

Laboratory class 0

Nome:

1 - Sheldon Goulart de Alcântara

Getting started with Python, IPython Notebook, Anaconda, and Scikit-learn

Welcome to the Machine Learning course. In this course we will introduce the basic concepts of machine learning. The goal of this course is for your effort to be spent on learning the fundamental concepts and algorithms behind machine learning in a hands-on fashion. These concepts transcend any single package. What you learn here you can use whether you write code from scratch, use any existing ML packages out there, or any that may be developed in the future.

The learning approach in this course is to start from use cases and then dig into algorithms and methods, what we call a case-studies approach. The lab classes are focused on understanding how ML can be used in various cases studies, and the practical assignments will dig into the details of algorithms and methods for each of the main ML areas. In the lab classes, you will not be implementing algorithms from scratch, but rather building intelligent applications that use ML. In the practical assignments, we will be implementing and comparing a wide range of algorithms. To make it easy to implement the use cases we will be covering, we are recommending a particular set of software tools, but you can successfully complete the course with other tools out there.

Why Python

In this course, we are going to use the Python programming language to build several intelligent applications that use machine learning. Python is a simple scripting language that makes it easy to interact with data. Furthermore, Python has a wide range of packages that make it easy to get started and build applications, from the simplest ones to the most complex. Python is widely used in industry, and is becoming the de facto language for data science in industry.

We will also use the IPython Notebook in our videos. The IPython Notebook is a simple interactive environment for programming with Python, which makes it really easy to share your results. Think about it as a combination of a Python terminal and a wiki page. Thus, you can combine code, plots and text to explain what you did.

A prerequisite of attending this course is that you have learnt at least one programming language in the past. It is not our objective to teach you python. At this class we expect our students to be able pick up a language as they go. If you have not experienced python before it may be worth your while spending some time understanding the language. There are resources available for you to do this here (<https://docs.python.org/2/tutorial/> (<https://docs.python.org/2/tutorial/>)) that are based on the standard console. An introduction to the Jupyter Notebook (formerly known as the IPython Notebook) is available here (<http://ipython.org/ipython-doc/2/notebook/index.html> (<http://ipython.org/ipython-doc/2/notebook/index.html>)).

Lab classes vs Programming assignments/Research Project

In lab classes, the focus is on exploring each case study, without having to implement your own algorithms from scratch, and benefiting from the performance advantages that Scikit-learn provides. In programming assignments, you will be implementing many of these algorithms from scratch, having had the foundation of seeing them perform in practice on real applications.

In other words, in lab classes, we focus on exploring the use cases we'll tackle throughout the course. A huge goal is to familiarize ourselves with the core ML concepts that we will use during programming assignments and research project. In those, there will be much more implementation of ML algorithms, so the specific ML package becomes less important. However, in lab classes, we want to move quickly through all the use cases, and Scikit-learn will help us do just that.

Learning outcomes

This reading will walk you through the steps you will need to follow to install and get started with Python, IPython Notebook, Anaconda, and Scikit-learn.

Installing Python, IPython Notebook, and Pandas

Starting IPython Notebook

Writing variables, functions and loops in Python

Doing basic data manipulations in Python with Pandas' Frames

Getting started using these resources

All resources needed during lab classes are already installed in the laboratory machines. However, if you want to use them in your own computer, following we present how to do it.

Download and install Python, IPython Notebook through Anaconda: <https://www.continuum.io/downloads> (<https://www.continuum.io/downloads>)

Download and install Scikit-learn distribution: <http://scikit-learn.org/stable/> (<http://scikit-learn.org/stable/>)

Assignment question 1

Who invented Python and why? What was the language designed to do? What is the origin of the name "python"? Is the language a compiled language? Is it an object oriented language?

- O Python foi concebido no final de 1989 por Guido van Rossum no Instituto de Pesquisa Nacional para Matemática e Ciência da Computação (CWI);
- Python foi feita com base na linguagem ABC, possui parte da sintaxe derivada do C, compreensão de listas, funções anônimas e função map de Haskell;
- O nome Python teve a sua origem no grupo humorístico britânico Monty Python, criador do programa Monty Python's Flying Circus;
- A linguagem é interpretada e orientada a objetos.

Assignment question 2

Read on the internet about the following python libraries: Scikit-learn, numpy, matplotlib, scipy and pandas. What functionality does each provide python? What is the pylab library and how it is related to each other libraries?

- Scikit-learn => biblioteca projetada para interagir com as bibliotecas python numéricas e científicas NumPy e SciPy para classificação, regressão e agrupamento;
- Numpy => pacote com suporte a arrays e matrizes multidimensionais, possuindo uma larga coleção de funções matemáticas para trabalhar com estas estruturas;
- Matplotlib => módulo do pacote Numpy que fornece uma API orientada a objetos para incorporar gráficos em aplicativos que usam kits de ferramentas GUI;
- Scipy => uma extensão da biblioteca NumPy, onde fornece uma manipulação conveniente e rápida de um array N-dimensional;
- Pandas => biblioteca que fornece uma estrutura de dados de alto desempenho de fácil utilização e ferramentas de análise de dados;
- Pylab => módulo da linguagem Python vinculada ao pacote NumPy que permite gerar gráficos de duas dimensões de excelente qualidade, permitindo edição interativa, animações, inúmeros tipos de gráficos diferentes, anotações em sintaxe Latex e salvamento das imagens geradas em diversos formatos diferentes.

Assignment question 3

What is Jupyter Notebook and why was it inveted? Give some examples of functionality it gives over standard python. What is the jupyter project? Name two languages involved in the Jupyter project other than python.

- O Jupyter Notebook é uma aplicação web de código aberto que permite criar e compartilhar documentos que contenham código ao vivo, equações, visualizações e texto explicativo. Os usos incluem: limpeza e transformação de dados, simulação numérica, modelagem estatística, aprendizado de máquinas e muito mais. Jupyter fornece suporte para R e Scala.

Download the data and sample code and familiarize yourself with the notebooks

Before doing the assignments in this course, familiarize yourself with the two following notebooks:

Download the notebook that covers getting started with Python: Getting started with iPython Notebook

Download the notebook that covers getting started with Pandas: Getting Started with Pandas

Download the simple people dataset: people-example.csv

Save all these files in the same directory. If you are not sure where to save the files? See this guide:

Assignment question 4

Describe the following functions of SFrames:

`head()`

`tail()`

`show()`

`apply()`

- Head() -> mostra o início da tabela;
- Tail() -> mostra o final da tabela;
- Show() -> gera um gráfico com os dados da tabela;
- Apply() ->

Congratulations!!! Now you are ready to get started! From here, you will be ready to do all the assignments in the course, and build awesome intellingent applications that use machine learning!