

# Aprendizagem Automática 2022/23

## Ficha prática 1

Nesta aula deverá instalar as ferramentas de software adaptadas ao sistema operativo do seu computador.

A disciplina de aprendizagem automática usa, como ferramentas de base, a linguagem de programação **Python** e a biblioteca **scikit-learn**, que é uma das muitas bibliotecas disponíveis num ambiente Python.

Estas são ferramentas open source de acesso livre, instalável em Linux (Ubuntu, entre outras distribuições), MacOS, e Windows. Este ambiente tem um conjunto de algoritmos de aprendizagem automática, com documentação detalhada sobre cada algoritmo.

Existe um conjunto de outras bibliotecas que são dependências do scikit-learn que têm grande utilidade em aprendizagem automática, tais como: **numpy**, **scipy**, **matplotlib**, e **pandas**. Uma outra ferramenta útil é o **IPython**, que é um interpretador interativo para várias linguagens de programação, especialmente orientado para a linguagem Python. Um projeto spin-off do interpretador IPython, é o projeto **Jupyter Notebook** que é um ambiente web-based interativo do tipo REPL (read-eval-print loop) especialmente útil para correr, fazer debug e partilhar código python.

## Instalação de software

### Python

Instale a versão 64 bits do Python3 seguindo as instruções em <https://www.python.org>.

No sistema Linux/Ubuntu pode utilizar o comando

```
$ sudo apt-get install python3
```

Verifique as opções disponíveis para a instalação do scikit-learn, em função do seu sistema operativo/distribuição em <https://scikit-learn.org/stable/install.html>. Aconselha-se a que use um sistema gestor de pacotes python, como por exemplo o **pip**.

No sistema Linux/Ubuntu faça a instalação usando pip. Para tal, instale primeiro o pacote do pip, com,

```
$ sudo apt-get install python3-pip
```

## Scikit-learn

Após a instalação dos pacotes python3 e python3-pip já pode instalar as diversas bibliotecas no ambiente Python usando o gestor de pacotes pip.

```
$ pip3 install numpy scipy matplotlib ipython scikit-learn pandas
```

**Nota:** por vezes, os sistemas operativos podem ter várias versões instaladas de Python e pip pelo que deve verificar em qual o scikit-learn ficou instalado.

Para verificar a instalação, pode executar os seguintes passos na linha de comando

```
$ python3 -m pip show scikit-learn
```

```
$ python3 -m pip freeze
```

```
$ python3 -c "import sklearn; sklearn.show_versions()"
```

A biblioteca scikit-image será necessária para correr exemplos relacionados com processamento e análise de imagens. Instale também esta biblioteca específica

```
$ pip3 install scikit-image
```

## Sistema Windows

No sistema operativo Windows sugere-se a instalação da distribuição Python(x,y) disponível em <http://python-xy.github.io/>. O Python(x,y) vem com NumPy, SciPy, matplotlib, pandas, IPython e scikit-learn, todas (ou quase todas) as bibliotecas úteis ao desenvolvimento de aplicações de ciência de dados e aprendizagem automática.

## Exercício #1 - teste do scikit-learn

Para testar a utilização das funções existentes execute exemplos disponíveis em [https://scikit-learn.org/stable/auto\\_examples/index.html](https://scikit-learn.org/stable/auto_examples/index.html). Sugere-se o teste com

- plot\_hog.py : exemplo para extração de atributos de imagens
- plot\_digits\_classification.py : classificação de algarismos utilizando uma técnica de SVM (support vector machines)

Poderá ainda executar os ficheiros Python através da linha de comando. O seguinte comando executa o código existente no ficheiro plot\_hog.py

```
$ python3 -i plot_hog.py
```

Pode encontrar os exemplos online em:

[https://scikit-learn.org/stable/auto\\_examples/classification/plot\\_digits\\_classification.html](https://scikit-learn.org/stable/auto_examples/classification/plot_digits_classification.html)

e

[https://scikit-image.org/docs/stable/auto\\_examples/features\\_detection/plot\\_hog.html](https://scikit-image.org/docs/stable/auto_examples/features_detection/plot_hog.html)

**Nota:** estes exemplos estão associados a versões das bibliotecas (os links acima estão associados às versões “stable”) pelo que se usar versões anteriores das bibliotecas scikit-learn e scikit-image deverá procurar os exemplos nos mesmos sites nos links das respetivas versões. Estas bibliotecas têm sido bastante estáveis, no entanto, por vezes existem alterações de funcionalidade que põem em causa a compatibilidade entre versões, podendo nessas situações ocorrer erros quando não se usam as versões corretas dos exemplos.

Teste o exemplo `plot_digits_classification.py`. Confirme que obteve como output uma matriz de confusão (*confusion matrix*). Não é importante nesta fase compreender os resultados obtidos. No entanto, trata-se de uma das formas mais completas de apresentar o resultado de um classificador - nas linhas a verdadeira classe do objeto/instância, e nas colunas o resultado do classificador. Os valores fora da diagonal principal representam erros: objetos da classe X que foram classificados como Y, enquanto os valores na diagonal principal representam resultados corretos do classificador.

## Exercício #2 - Jupyter notebook

Um outro software útil para a disciplina é o Jupyter Notebook. O Jupyter Notebook é um ambiente interativo para a execução de código no browser. É uma ferramenta excelente para análise exploratória de dados e é muito utilizada por cientistas de dados. O Notebook Jupyter facilita a incorporação de código, texto e imagens numa mesma interface.

1. Explore sumariamente o software Jupyter em [jupyter.org](http://jupyter.org)
2. Instale o Jupyter Notebook. No sistema Linux pode utilizar o comando

```
$ sudo apt-get install jupyter
```

3. Corra e observe a demo `Lorenz.ipynb`

## Atividade extra #3 - Linguagem Python

O Python é considerada uma linguagem de eleição para aprendizagem automática, ciência dos dados, e inteligência artificial. Se este é o primeiro contacto que tem com a linguagem

Python o Kaggle disponibiliza diversos cursos hands-on. O curso de introdução à linguagem (<https://www.kaggle.com/learn/python>) permite aprender/relembrar a linguagem com foco na sua futura utilização para problemas de ciência dos dados. O curso está dividido em 7 lições e foi desenvolvido com recurso a notebooks Jupyter:

1. Hello, Python
2. Functions and getting Help
3. Booleans and Conditionals
4. Lists
5. Loops and List Comprehensions
6. Strings and Dictionaries
7. Working with External Libraries

Leia e realize os exercícios propostos nas diferentes lições.

### **Recursos para consultar**

- Python: <https://www.python.org/>
- Scikit-learn: <https://scikit-learn.org/stable/>
- Jupyter: <https://jupyter.org/>
- Kaggle: <https://www.kaggle.com/>