

Programming Lab II

Lez 1

Introduzione, python extra e numpy

Laura Nenzi

Benvenuti

Corso: Laboratorio di programmazione II (3 CFU) 

→ Modulo del corso: INTRODUZIONE ALLA PROGRAMMAZIONE E
LABORATORIO (15 CFU) che comprende Modulo prof. Caravagna (9 CFU) e
Laboratorio di programmazione I (3 CFU)

Docente: Laura Nenzi (io )

Sito Web (repository): https://github.com/lauranenzi/ProgrammingLab_II

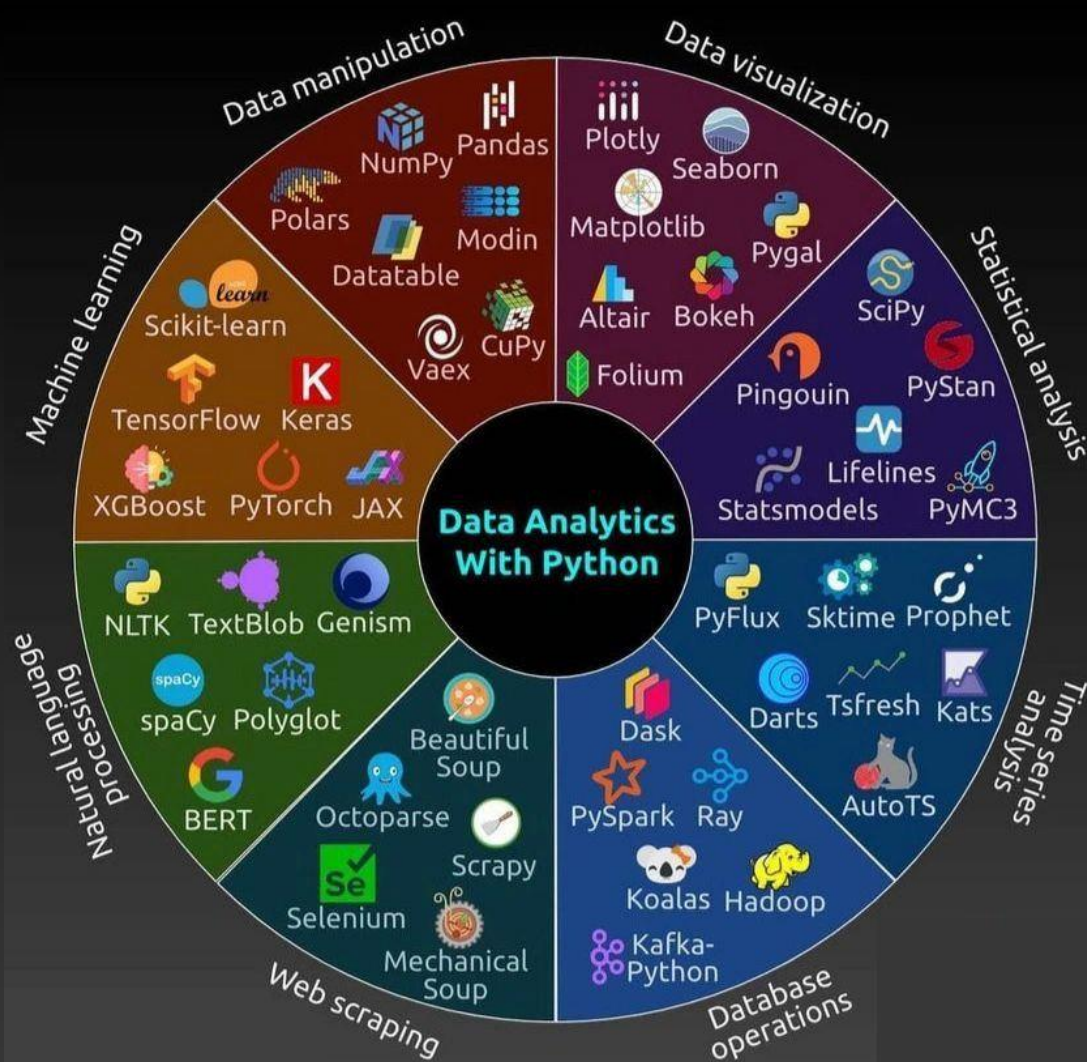
Ricevimento: libero, scrivetemi a lnenzi@units.it

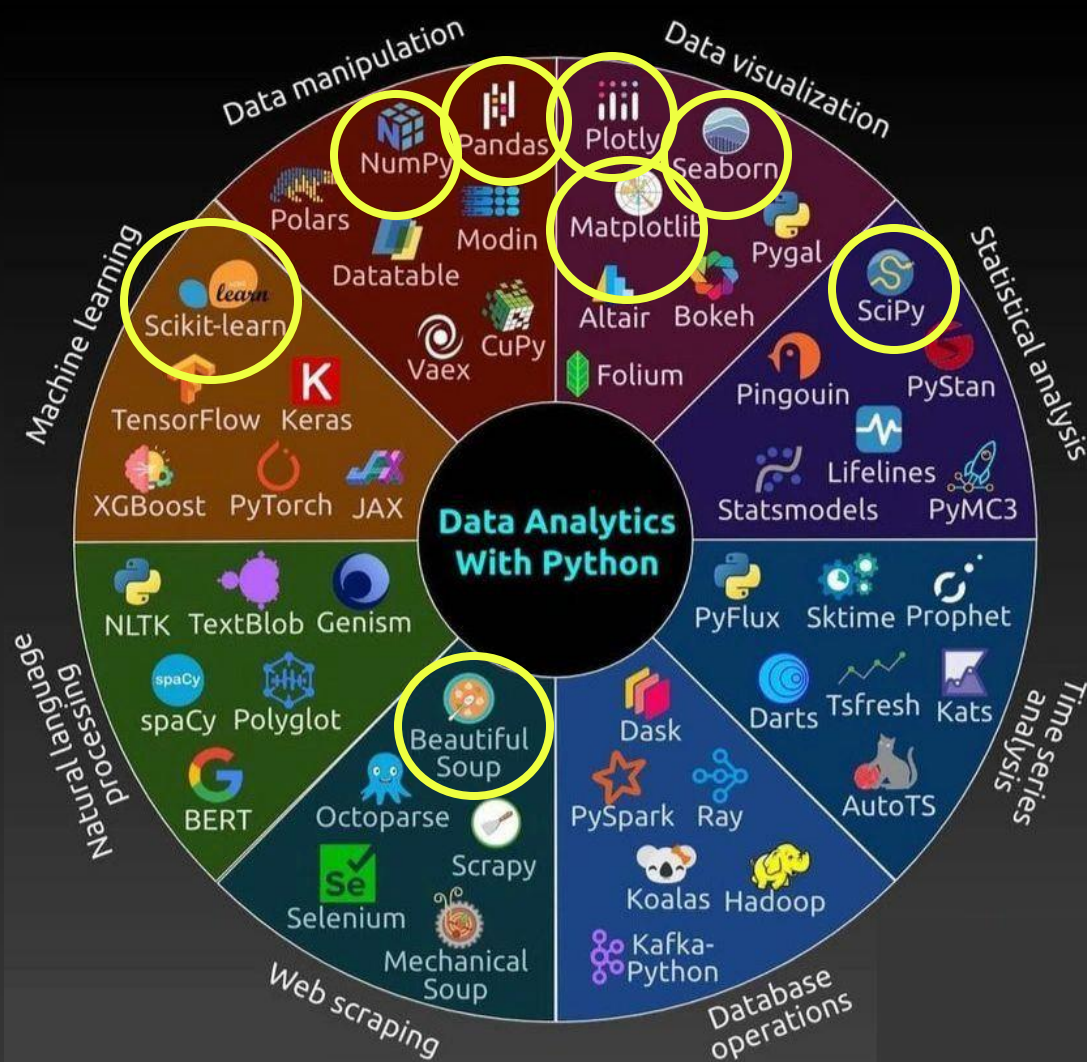
Tutors: Michel El Saliby, michel.elsaliby@phd.units.it

Argomenti del corso

- 1) Numpy per il calcolo numerico
- 2) Pandas per la struttura e manipolazione dei dati
- 3) Librerie per fare grafici: Matplotlib e Seaborn
- 4) SciPy per ottimizzazione e computazioni statistiche
- 5) Scikit-learn per la costruzione di modelli lineari
- 6) Plotly: libreria per fare anche grafici iterativi
- 7) (se ci sarà tempo) BeautifulSoup: per estrarre dati da pagine web (Web Scraping).

Sono tutte librerie open source, come quasi tutte le librerie

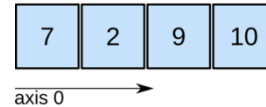




NumPy

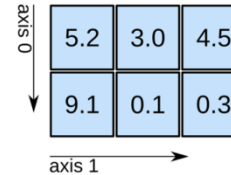
- Array N-dimensional potenti con operazioni di vettorizzazione, indicizzazione e broadcasting
- Strumenti per il calcolo numerico: funzioni matematiche, generatori di numeri casuali, operazioni di algebra lineare, e molto altro
- Performante: Il nucleo di NumPy è scritto in C altamente ottimizzato
- Interoperabile tra diversi hardware e piattaforme

1D array



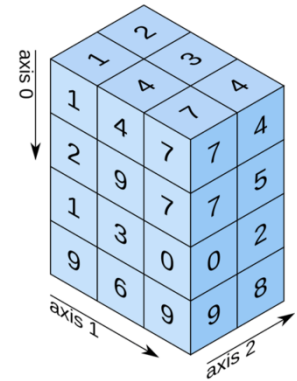
shape: (4,)

2D array



shape: (2, 3)

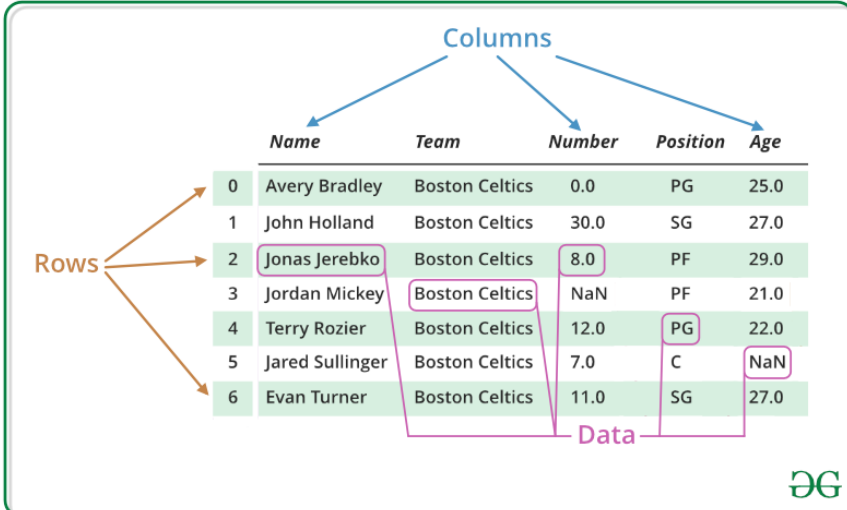
3D array



shape: (4, 3, 2)

pandas

- Si basa su NumPy e fornisce strumenti potenti per lavorare con dati strutturati come tabelle, serie temporali e dataset di grandi dimensioni.
- La struttura dati principale di Pandas è il DataFrame: tabella bidimensionale con righe e colonne, simile a un foglio Excel o a una tabella SQL
- Legge facilmente file **CSV**, **Excel**, **JSON**, **SQL**, **Parquet**, ecc.
- Manipolazione dei dati: filtraggio, modifiche
- Compatibile con Matplotlib e Seaborn per creare grafici rapidamente.

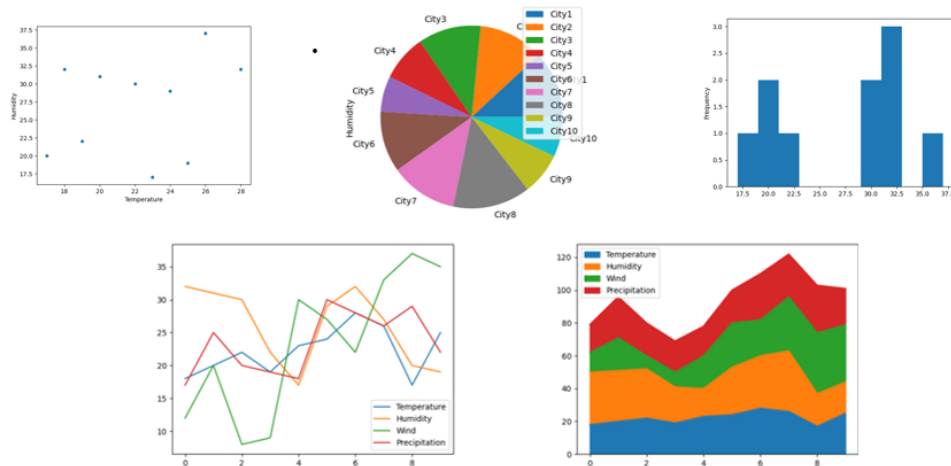


The diagram illustrates a DataFrame structure. At the top, the word "Columns" is written in blue, with arrows pointing to the column headers: "Name", "Team", "Number", "Position", and "Age". On the left, the word "Rows" is written in orange, with arrows pointing to the row indices: 0, 1, 2, 3, 4, 5, and 6. A pink box labeled "Data" is drawn around the data cells of the table, specifically highlighting the values for Jonas Jerebko (8.0), Jordan Mickey (NaN), Terry Rozier (PG), Jared Sullinger (C), and Evan Turner (SG).

	Name	Team	Number	Position	Age
0	Avery Bradley	Boston Celtics	0.0	PG	25.0
1	John Holland	Boston Celtics	30.0	SG	27.0
2	Jonas Jerebko	Boston Celtics	8.0	PF	29.0
3	Jordan Mickey	Boston Celtics	NaN	PF	21.0
4	Terry Rozier	Boston Celtics	12.0	PG	22.0
5	Jared Sullinger	Boston Celtics	7.0	C	NaN
6	Evan Turner	Boston Celtics	11.0	SG	27.0

OG

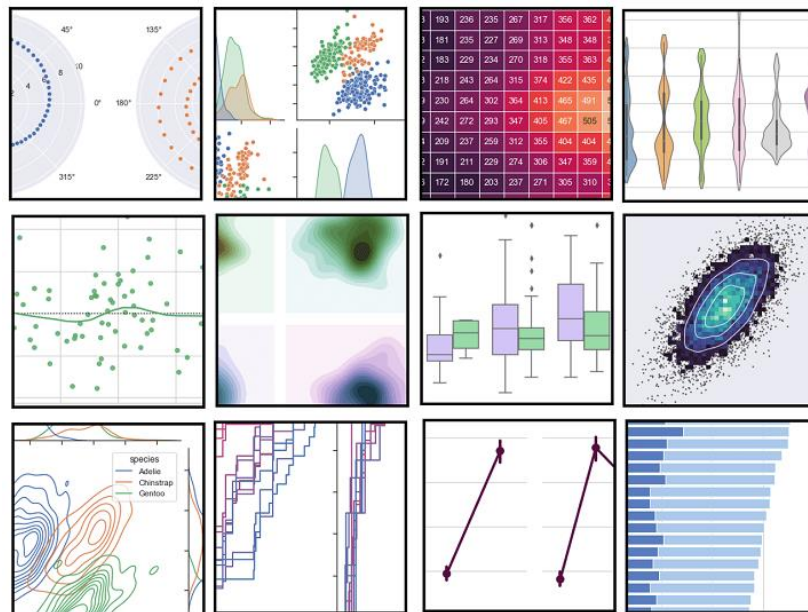
matplotlib



- Crea grafici 2D statici come linee, barre, istogrammi, scatter plot, ecc
- Rappresenta più di due dimensioni con colori, grandezza punti,..
- Crea grafici 3D con l'aggiunta di un modulo
- Integrazione con NumPy e Pandas: perfetto per dataset tabulari e array numerici.
- Può gestire dati di tipo misto



seaborn



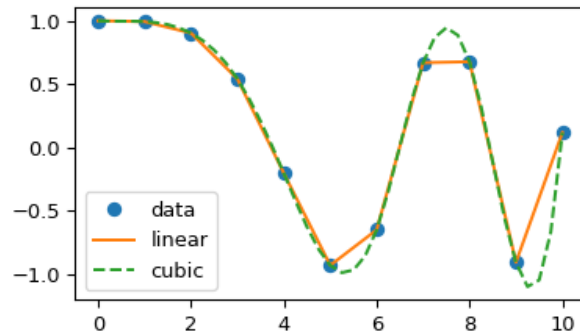
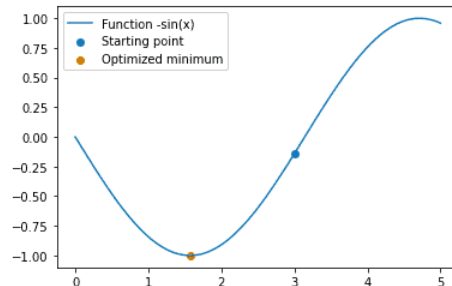
- Crea grafici statistici più avanzati come boxplot, violinplot, heatmap, ...
- I dati devono essere omogenei
- Rappresenta più di due dimensioni con colori, grandezza punti,..
- Richiede meno memoria ed è più veloce di matplotlib
- Statistiche integrate: può calcolare automaticamente regressioni, distribuzioni e relazioni tra dati.



SciPy

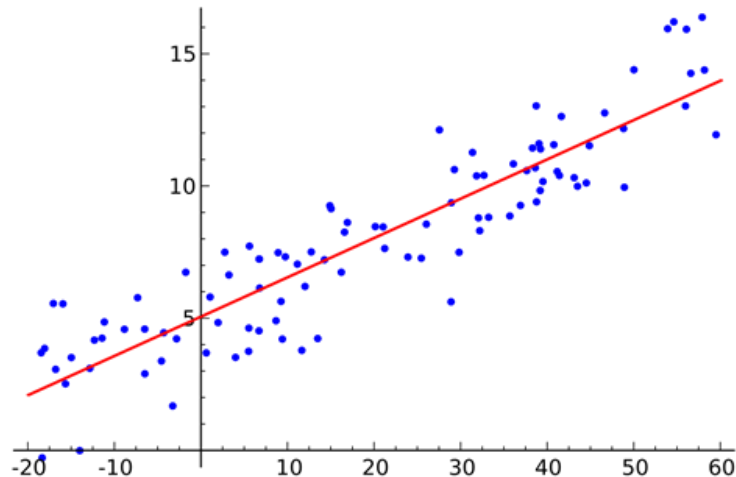
- Utilizzata per calcoli scientifici, analisi numerica e ottimizzazione.
- Si basa su NumPy e fornisce strumenti avanzati in particolare per:
 - algebra lineare,
 - ottimizzazione
 - statistica e probabilità
 - integrazione
 - interpolazione e fitting

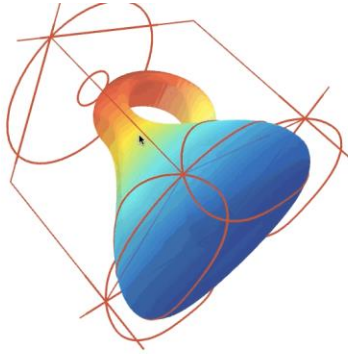
```
import matplotlib.pyplot as plt
plt.plot(x, f(x))
plt.scatter(start, f(start))
plt.scatter(optimized, f(optimized))
plt.legend(['Function -sin(x)', 'Starting point', 'Optimized minimum']);
```





- Una delle librerie più popolari per il Machine Learning in Python.
- Offre strumenti semplici ed efficienti per apprendimento automatico, analisi dei dati e modellazione predittiva.
- Si basa su NumPy, SciPy e Matplotlib ed è utilizzata in molti settori per costruire modelli di classificazione, regressione, clustering e riduzione della dimensionalità.
- Noi useremo solo la regressione lineare per fare un modello predittivo





[see the interactive plot](#)



- Crea grafici interattivi e dinamici
- È particolarmente utile per visualizzazioni avanzate, dashboard web e grafici 3D.

BeautifulSoup

- Utilizzata per estrarre dati da pagine web (Web Scraping).
- Permette di analizzare e manipolare il codice HTML e XML, facilitando la navigazione nei dati.
- Perfetta per analisi automatizzate: raccolta di dati da news, e-commerce, social media, ecc.



doc.html

Soup
Preparation
→



```
soup =  
BeautifulSoup('doc.html', 'html.parser')
```

Strumenti



Visual Studio Code



GitHub



Materiale del corso

- Sito del corso: https://github.com/lauranenzi/ProgrammingLab_II
- Libro: Python Data Science Handbook, Second Edition, Jake VanderPlas
- Tutorial on scientific python: <https://lectures.scientific-python.org/index.html>
(sul sito c'è anche il pdf del tutorial)

Altri Materiali Online

- Video tutorial su Jupyter notebook in italiano:
<https://www.youtube.com/watch?v=ExAX5cLtSu4>
- Colab: <https://research.google.com/colaboratory/>
- Introduzione a JupyterLab in Colab:
https://colab.research.google.com/github/bebi103a/bebi103a.github.io/blob/master/lessons/00/intro_to_jupyterlab.ipynb

Modalità d'esame

- 20% Repository su GitHub
- 80% Esame scritto di 2 ore, l'ultima lezione faremo un'esercitazione\simulazione d'esame
- L'orale è facoltativo se lo scritto è sufficiente, obbligatorio se il voto è quasi sufficiente ed obbligatorio a discrezione del docente se vengono rilevate delle irregolarità
- La prova orale si sostiene solo nell'appello in cui si svolge lo scritto, ed in date e orari concordati col docente.