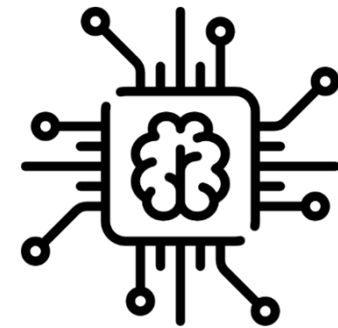


## Supervised Learning: Regression

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

### Agenda

1. Exercise
  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>



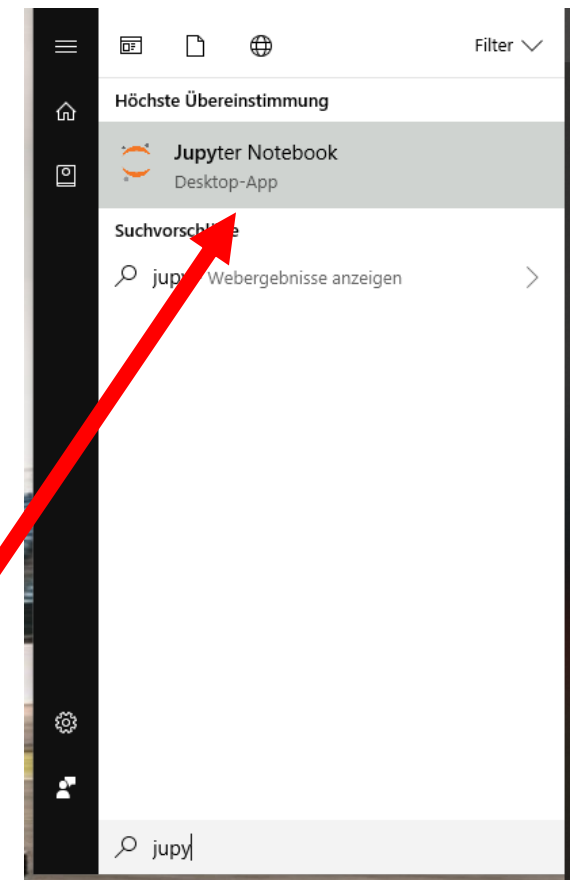
## Tools - Jupyter Notebook

- 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!**

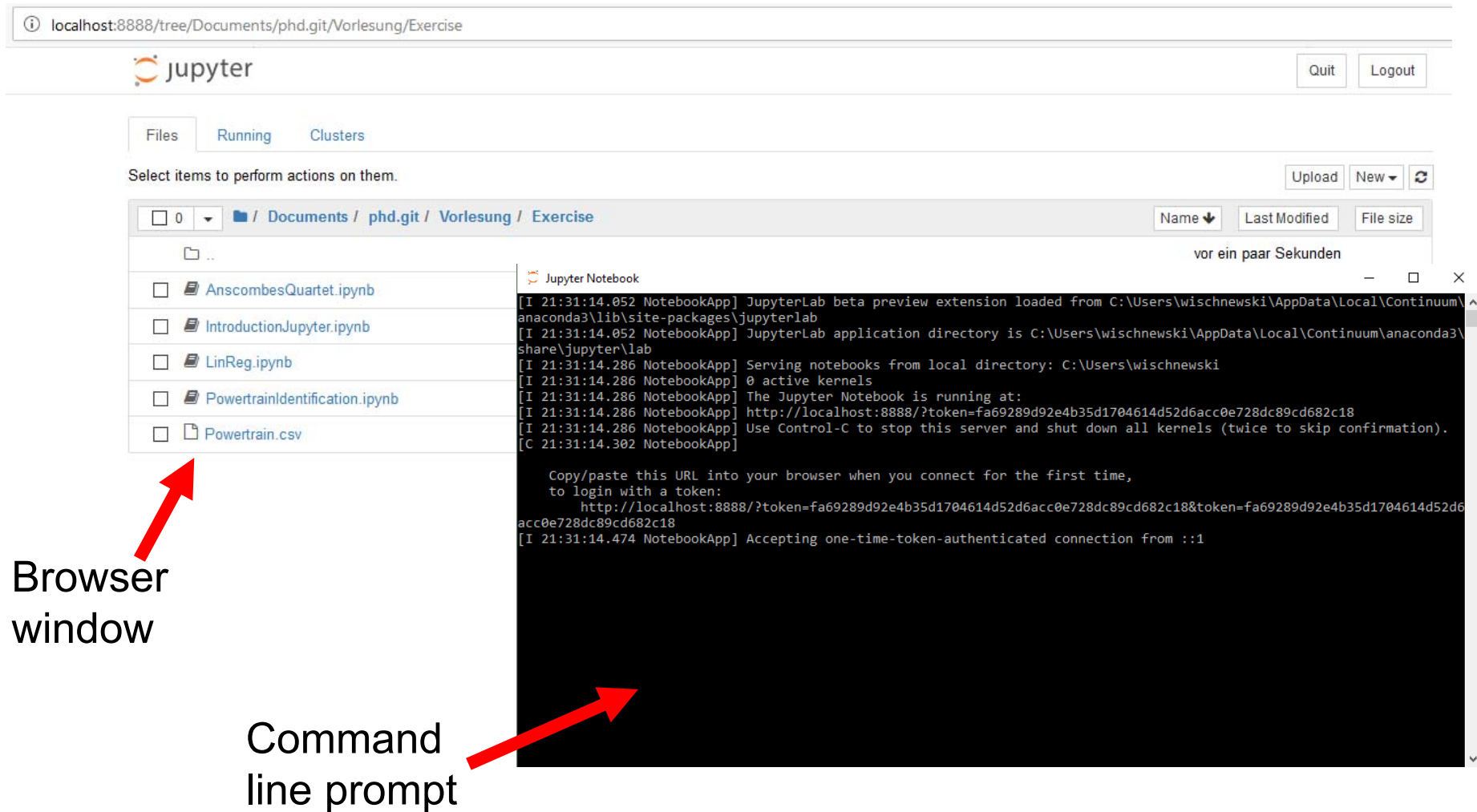


## Tools - Use the jupyter notebooks

- 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



localhost:8888/tree/Documents/phd.git/Vorlesung/Exercise

jupyter

Quit Logout

Files Running Clusters

Select items to perform actions on them.

Upload New

0 / Documents / phd.git / Vorlesung / Exercise

Name Last Modified File size

AnscombesQuartet.ipynb

IntroductionJupyter.ipynb

LinReg.ipynb

PowertrainIdentification.ipynb

Powertrain.csv

Jupyter Notebook

```
[I 21:31:14.052 NotebookApp] JupyterLab beta preview extension loaded from C:\Users\wischnewski\AppData\Local\Continuum\
anaconda3\lib\site-packages\jupyterlab
[I 21:31:14.052 NotebookApp] JupyterLab application directory is C:\Users\wischnewski\AppData\Local\Continuum\anaconda3\
share\jupyter\lab
[I 21:31:14.286 NotebookApp] Serving notebooks from local directory: C:\Users\wischnewski
[I 21:31:14.286 NotebookApp] 0 active kernels
[I 21:31:14.286 NotebookApp] The Jupyter Notebook is running at:
[I 21:31:14.286 NotebookApp] http://localhost:8888/?token=fa69289d92e4b35d1704614d52d6acc0e728dc89cd682c18
[I 21:31:14.286 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 21:31:14.302 NotebookApp]

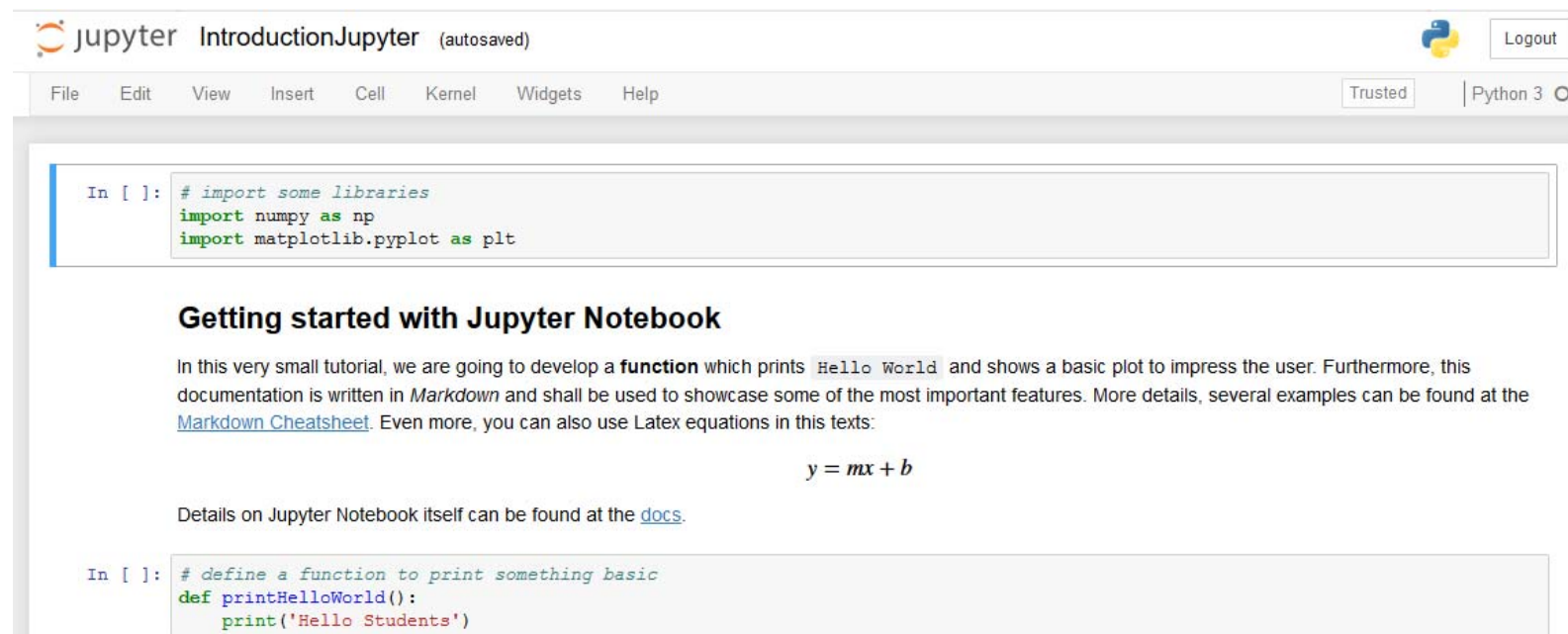
Copy/paste this URL into your browser when you connect for the first time,
to login with a token:
http://localhost:8888/?token=fa69289d92e4b35d1704614d52d6acc0e728dc89cd682c18&token=fa69289d92e4b35d1704614d52d6
acc0e728dc89cd682c18
[I 21:31:14.474 NotebookApp] Accepting one-time-token-authenticated connection from ::1
```

Browser window

Command line prompt

## Tools - Use the jupyter notebooks

- Open one of the notebooks by clicking on the name
- Run each cell separately 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



The screenshot shows a Jupyter Notebook interface. At the top, there's a header bar with the Jupyter logo, the text "IntroductionJupyter (autosaved)", a Python logo, and a "Logout" button. Below this is a menu bar with "File", "Edit", "View", "Insert", "Cell", "Kernel", "Widgets", and "Help". On the right of the menu bar, there's a "Trusted" status indicator and a "Python 3" environment selector. The main area contains two cells. The first cell is a code cell with the following Python code:

```
In [ ]: # import some libraries
import numpy as np
import matplotlib.pyplot as plt
```

The second cell is a markdown cell with the following content:

### Getting started with Jupyter Notebook

In this very small tutorial, we are going to develop a **function** which prints `Hello World` and shows a basic plot to impress the user. Furthermore, this documentation is written in *Markdown* and shall be used to showcase some of the most important features. More details, several examples can be found at the [Markdown Cheatsheet](#). Even more, you can also use Latex equations in this texts:

$$y = mx + b$$

Details on Jupyter Notebook itself can be found at the [docs](#).

The third cell is another code cell with the following Python code:

```
In [ ]: # define a function to print something basic
def printHelloWorld():
    print('Hello Students')
```

## Excercises

- Comparison of regression algorithms
- Real world data analysis