

Introduction to Data analysis and Machine Learning in Physics

Jörg Marks

Studierendentage, 8-12 April 2024
9:00 - 12:00

Outline

- **Day 1**
 - ▶ Introduction, software and data fitting
- **Day 2**
 - ▶ Machine learning - basics
- **Day 3**
 - ▶ Machine learning - multivariate analysis
- **Day 4**
 - ▶ Machine learning - decision trees
- **Day 5**
 - ▶ Machine learning - convolutional neural networks

Course Information (1)

- **Organization** and **Objective**
 - ▶ 1 ETC: Compulsory attendance is required
Active participation in the exercises
 - ▶ Course in a tutorial style
 - ▶ Obtain basic knowledge for problem-oriented self-studies

Course Information (2)

- Course requirements
 - ▶ Python knowledge needed / good C++ knowledge might work
 - ▶ Userid to use the CIP Pool of the faculty of physics
- Course structure
 - ▶ Course in INF 226, Glasbox 2.108 without terminal access but using the CIP server via `ssh`
 - ▶ You need your own laptop to participate and access the `jupyter3 hub` of the faculty of physics
 - ▶ Lectures are interleaved with tutorial/exercise sessions in small groups (up to 4 persons / group)

- Course homepage which includes and distributes all material

https://www.physi.uni-heidelberg.de/~marks/ml_einfuehrung/

/Folien

Transparencies of the lectures

/Beispiele

iPython and python3 files shown in the lectures

/Aufgaben

Exercises to be solved during the course

/Loesungen

Solutions of the exercises

Course Information (3)

TensorFlow and Keras are now also installed in the CIP jupyter hub. In addition, with a google account you can run jupyter notebooks on Google Colab:

<https://colab.research.google.com/>

Missing python libraries can be included by adding the following to a cell (here for the pypng library):

```
!pip install pypng
```

Course Information (4)

- Your installation at home:
 - ▶ Web Browser to access jupyter3
 - ▶ Access to the CIP pool via an ssh client on your home PC
- No requirements for a special operating system
- Software:
 - ▶ firefox or similar
 - ▶ Cisco AnyConnect
 - ▶ ssh client (MobaXterm on Windows, integrated in Linux/Mac)
- Local execution of python / iPython
 - ▶ Install **anaconda3** and download / run the iPython notebooks (also python scripts are available)
https://www.physi.uni-heidelberg.de/~marks/ml_einfuehrung/Folien/anaconda3SetupHints.pdf
- Hints for software installations and CIP pool access
https://www.physi.uni-heidelberg.de/~marks/root_einfuehrung/Folien/CIPpoolAccess.PDF

Course Information (5)

Here are the relevant instruction for macOS using pip:

Assumptions: homebrew is installed.

Install python3 (see <https://docs.python-guide.org/starting/install3/osx/>)

```
$ brew install python
$ python --version
Python 3.8.5
```

Make sure pip3 is up-to-date (alternative: conda → don't mix conda and pip installations)

```
$ pip3 install --upgrade pip
```

Install modules needed:

```
$ pip3 install --upgrade jupyter matplotlib numpy pandas
scipy scikit-learn xgboost iminuit tensorflow tensorflow_datasets Keras
```

Topics and file name conventions

0. Introduction (this file) (introduction.pdf)
1. Introduction to Python (01_intro_python_*)
2. Data Modeling and Fitting (02_fit_intro_*)
3. Machine Learning Basics (03_ml_intro_*)
4. Decisions Trees (04_decision_trees_*)
5. Convolutional Neural Networks (05_neural_networks_*)

Programm Day 1

- Technicalities
- Summary of NumPy
- Plotting with matplotlib
- Input / output of data
- Summary of pandas
- Fitting with iminuit and PyROOT
- Transparencies with activated links, examples and exercises
 - ▶ Software: [01_intro_python.pdf](#)
 - ▶ Fitting: [02_fit_intro.pdf](#)

Programm Day 2

- Introduction to machine learning [03_ml_intro.pdf](#)
 - ▶ Tensorflow / Keras, datasets
 - ▶ Supervised learning
 - ▶ Classification

Programm Day 3

- Multivariate analysis [03_ml_intro_mva.pdf](#)
 - ▶ Regression
 - ▶ Linear regression
 - ▶ Logistic regression
 - ▶ Softmax regression (multi-class classification)

Programm Day 4

- Decision trees [04_decision_trees.pdf](#)
 - ▶ Bagging and boosting
 - ▶ Random forest
 - ▶ XGBoost

Programm Day 5

- Convolutional neural networks [05_neural_networks_cnn.pdf](#)