Foundations of Machine Learning

Course overview

Learning objectives:

- Build predictive models from training data
- Correctly evaluate predictive models
- Analyze and compare the performance of different models
- Reason about the mathematical foundations of data mining techniques
- Recognize when a predictive model is under/overfitting
- Combine the above with dimension-reduction techniques
- Visualize and explore data using embeddings and clustering

Lecturers:

- Joaquin Vanschoren (j.vanschoren@tue.nl) MF 7.104a
- Vlado Menkovski (v.menkovski@tue.nl) MF 7.097b
- Anne Driemel (a.driemel@tue.nl) MF 7.073

Contact hours:

- Mondays, 10:45 - 12:30: Plenary Lectures (Flux 1.02)
- Thursdays, 13:45 - 15:30: Tutorials and Feedback (Flux 1.02)
  - These are NOT labs
- Thursdays, 15:45 - 17:30: Plenary Lectures (Flux 1.02)

Course materials:

- See Canvas for:
  - Syllabus
  - Announcements, discussions
  - Assignments, grades
- Lecture materials (Notebooks+PDFs) on GitHub
  - https://github.com/joaquinvanschoren/ML-course
  - The README contains pointers to relevant books

Evaluation:

- No exam, only assignments (4 team assignments, 1 individual)
- Preliminary(!) overview:
  - 1: Linear Models, Model selection, Ensembles (15 points)
    - Released Feb 8 , Deadline Mar 1
  - 2: Kernel methods and Bayesian Inference (15 points)
    - Released Mar 1, Deadline Mar 15
  - 3: Dimensionality reduction, Embeddings (15 points)
    - Released Mar 15, Deadline Mar 29
  - 4: Deep learning (15 points)

        * Released Mar 29, Deadline Apr 12
    - Individual assignment: Data challenge (30 points)
        * Released Mar 22, Deadline Apr 20

- Team assignments in teams of 2 students
    - Free choice, form groups on Canvas before Feb 9th!
        * Can be the same for all 4 team assignments
        * You can use Canvas Discussions or Chat to find teammates
    - You're allowed to find a new teammate (yourselves) after each assignment
- Passing grade: 6/10 over all assignments
    - 2/3 team assignments, 1/3 individual assignment

Course contents

May still change!

Contents: Week 1

- Machine learning concepts
- Build first models with `scikit-learn`
- k-Nearest Neighbors
- Tutorials:
    - Linear Algebra
    - Data Analysis with Python
- Linear Models
    - Linear regression, ridge regression, lasso
    - Logistic regression, linear SVMs

Contents: Week 2 (after spring break)

- Model evaluation and selection
    - Overfitting, cross-validation, ROC analysis, Bias-Variance analysis
    - Hyperparameter optimization
- Tutorials:
    - Data Analysis with Python (continued)
    - Feature engineering
- Ensemble learning
    - Decision trees
    - Bagging, (Gradient) Boosting, Stacking

Contents: Week 3

- Kernel methods
  - Support Vector Machines, Kernelization
- Tutorials:
  - Feature engineering (continued)
- Bayesian Learning
  - Naive Bayes
  - Gaussian processes

Contents: Week 4

- Constructing pipelines
  - Preprocessing, feature engineering, learning
  - Building machine learning systems
- Tutorials:
  - Analysing images
- Dimensionality reduction 1
  - PCA, MDS, Isomap

Contents: Week 5

- Dimensionality reduction 2
  - Random projections, Locality-sensitive hashing
- Locality-sensitive hashing
  - Jaccard Similarity, MinHashing

Contents: Week 6

- Clustering
  - Lloyd's algorithm, kMeans++, Gonzales' algorithm
- Introduction to Deep Learning
  - Artificial neurons, gradient descent, backprop

Contents: Week 7

- Multilayer Perceptron
- Convolutional Neural Networks

Contents: Week 8

- Recurrent Neural Networks
- Q&A