", Thomson
- R.Szeliski, „Computer Vision: Algorithms and Applications", Springer
- Eichhardt, D. Chetverikov, Z. Jankó, „Image-guided ToF depth upsampling: a survey", Machine Vision and Applications, vol. 28, pp. 267–282, 2017.
- M. Grzegorzek et al. (Eds), „Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications", Springer, 2013.
- Wikipedia, „Lidar", https://en.wikipedia.org/wiki/Lidar, 2018.

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, hyper-parameter 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

- Ian H. Witten, Eibe Frank, Mark A. Hall (2011). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.

Recommended

- Jake VanderPlas (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media.
- Sebastian Raschka, Vahid Mirjalili (2017). Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow. Packt Publishing.

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

- Goodfellow, Y. Bengio, A. Courville, **Deep Learning**, MIT Press, 2016, ISBN: 9780262035613
- F. Chollet, **Deep Learning with Python**, Manning Publications Co., 2017, ISBN: 9781617294433
- I. Lieder, Y. Resheff, T. Hope, **Learning TensorFlow: A Guide to Building Deep Learning System**, O'Reilly Media, 2017, ISBN: 9781491978511

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

Literature

Compulsory

- R. C. Martin: Clean Code: A Handbook of Agile Software Craftsmanship, Prentice Hall 2008. ISBN-13: 978-0132350884
- J. Humble, D. Farley: Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation, Addison-Wesley 2010. ISBN-13: 978-0321601919
- E. Gamma, R. Helm, R. Johnson, J. Vlissides: Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley, 1994. ISBN-13: 978-0201633610
- M. R. Blaha, J. R. Rumbaugh: Object-Oriented Modeling and Design with UML, Pearson, 2004. ISBN-13: 978-0130159205
- L. Bass, P. Clements, R. Kazman: Software Architecture in Practice. 3rd ed. Addison-Wesley

Professional, 2012. ISBN-13: 978-0321815736

Recommended

- F. P. Brooks: The Mythical Man-Month: Essays on Software Engineering, Addison-Wesley 1995. ISBN-13: 978-0201835953

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

- Charu C. Aggarwal. Data Streams - Models and Algorithms. Springer, 2007.
- A. Bifet. Adaptive Stream Mining: Pattern Learning and Mining from Evolving Data Streams. IOS Press, 2010.
- B. Ellis. Real-Time Analytics: Techniques to Analyze and Visualize Streaming Data. Wiley, 2014.
- A. G. Psaltis. Streaming Data: Understanding the real-time pipeline. Manning Publications, 2016.
- Henrique C. M. Andrade, Buğra Gedik, Deepak S. Turaga. Fundamentals of Stream Processing: Application Design, Systems, and Analytics. Cambridge, 2014.

Name of the course: **Theory of programming**

Description

Basic notions of programming. The syntax and semantics of nondeterministic programs. Partial and total correctness. Weakest precondition. The notion of loop invariant. Verification rules of program constructs. Verification: a method for proving total correctness of deterministic and nondeterministic programs. Synthesizing correct sequential programs by using the derivation rules. The correctness of concurrent programs, verification rules of the new statements (parbegin/parend, await). Owicki-Gries method for proving the total correctness of parallel programs, deadlock freedom and interference freedom.

Literature

Compulsory

- K. R. Apt, E.-R. Olderog. Verication of Sequential and Concurrent Program. Springer-Verlag, 1997. ISBN 978-1-84882-744-8

- S. Owicki, D. Gries. An axiomatic proof technique for parallel programs. Acta Inf., 6, pp. 319-340, 1976
- E. W. Dijkstra. A Discipline of Programming. Prentice-Hall, Englewood Cliffs, New York, 1976. ISBN-13: 978-0132158718
- D. Gries, The Science of Programming, Springer, Berlin, 1981. ISBN: 978-1-4612-5983-1

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.
Introduction to Web Technologies and Web Engineering: specialties, characteristics, categories of web applications. Web Architectures: multi-tier, data-centric architectures, Requirement Analysis of Web Applications Specialties of Large Enterprise and Small and Medium Enterprise Web Applications Development Process of Web Applications Model-Based Web Application Design and Development, WebML Testing, Quality Management. Design of Web 2.0 és Enterprise 2.0 Applications Web Business Models Web project management Design of Mobile Web Applications Semantic Web Applications, integration to Web Information Systems Web Application Models, Cloud computing Service Oriented Architectures, Web Information Systems.

Literature

Compulsory

- Kappel, G., Pröll, B., Reich, S., Retschitzegger W. (Eds.): Web Engineering: The Discipline of Systematic Development. John Wiley & Sons Inc., Chichester (2006).
- Mendes, E., Mosley, N. (Eds.): Web engineering. Springer-Verlag, Berlin (2005).
- Murugesan, S., Deshpande, Y. (Eds.): Web Engineering: Managing Diversity and Complexity of Web Application Development. LNCS 2016, Springer-Verlag, Berlin (2001).