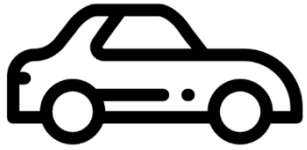
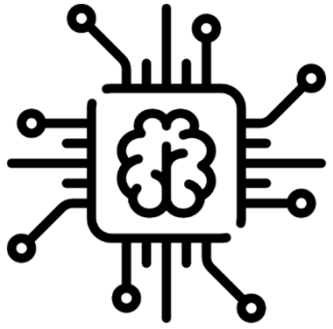
Supervised Learning: Regression

Johannes Betz / Prof. Dr. Markus Lienkamp / Prof. Dr. Boris Lohmann
(Alexander Wischnewski, M. Sc.)

Agenda

1. Excercise
    1. Task: Tools
    2. Task: Comparison of regression algorithms
    3. Task: Real world data analysis





Tools - Scikit – Learn – Example for a Python ML Library

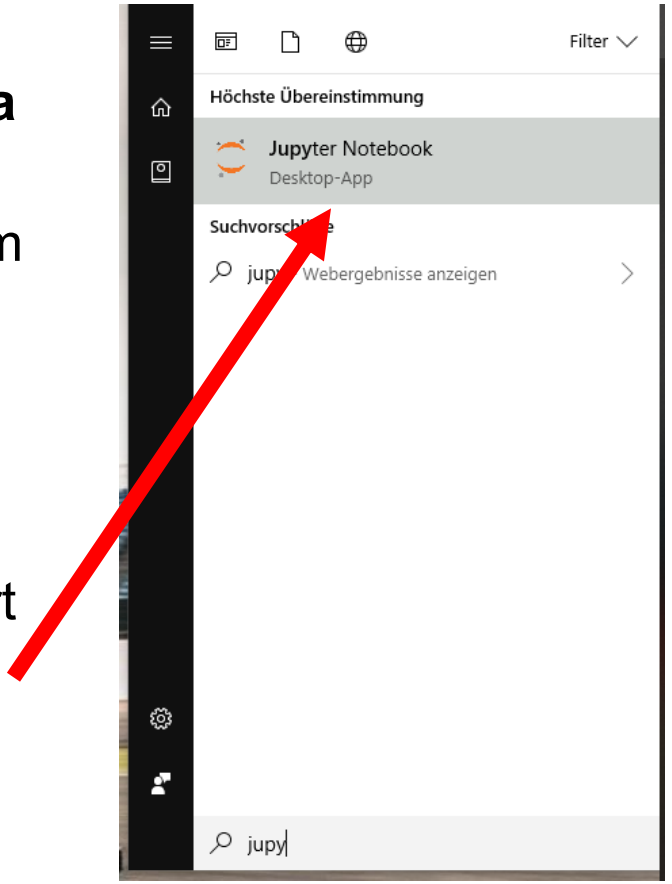
- Actively developed machine learning library for python
- Extension to scipy
- Widely used and well documented algorithms for
  - Classification
  - Regression
  - Clustering
  - Model selection
  - Data preprocessing
- http://scikit-learn.org/stable/documentation.html

Tools - Jupyter Notebook

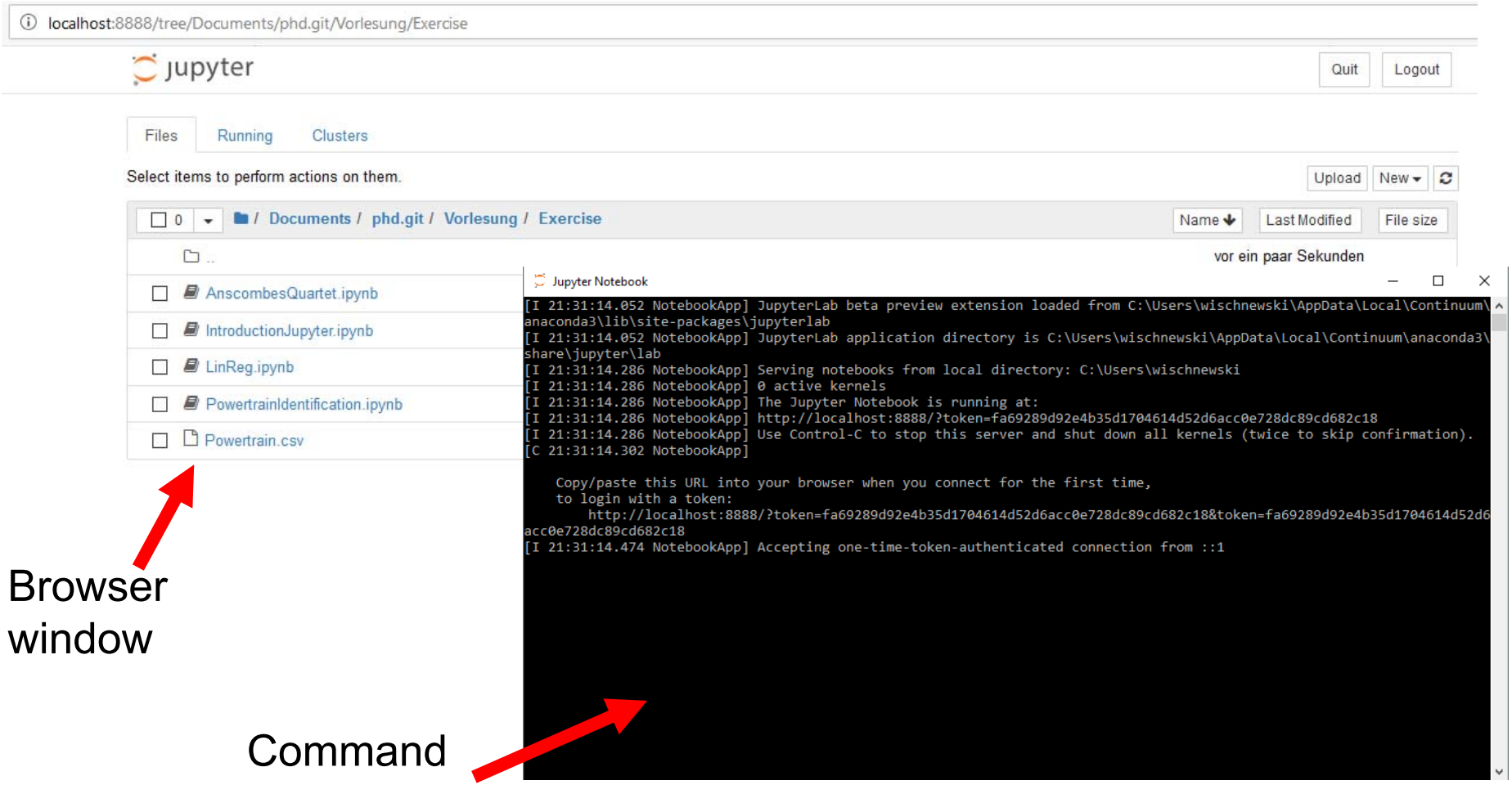
- Lightweight tool to develop python code
- Creates so called ‚Notebooks', which store
  - Code
  - Results
  - Plots
  - Documentation in Markdown
- Ideal for small to medium sized projects, especially student thesis
- Can be downloaded and installed via the tutorial:
  - http://jupyter.org/install
- Docs are available here: https://jupyter-notebook.readthedocs.io/en/stable/
- **There will be example code for the homework available via jupyter notebooks at moodle!**

Tools - Use the jupyter notebooks

- Install Jupyter according to the tutorial http://jupyter.org/install using the **Anaconda Distribution for python3**!
- Download the Notebooks and Datasets from Moodle
- Create a new folder within your *Documents* folder
- Copy everything into this folder
- Start Jupyter Notebook by pressing the start button, type Jupyter and execute *Jupyter Notebook*
- This opens a **command prompt** and a **browser window**. Both have to stay open during your work!



Tools - Use the jupyter notebooks

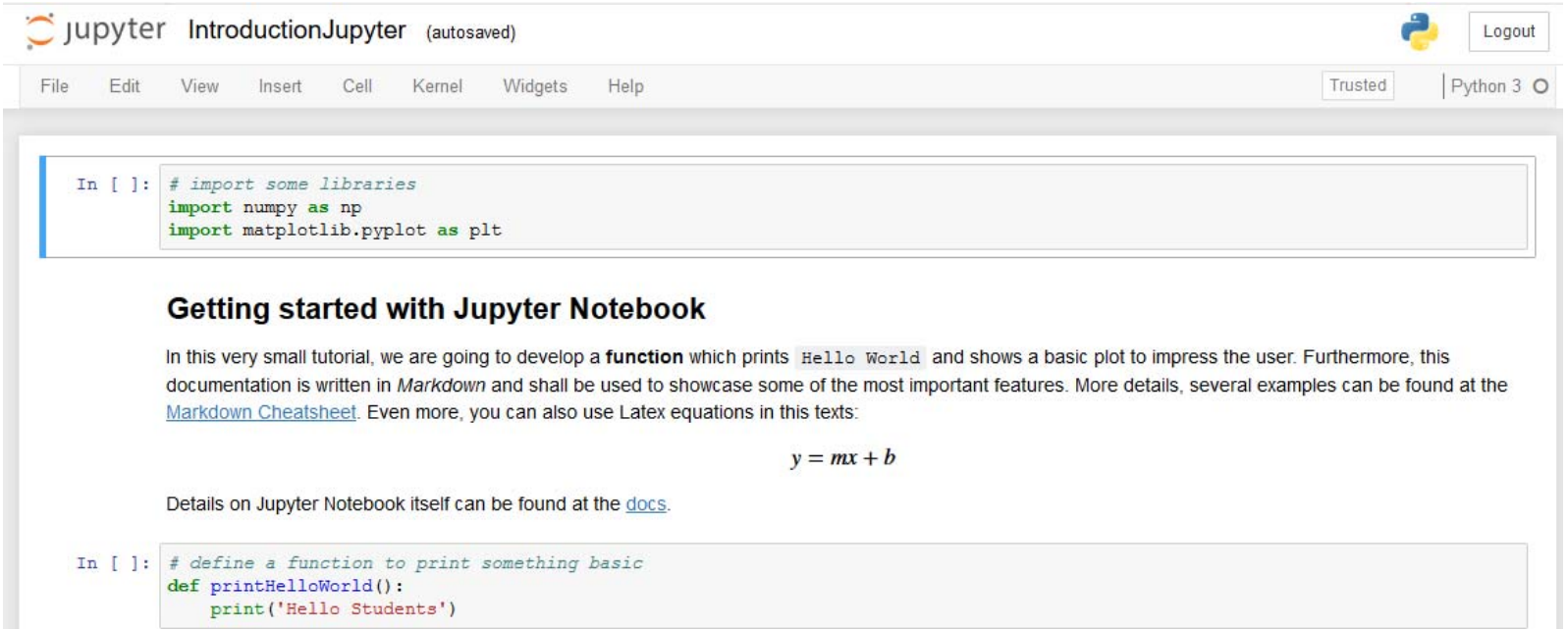


Browser window

Command line prompt

Tools - Use the jupyter notebooks

- Open one of the notebooks by clicking on the name
- Run each cell seperatly by clicking into it and press *Strg+Enter* (Take care of the order! Things needed later have to be run first)
- Or run the whole notebook by clicking Kernel → Restart & Run all → Restart and Run all Cells



Excercises

- Comparison of regression algorithms
- Real world data analysis