

# Programming Setup

Joaquin Vanschoren, Mykola Pechenizkiy, Anne Driemel

## 1 Using R and Python

R and Python are both great ways to quickly learn how to do machine learning in practice. They can be learned rather quickly and enable you to do almost anything you want. Mastering them is increasingly crucial for a career in data science.

### 1.1 Jupyter Notebooks

- We will use Jupyter Notebooks in some of the course lectures
- Literate programming: Explanations + executable code examples
- Contain the full code of what we show in class, so you can try it at home and do your own experiments

These lectures are made available in 3 formats:

- HTML: for online viewing (navigation: spacebar, arrows, ESC)
- Notebook: download the .ipynb files and open them with Jupyter
- PDF: for offline reference

### 1.2 Learning R

There exist many ways to learn R quickly:

- DataCamp (beginner to advanced): <https://www.datacamp.com>
- Codeschool (beginner, step by step): <http://tryr.codeschool.com>
- Data Carpentry (intermediate): <http://www.datacarpentry.org/R-ecology>
- R Studio guide (installation, environments): <https://www.rstudio.com/resources/training/online-learning/#R>

### 1.3 Learning Python

Likewise, there are many ways to quickly learn Python:

- DataCamp (beginner to advanced): <https://www.datacamp.com>
- CodeCademy (beginner, step by step): <https://www.codecademy.com/learn/python>
- Google's Python Class (intermediate): <https://developers.google.com/edu/python>
- Data Carpentry (intermediate): <http://www.datacarpentry.org/python-ecology>
- Many more...

### 1.4 Setting up Python and R environments

#### 1.4.1 Install Anaconda (Python+Jupyter+more)

- [Install Anaconda](#), choose your OS and the latest Python 3
- Note: This will install a separate Python environment
  - Anaconda puts a line in your .bash\_profile to use this version by default

You should now be able to:

- Start a Python shell by typing `python` in your command shell
- Start a Jupyter notebook by typing `jupyter notebook`

### 1.4.2 Installing Python extensions

Whenever you need a package that is not already installed:

- Try conda: `conda install <package>`
- Try pip: `pip install <package>`
- Necessary: `conda install numpy scipy scikit-learn jupyter matplotlib`

Some packages need to be installed from GitHub source code:

- Create a folder (e.g. folder 'github' inside the anaconda dir)
- Then clone the package and install it:

```
git clone <git URL>
cd <new directory>
python setup.py install
```

- OpenML (for ML resources): <https://github.com/openml/python.git>
- RISE (for slideshows): <https://github.com/damianavila/RISE.git>

### 1.4.3 Jupyter Extensions

To export PDFs:

- [Install pandoc](#)

### 1.4.4 Installing R

- You can install the Anaconda-supported R (not always the latest version), which includes the Jupyter R kernel:

```
conda install -c r r
conda install -c r r-essentials
```

- OSX users: there seems to be a transient bug in Anaconda's Jupyter R kernel. If you experience issues, delete and reinstall anaconda (without R) and install R yourself.
- You can also [install R yourself](#). If so, also install the Jupyter R kernel:

```
install.packages('devtools') # If necessary
library(devtools)
install_github('IRkernel/repr')
install_github('IRkernel/IRdisplay')
install_github('IRkernel/IRkernel')
IRkernel::installspec()
```

- You can now start the R shell by typing `R` on the command line (or use RStudio)

### 1.4.5 Installing R extensions

Whenever you need a package that is not already installed:

- start your R shell in admin mode, e.g. `sudo R`
- `install.packages(<package name>)`. Case sensitive!

Some packages need to be installed from GitHub source code:

```
install.packages('devtools') # If necessary
library(devtools)
install_github('<package name>')
```

E.g. for OpenML:

```
install_github('openml/r')
```