Natural Language Processing

Course code: MLC_NLP

This course is focused on the analysis and processing of text data. We expect knowledge of basic principles of machine learning to the same extent as the Introduction to Machine Learning course provides. Special attention will be paid to text preprocessing and vectorization, which is crucial for NLP. We will further focus on text classification, language modeling and text synthesis.

Required skills

- basic knowledge of programing in Python
- high school level of mathematics
- Basics of machine learning on the level of our course Introduction to machine Learning

Course outline

- Introduction to natural language processing
- Chapters from computational linguistics (corpus, tokenization, morphological, syntactic and semantic analysis, entropy, perplexity)
- Text document vectorization (bag of words, one-hot encoding, TF-IDF)
- Practical taks on text classification
- Word embedding (word2vec, GloVe)
- Introduction to language modelling (n-gram models, smoothing, neural network based language models)
- Practical task on language modelling (implementation of a language detection algorithm based on language models)
- Neural network based text generator