Quantitative Methods and Machine Learning

Introduction

Mathias Müller

University of Zurich[UZH]

Institute of Computational Linguistics

Course Logistics

Course Organization

- Who we are
- Semester plan, milestones
- how we put together your grade
- Exercises
- Shared task
- Olat
- Requirements, be up to speed with
- Working together policy
- Who to ask if you need help

University of Zurich[UZH]

Institute of Computational Linguistics

Introduction to Machine Learning

What is Machine Learning?

The Real World vs. Models

Features

- Feature extraction
- Feature engineering
- Discriminative features

Intuition about Geometric Spaces

- Visualizing data points
- Geometric distance
- High-dimensional spaces

Data, Samples, Labels

- Data selection / collection,
- annotation, IAA, gold standard
- Training data, development / validation , testing splits
- Sample representativeness
- Bias
- Outliers
- Distributions, means, variance

Evaluation

- Baseline
- Hypothesis (- Testing)
- Improvements due to chance?
- N-fold cross-validation
- Performance metrics
- Precision, recall, f-measure (recap!)
- Confusion matrices

Supervised vs. unsupervised learning

- Classification/regression,
- Clustering, association
- Semi-supervision

Classification vs. regression

Example NLP Classification Problem

- Sentiment classification
- https://www.kaggle.com/c/word2vec-nlp-tutorial/data would be an option