

Machine Learning for Health Informatics: Assignments

Marcus Bloice
Tutor

Email: marcus.bloice@medunigraz.at

GitHub: <http://bit.do/MLHI>

<https://github.com/mdbloice/Machine-Learning-for-Health-Informatics>

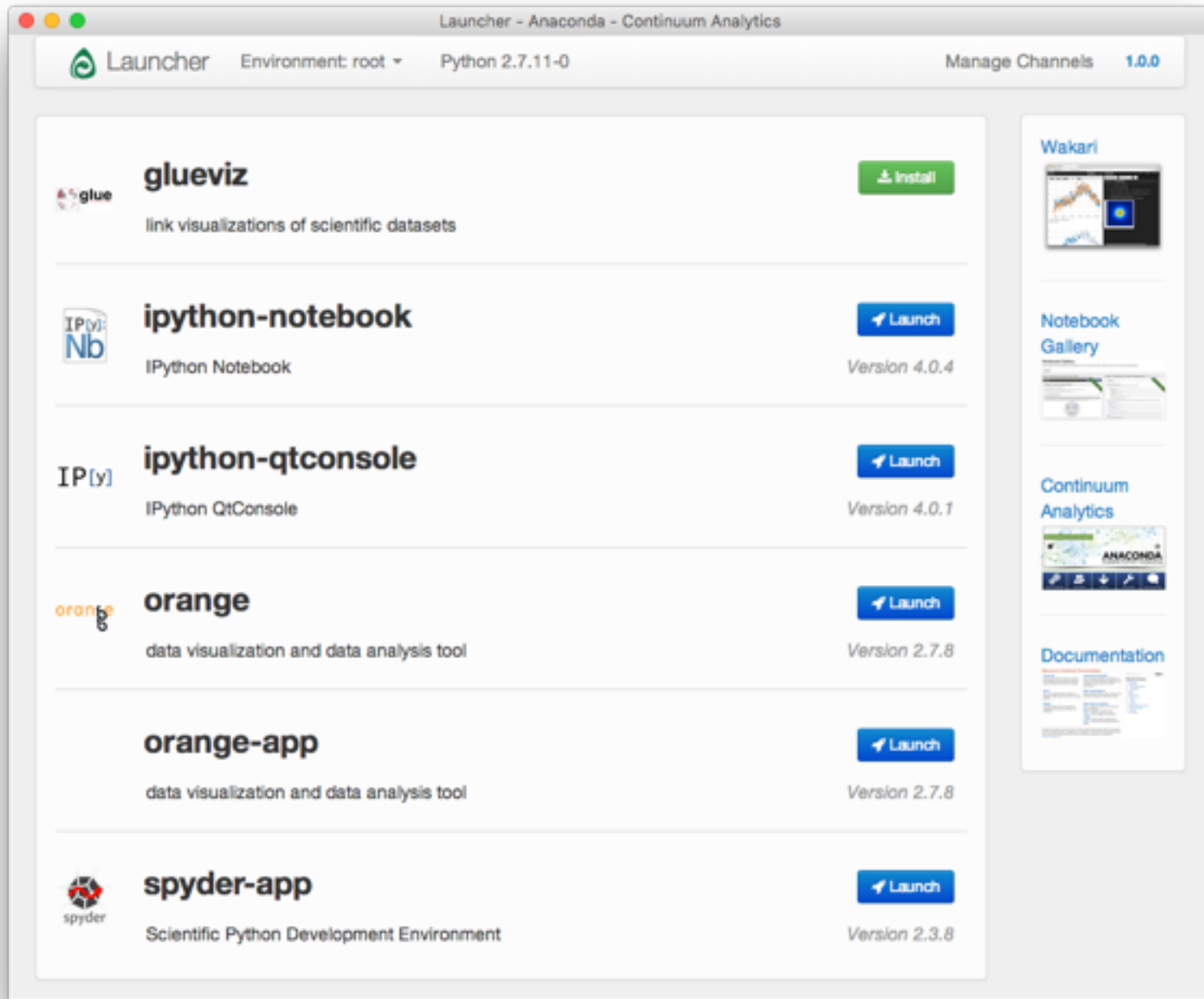


- Assignments for Übung
- **Three programming assignments** in the semester
- Assignments will be written in **Python**
- They will take the form of **Jupyter Notebooks**
- The assignments will require you to **apply algorithms from the lecture to health data**
- First assignment has **no deliverable**:
 - **Get all requirements installed!**
- Next assignment in ≈ 1 month

- What is a Jupyter Notebook?
- Jupyter is a **browser-based IDE**, projects take the form of **interactive notebooks**: Demo later
- It is easy to share your notebooks with others
- During the semester I will make available a few **semi-complete notebooks on GitHub**:
 - <http://bit.do/MLHI>
- Your task would be to **finish the notebooks**
- Send the notebooks to me **by email**

- You will need an installation of
 - Python
 - Jupyter
- And you will need the following libraries:
 - NumPy (linear algebra)
 - SciKit-Learn (Machine Learning framework)
 - Matplotlib (plotting library)
 - SciPy, Pandas, maybe one or two more
- However: the easiest way is to install the **Anaconda scientific Python distribution**

- "Anaconda is a completely free Python distribution. It includes more than 400 of the most popular Python packages for science, math, engineering, and data analysis."
- It also includes **an IDE called Spyder** and several other utilities such as **Jupyter**
- Everything is **configured and ready to use**
- Get it here:
 - <https://www.continuum.io/downloads>
- **Use the Python 2.7 version!!!**



The screenshot shows the Anaconda Launcher window titled "Launcher - Anaconda - Continuum Analytics". The interface includes a header with the Anaconda logo, "Environment: root", "Python 2.7.11-0", and "Manage Channels 1.0.0". The main content area lists several applications:

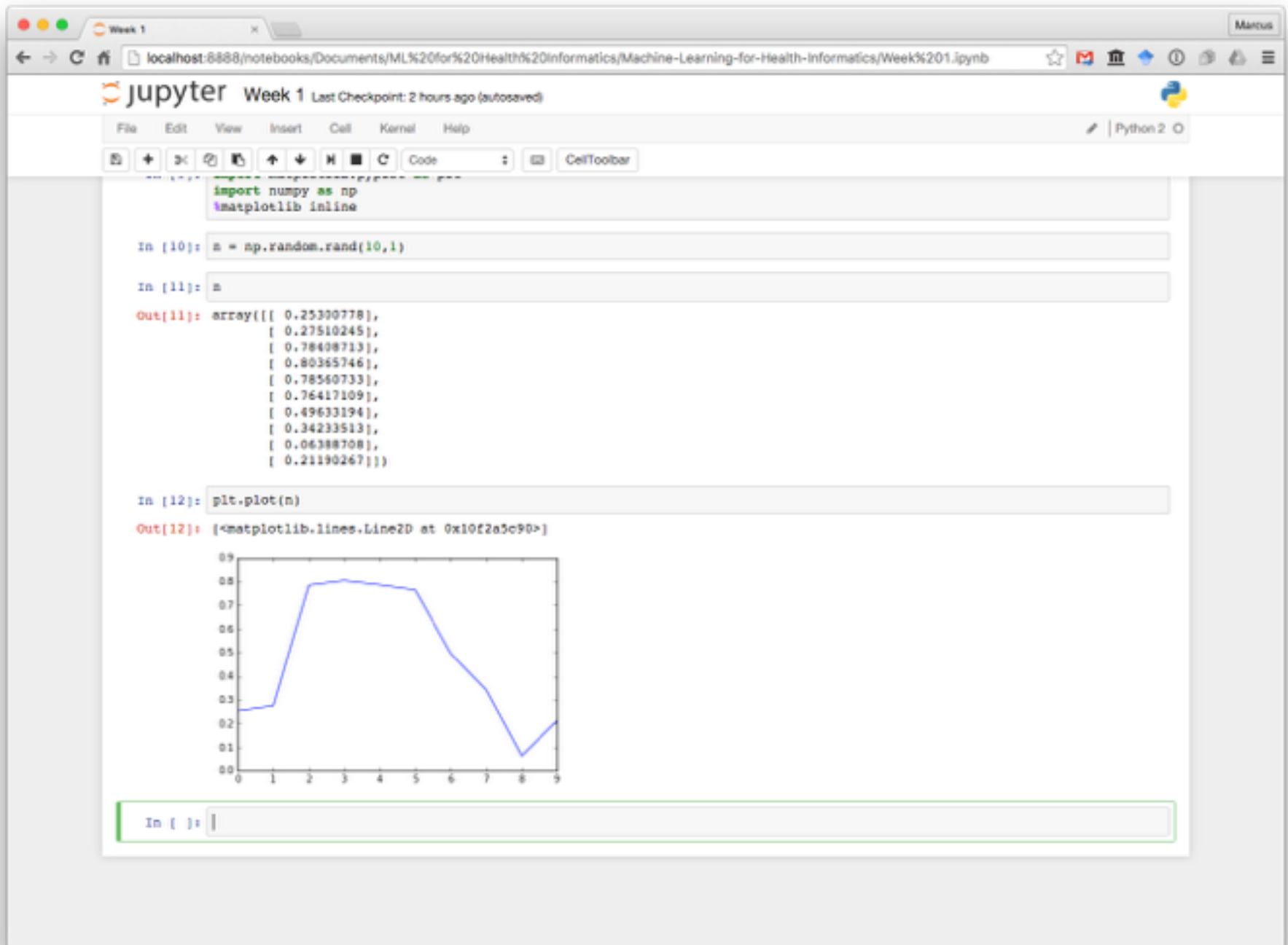
- glueviz**: link visualizations of scientific datasets. Action: [Install](#)
- ipython-notebook**: IPython Notebook. Version 4.0.4. Action: [Launch](#)
- ipython-qtconsole**: IPython QtConsole. Version 4.0.1. Action: [Launch](#)
- orange**: data visualization and data analysis tool. Version 2.7.8. Action: [Launch](#)
- orange-app**: data visualization and data analysis tool. Version 2.7.8. Action: [Launch](#)
- spyder-app**: Scientific Python Development Environment. Version 2.3.8. Action: [Launch](#)

The right sidebar contains links to "Wakari", "Notebook Gallery", "Continuum Analytics", and "Documentation".

The screenshot displays the Spyder IDE environment with the following components:

- Editor:** A Python script named `Segment.py` is open. It contains code for image processing using `skimage` and `matplotlib`. The script defines an image path, opens it, and performs various operations like histogramming and edge detection.
- Python console:** Shows the output of the script execution. It includes the Python version (2.7.11), Anaconda version (2.3.0), and a `TypeError` message: `TypeError: arange: scalar arguments expected instead of a tuple.` This error occurs at line 1 of the `import numpy as np` statement.
- Variable explorer:** A window showing the current state of variables. It contains a table with the following data:

Name	Type	Size	Value
n	int64	(0,)	array([], dtype=int64)
- History log:** A window showing the execution history of the script, including the `import numpy as np` statement and the `np.arange` function calls.



The screenshot shows a Jupyter Notebook interface in a web browser. The browser address bar shows the URL: `localhost:8888/notebooks/Documents/ML%20for%20Health%20Informatics/Machine-Learning-for-Health-Informatics/Week%201.ipynb`. The notebook title is "jupyter Week 1" with a subtext "Last Checkpoint: 2 hours ago (autosaved)". The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Help) and a toolbar with icons for file operations and code execution. The code cell contains the following Python code:

```
import numpy as np
import matplotlib inline

In [10]: n = np.random.rand(10,1)

In [11]: n
Out[11]: array([[ 0.25300778],
 [ 0.27510245],
 [ 0.78408713],
 [ 0.80365746],
 [ 0.78560733],
 [ 0.76417109],
 [ 0.49633194],
 [ 0.34233513],
 [ 0.06388708],
 [ 0.21190267]])

In [12]: plt.plot(n)
Out[12]: [<matplotlib.lines.Line2D at 0x10f2a5c90>]
```

The output of the plot shows a line graph with the x-axis ranging from 0 to 9 and the y-axis ranging from 0.0 to 0.9. The plot displays a blue line representing the values of the array `n`. The line starts at approximately 0.25, rises to a peak of about 0.8 between x=2 and x=5, and then falls to a minimum of about 0.07 at x=8, ending at approximately 0.2 at x=9.

At the bottom of the notebook, there is an empty input cell with the prompt `In []:`.

1. Install Anaconda

2. Get the `Week1.ipynb` file from the GitHub repository: <http://bit.do/MLHI>

3. Make sure **each cell can run without errors**

- There is **NO deliverable** for this assignment!
- If all runs well, then you will be **ready to start machine learning using Python!**
- If a package **does not exist**, use `conda` to install it: for example seaborn:
`conda install seaborn`

Jupyter Demo

GitHub: <http://bit.do/MLHI>

