

Data Mining and Machine Learning

Lessons 3 hours, Lab 2 hours, 4 **Points**

Previsories: 61760 Probability for software engineering, 61747-Data structure

The course is devoted to principles and advanced uses of data mining. During the course, subjects are studied in the classification and embedding of data in statistical and computer imaging. Also, the methods for supervised and unsupervised learning and methods of reducing the dimension and deep learning. The methods are demonstrated using text mining and market models. The course is accompanied by a lab and a final project performed on MATLAB or PHYTON systems.

Course Items

1. Introduction. Text mining problems
2. Statistical methods. Correlation and linear regression
3. Clustering methods:
 - Distribution methods (k-means, EM)
 - Hierarchical clustering
4. Classification methods:
 - k-NN classifier
 - Decision trees
 - SVM
5. Artificial neural network, Logistic regression. Perceptron
6. Back- Propagation approach.
7. Introduction to Deep Learning:
 - Convolutional Neural Networks
 - LSTM as RNN (Recurrent Neural Networks)
 - Word Embedding Networks (optionally)

Literature

1. Goodfellow I., Bengio Y., Courville A., Deep Learning (Adaptive Computation and Machine Learning series) Hardcover – November 18, 2016
2. Rogers S., Girolami M., A First Course in Machine Learning, Second Edition, August 15, 2016 by Chapman and Hall/CRC
3. Witten, I.H. and Frank, E., *Data Mining Practical Machine Learning Tools and Techniques*, 3nd ed., The Morgan Kaufmann Series in Data Management Systems, 2011.
4. Bishop, C.M., *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, 2007.
5. Stork, D.G. and Yom-Tov, E., *Computer Manual in MATLAB to Accompany Pattern Classification*, 2nd ed., Wiley-Interscience, 2004.

6. Tan, P.N. and Steinbach, M. and Kumar, V., *Introduction to Data Mining*, Addison Wesley, 2005.

Course requirements and grading composition

- a valid grade will only be calculated in case of a passing grade on the final exam(55)
- Home works: 10%
- Project summary (: 40%
- Final exam: 50%

Teaching outcomes

Upon successful completion of this course, students will be able to:

1. Have a good understanding of the fundamental issues and challenges of machine learning and data mining methodologies: data, model and approach selection, results verification.
2. Have a perception of the strengths and weaknesses of the conventional machine learning approaches.
3. Appreciate the fundamental precise relationships within and across algorithms and the paradigms of supervised and unsupervised learning.
4. Be able to conceive and implement basic machine learning approaches in real-world applications.
5. Adapt analysis and tools learned in class into given modified scenarios.