

# DATA MINING

---

Python

# Python

- In the last few years there is an increasing community that creates **Data Mining tools in Python**
  - There are also tools in other languages but we will use Python whenever we can for a common point of reference.
- We will use **Iron Python** that interfaces with .NET and we can run Notebooks in a browser.
  - You can also use any editor and compile and run from a terminal

# Installing Python

- Installing libraries in Python is complex, so you should download the **Anaconda Scientific Python** distribution which will install most of the libraries that we will use.
  - There are two versions, Python 2.7 and Python 3.0 and they are not compatible. We will use Python 3.0

# Resources

- There are tons of resources online for Python.
- For an introduction you can also look at the slides of the [Introduction to Programming](#) course by prof. N. Mamoulis

# The Anaconda Navigator

The screenshot displays the Anaconda Navigator desktop application. The window title is "Anaconda Navigator" and it includes a menu bar with "File" and "Help". The main header features the "ANACONDA NAVIGATOR" logo and a "Sign in to Anaconda Cloud" button. A left sidebar contains navigation options: Home, Environments, Projects (beta), Learning, and Community. The main content area, titled "Applications on root", shows a grid of application cards. The "jupyter notebook" card is highlighted with a red border. Below the grid, there are links for "Documentation", "Developer Blog", and "Feedback", along with social media icons for Twitter, YouTube, and GitHub.

Applications on  Channels Refresh

Application	Version	Action
jupyterlab	0.27.0	Launch
<b>jupyter notebook</b>	<b>5.0.0</b>	<b>Launch</b>
qtconsole	4.3.1	Launch
spyder	3.2.3	Launch
glueviz	0.10.4	Install
orange3	3.4.1	Install
rstudio	1.0.153	Install

Documentation  
Developer Blog  
Feedback

Twitter YouTube GitHub

Home x

localhost:8888/tree

# jupyter

Files Running Clusters

Select items to perform actions on them. Upload New ▾ ↻

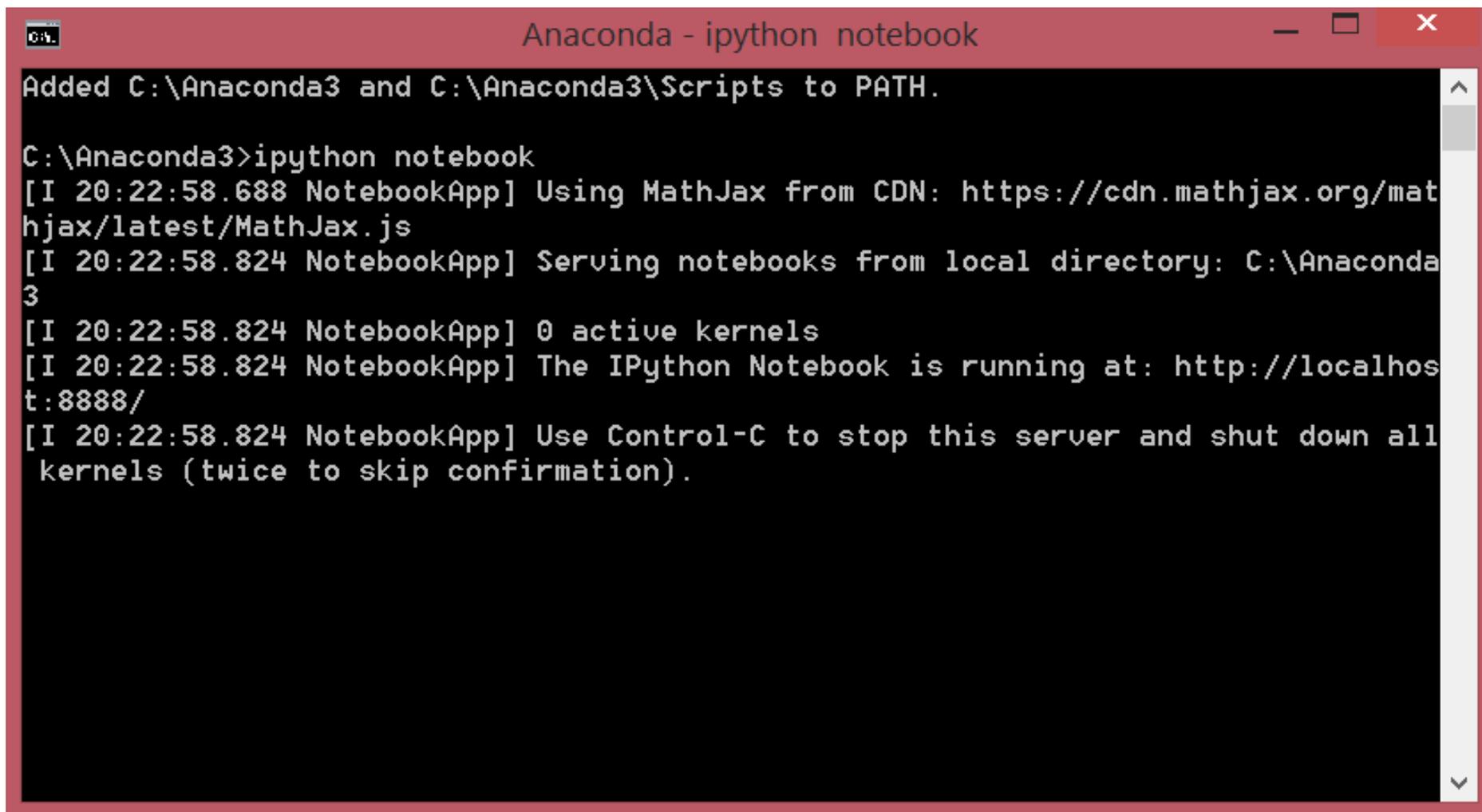
<input type="checkbox"/>	▼	🏠
<input type="checkbox"/>	📁	DLLs
<input type="checkbox"/>	📁	Doc
<input type="checkbox"/>	📁	Examples
<input type="checkbox"/>	📁	Lib
<input type="checkbox"/>	📁	Library
<input type="checkbox"/>	📁	Menu
<input type="checkbox"/>	📁	Scripts
<input type="checkbox"/>	📁	Tools
<input type="checkbox"/>	📁	conda-meta
<input type="checkbox"/>	📁	envs
<input type="checkbox"/>	📁	include
<input type="checkbox"/>	📁	info
<input type="checkbox"/>	📁	libs
<input type="checkbox"/>	📁	licenses
<input type="checkbox"/>	📁	node-webkit
<input type="checkbox"/>	📁	pkgs
<input type="checkbox"/>	📁	share
<input type="checkbox"/>	📁	tcl

# The Anaconda Navigator

The screenshot displays the Anaconda Navigator desktop application. The window title is "Anaconda Navigator" and it includes a menu bar with "File" and "Help". The main header features the "ANACONDA NAVIGATOR" logo and a "Sign in to Anaconda Cloud" button. A left sidebar contains navigation options: Home, Environments, Projects (beta), Learning, and Community. The main content area, titled "Applications on root", shows a grid of application cards. The "qtconsole" card is highlighted with a red border. Each card includes an icon, the application name and version, a brief description, and a button to either "Launch" or "Install".

Application	Version	Description	Action
jupyterlab	0.27.0	An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.	Launch
jupyter notebook	5.0.0	Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.	Launch
qtconsole	4.3.1	PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.	Launch
spyder	3.2.3	Scientific PYTHON Development Environment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features	Launch
glueviz	0.10.4	Multidimensional data visualization across files. Explore relationships within and among related datasets.	Install
orange3	3.4.1	Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.	Install
rstudio	1.0.153	A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.	Install

# Starting iPython notebook



```
Added C:\Anaconda3 and C:\Anaconda3\Scripts to PATH.  
C:\Anaconda3>ipython notebook  
[I 20:22:58.688 NotebookApp] Using MathJax from CDN: https://cdn.mathjax.org/mathjax/latest/MathJax.js  
[I 20:22:58.824 NotebookApp] Serving notebooks from local directory: C:\Anaconda3  
[I 20:22:58.824 NotebookApp] 0 active kernels  
[I 20:22:58.824 NotebookApp] The IPython Notebook is running at: http://localhost:8888/  
[I 20:22:58.824 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
```

# Changing the notebook default directory

- From the Anaconda terminal type the command:
  - `jupyter notebook --generate-config`
- This will generate `.jupyter/jupyter_notebook_config.py` file under your home directory.
- Find, un-comment and modify the line `# c.NotebookApp.notebook_dir = ''` in the config file to point to the desired directory

# Installing Packages

- You can install packages from the Anaconda terminal using the command:
  - `conda install <name of package>`
- For example, [Seaborn](#) is a package for Statistical Data Visualization.
  - `conda install seaborn`
- [panda-datareader](#) is a package for loading online datasets.
  - `conda install panda_datareader`