

Machine Learning Workshop

The TAD center is giving a two half-day machine learning workshop. In this workshop we will present advanced subjects in machine learning and discuss how to apply them. The workshop will not contain hands on sessions, as it is aimed to expose new subjects to the audience. For discussions on applying these ideas in concrete cases, we welcome people to consult with us.

The intended audience are researchers and graduate students with experience in data analysis using multivariate statistics methods. The workshop is condensed and we assume some prior knowledge on the subject.

Day I – Machine Learning

1. Overview of AI landscape with an emphasis on ML: (30min)
 - a. AI, ML, Statistics and their relations
 - b. supervised vs unsupervised
 - c. classification vs regression
 - d. discriminatory vs generative
 - e. matching model types and data:
 - i. what usually works on tabular data
 - ii. what works best for temporal/spatial relation within the data
 - iii. combine data type
 - f. black-box vs explainable models
2. Data preprocessing (50min)
 - a. Exploration and visualization
 - b. Missing values
 - c. Imbalanced data and re-weighting
 - d. Scaling
 - e. Dimensionality reduction

Break (5min)

3. Model evaluation and selection (55min)
 - a. Quality measures
 - b. Model selection - in-sample methods vs. data driven
 - i. Statistical tests and their assumptions
 - ii. Validation and cross validation
 - c. Bias variance tradeoff
 - d. Regularization
 - e. Hyper-parameter tuning

Break (10min)

4. Trees and ensemble methods (60min)

Day II – Deep Learning

1. Regression and classification with NN (20min)
2. Autoencoders and embedding (30min)

Break (10min)

3. Time series and sequential models (20min)
4. Case studies: word2vec, seq2seq (30min)

Break (10min)

5. Convolutional NN for image classification and segmentation (50min)

Break (10min)

6. Language models and transformers (60min)