

Python, Machine Learning y Modelos

Generalidades

PhD(c) MSc. Jesus Alvarado Huayhuaz

Universidad Peruana Cayetano Heredia
Ingeniería Biomédica

Outline

① Python

Rápido, IA! Calcule as coordenadas de dobradura para o sistema Bal Panda!

Opa. Tipo errado de IA.
Eu sou apenas um cara em uma roupa de robô. Isto é estranho.

② Librerías

③ ML

④ Desafíos

⑤ Conclusiones

⑥ Complementos



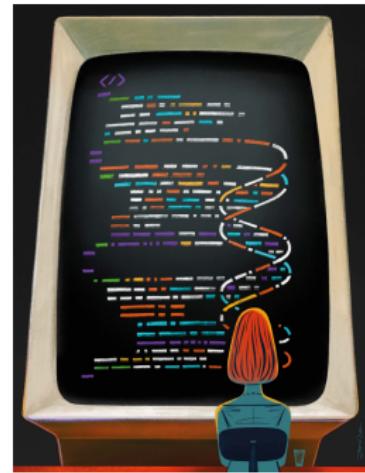
a

^aShane,J. (2022). Alta Books

10 códigos que transformaron la ciencia ¹



- ① Language pioneer: the Fortran compiler (1957)
- ② Signal processor: fast Fourier transform (1965)
- ③ Molecular cataloguers: biological databases (1965)
- ④ Forecast leader: the general circulation model (1969)
- ⑤ Number cruncher: BLAS (1979)
- ⑥ Microscopy must-have: NIH Image (1987)
- ⑦ Sequence searcher: BLAST (1990)
- ⑧ Preprint powerhouse: arXiv.org (1991)
- ⑨ Fast learner: AlexNet (2012)
- ⑩ Data explorer: IPython Notebook (2011)



¹ Nature 589, 344-348 (2021)

Zen de Python



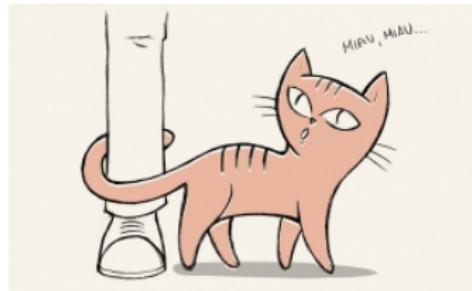
- ➊ Bello es mejor que feo.
 - ➋ Explícito es mejor que implícito.
 - ➌ Simple es mejor que complejo.
 - ➍ Complejo es mejor que complicado.
 - ➎ Plano es mejor que anidado.
 - ➏ Espaciado es mejor que denso.
 - ➐ La legibilidad es importante.
- ...



¿Por qué usamos Python? ²



Python es un lenguaje de *alto nivel*, pensado para ser relativamente sencillo de leer y escribir para las personas, y fácil de leer y procesar para las máquinas.



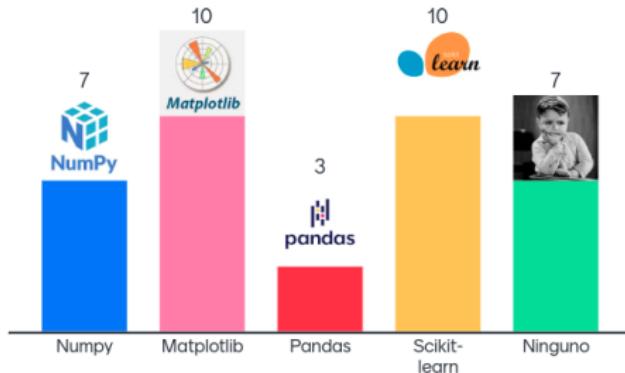
- False
- True
- as
- async
- class
- def
- elif
- except
- from
- if
- in
- lambda
- or
- raise
- try
- with
- None
- and
- assert
- await
- continue
- del
- else
- finally
- global
- import
- is
- nonlocal
- pass
- return
- while
- yield

²Severance, CR (2009). Python para todos

Librerías

Código para Mentimeter: 78688711

Con qué librería estás familiarizad@??



Librerías

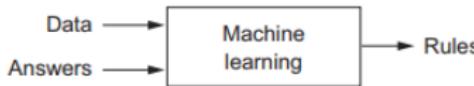
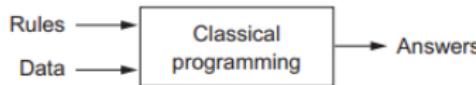
- 1 NumPy: Cálculo numérico con Python. Proporciona un conjunto de funciones y herramientas para trabajar con matrices y vectores.
- 2 Pandas: Análisis de datos que proporciona herramientas para manipular y analizar datos de diversas fuentes.
- 3 Matplotlib: Visualización de datos que permite crear gráficos de alta calidad, diagramas de dispersión, histogramas, entre otros.
- 4 SciPy: Cálculo científico que proporciona algoritmos numéricos y herramientas para la resolución de problemas matemáticos.
- 5 Scikit-learn: Una librería para aprendizaje automático que proporciona herramientas para la clasificación, regresión, agrupación y otras técnicas de análisis de datos.



Machine Learning



- ① 1840 - Charles Babbage: Máquina analítica
- ② 1843 - Ada Lovelace: "La máquina analítica no tiene ninguna aspiración de originar algo."
- ③ 1950 - Alan Turing: "La objeción de lady Lovelace".³
- ④ ML como un nuevo paradigma de programación.⁴



³Turing, AM. "Computing and Intelligence", Mind 59, 236 (1950):433-460

⁴Chollet, F. (2021). Deep learning with Python.

Machine Learning

Código para Mentimeter: 78688711

Machine learning es considerado un nuevo paradigma de programación porque:



Utiliza modelos estadísticos y algoritmos para aprender de los datos.



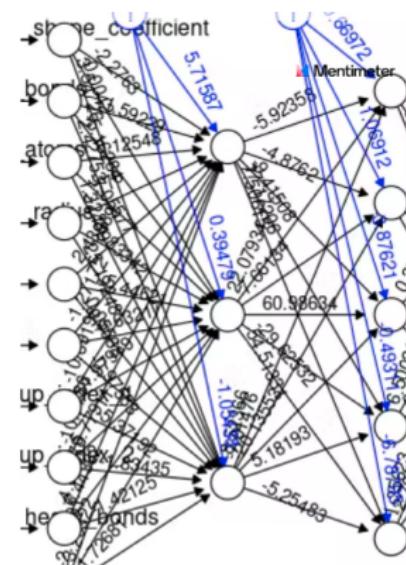
Se basa en el aprendizaje a partir de la memoria.



Es una técnica de programación tradicional ampliamente utilizada en la industria.

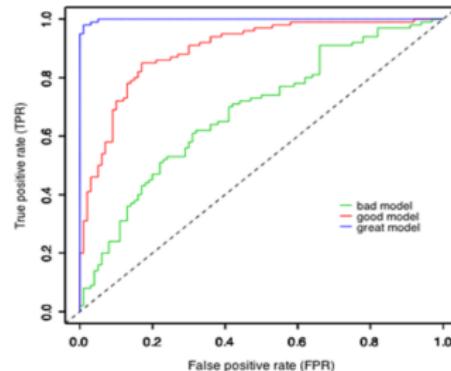
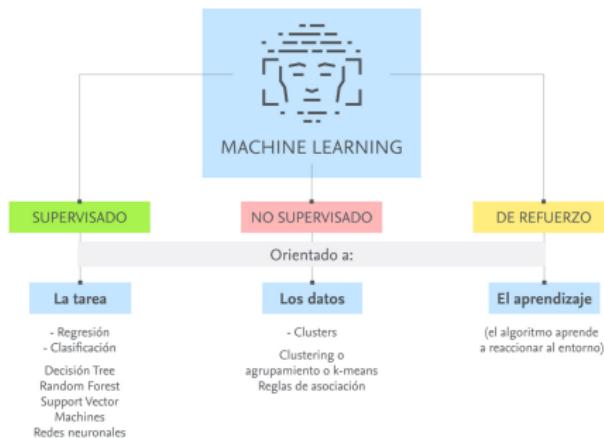


Utiliza una sintaxis de programación única y específica.



Modelos y evaluación

- Un modelo de ML es la salida de información que se genera cuando entrena su algoritmo de machine learning con datos. Después del entrenamiento, al proporcionar un modelo con una entrada, se le dará una salida. Por ejemplo, un algoritmo predictivo creará un modelo predictivo.
- Las métricas dependerán del tipo de modelo: AUC ROC, R2, MAE, etc.



¿Qué otros desafíos encontramos?

Principales desafíos de Machine Learning



- ① Datos: Calidad y cantidad.
 - ② Maldición de la dimensionalidad.
 - ③ Sobreajuste y subajuste.
 - ④ Costo computacional.
- Optimización de los modelos.
 - Hardware y servicios en la nube.

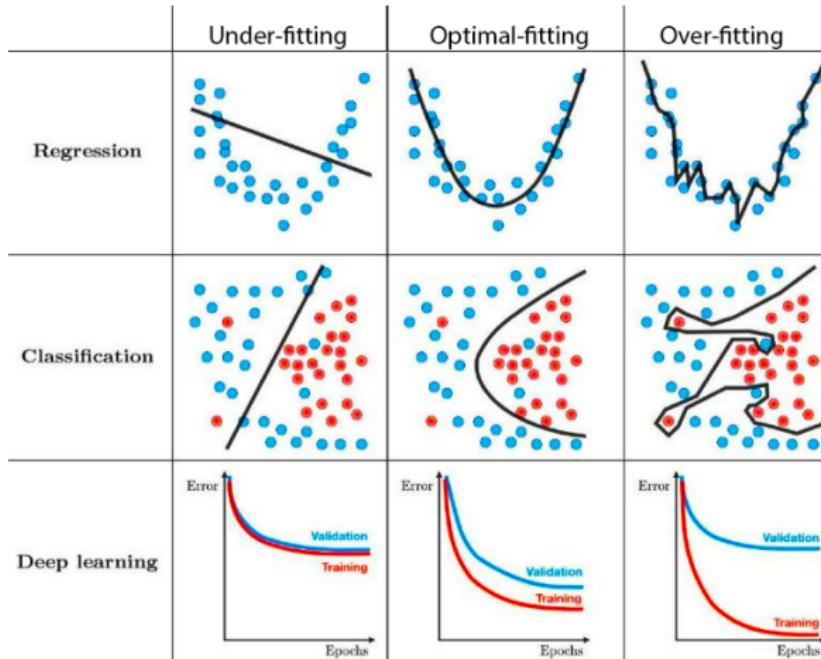


¿Qué otros desafíos encontramos?

Principales desafíos de Machine Learning



- Sobreajuste y subajuste [Mentimeter: 78688711]



Conclusiones

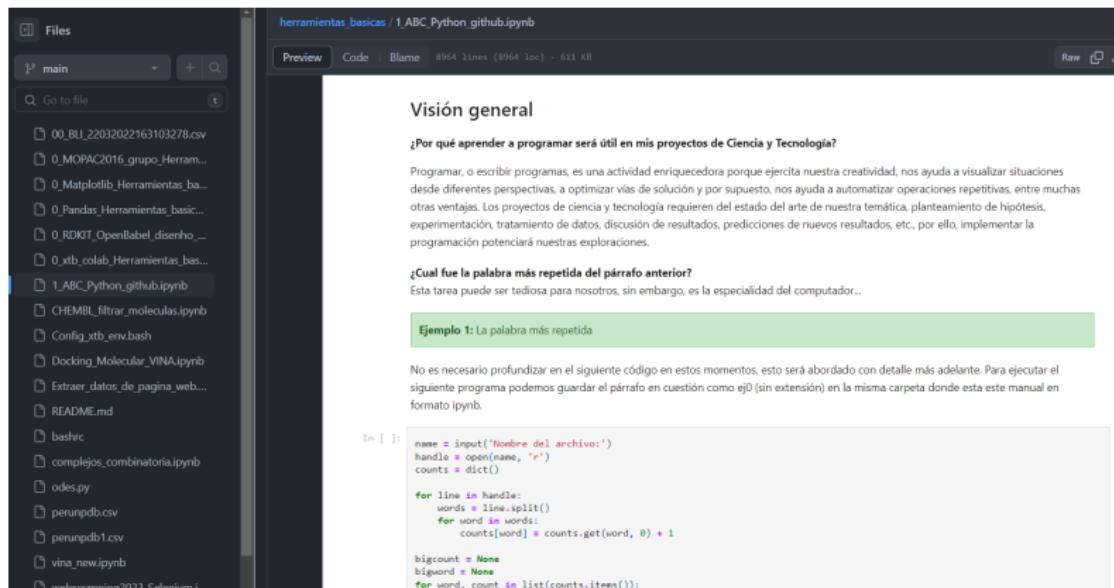


- ① Justificamos el uso de Python y sus principales librerías
- ② Abordamos el paradigma de programación en el aprendizaje automatizado
- ③ Conocimos diferentes modelos de ML y algunas métricas de evaluación



Material Complementario

- Notebook 1: Repaso de "Python para todos"



The screenshot shows a Jupyter Notebook interface. On the left, a sidebar displays a file tree with various Python files and CSV files. The main area shows a notebook titled "herramientas_basicas / 1.ABC_Python.github.ipynb". The first cell, titled "Visión general", contains text about the utility of learning to program for science and technology, mentioning its applications in optimization, automation of repetitive tasks, and hypothesis testing. It also asks for the most repeated word in the previous paragraph and notes that this task is tedious for humans but easy for computers. A green box highlights the question "¿Cuál fue la palabra más repetida del párrafo anterior?". The second cell, titled "Ejemplo 1: La palabra más repetida", contains Python code for reading a file and counting the frequency of each word. The code uses a dictionary to store word counts and iterates through each line of the file, splitting it into words and updating the count for each word.

```
In [ ]: name = input('Nombre del archivo: ')
handle = open(name, 'r')
counts = dict()

for line in handle:
    words = line.split()
    for word in words:
        counts[word] = counts.get(word, 0) + 1

bigcount = None
bigword = None
for word, count in list(counts.items()):
```

Material Complementario



- Notebook 2: Introducción a Python

The screenshot shows a Google Colab notebook interface. At the top, there's a toolbar with a 'co' logo, the notebook title 'Introduccion_a_Python.ipynb', and a star icon. Below the toolbar are menu options: Archivo, Editar, Ver, Insertar, and Entorno de ejecución. A sidebar on the left contains icons for file operations like new file, open, save, and delete, along with a search bar and a code/text switcher. The main content area is titled 'Tabla de Contenidos' and lists the following topics:

- [Variables](#)
- [Tipos de Datos](#)
- [Funciones](#)
- [Librerías](#)
- [Tablas](#)
- [Simple plots](#)
- [Multiple plots](#)
- [Regresión lineal](#)

Google
colab

Referencias



Jesús Bobadilla.

Machine learning y deep learning: usando Python, Scikit y Keras.
Ediciones de la U, 2021.



Francois Chollet.

Deep learning with Python.
Simon and Schuster, 2021.



Jeffrey M Perkel.

Ten computer codes that transformed science.
Nature, 589(7842):344–349, 2021.



Charles R Severance.

Python para todos.
2009.