Applied Machine Learning with Python
Part-1.Basics of Python for Machine Learning

Anthony FAUSTINE

May 2017

Outline

Introduction
What is Python ?

A very popular general-purpose programming language.

- Open source general-purpose language
- Dynamically semantics (rather than statically typed like Java or C/C++)
- Interpreted (rather than compiled like Java or C/C++)
- Object Oriented,

What can you use Python for?

- Web development (Django)
- Web Scraping (Beautiful Soup)
- Scripting Language.
- Scientific programming and Numeric Computing.
- Automation and Embedded Sytstem.
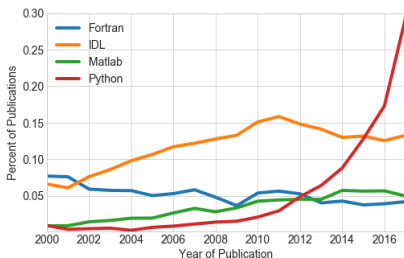- Desktop GUIs and 3D modelling.

But Why Python ?

FIGURE – Jake VanderPlas PyCon 2017

- Python is a "teaching language"

- ....created to "bridge the gap between the shell and C

- "never intended. . . to be the primary language for programmers."

Why is Python such an effective tool in science?

1. Interoperability with Other Languages
   - You can use it in the shell on microtasks, or interactively, or in scripts, or to write server software, or to build enterprise software with GUIs.

2. "Batteries Included" + Third-Party Modules
   - Python has built-in libraries for nearly everything . . . and there are third-party libraries for everything else.

3. Simplicity & Dynamic Nature
   - You can run your Python code on any architecture . . .

4. Open ethos well-fit to science
   - Easy to reproduce results with python

Python's Scientific Stack

Python Syntax

Practical Session 1

Numpy

NumPy is the fundamental Python package for scientific computing.

- Provide high-performance vector, matrix and higher-dimensional data structures.
- Implemented in C and Fortran for efficiency.
- Designed for scientific computation : linear algebra and Signal Analysis.
- Offers Matlab-ish capabilities within Python.

Numpy Arrays

NumPy provide a high-performance multidimensional array object, and tools for working with these arrays.

- A numpy array is a grid of values, all of the same type, and is indexed by a tuple of nonnegative integers.
- The essential difference between lists and NumPy arrays is functionality and speed.
  - Lists give you basic operation, but NumPy adds FFTs, convolutions, fast searching, basic statistics, linear algebra, histograms etc.
- Thus Numpy array is memory-efficient container that provides fast numerical operations.

Numpy Arrays

Practical Session 2

Matplotlib

Matplotlib is an excellent 2D and 3D graphics library for generating scientific figures.

- It provides both a very quick way to visualize data from Python and publication-quality figures in many formats.

Matplotlib

Practical Session 3

Data Analysis with Pandas
What is pandas ?

A Python package providing fast, flexible, and expressive data structures for data analysis.

- A fundamental high-level building block for doing practical, real world data analysis in Python.
- Designed to work with relational or labeled data or both $\Rightarrow$ a python version of Excel.

Data Analyis with Pandas

Practical Session 3

Machine Learning

What is Machine Learning ? :The process of extracting knowledge from data automatically with the goal of making predictions or inference.

- A classical example : Recommendations services like in Amazo or Netflix.
- Machine learning algorithms that learn to recognise what they see and hear are at the heart of Apple, Google, Amazon, Facebook, Netflix, Microsoft, etc.

Why Machine Learning

For many problems it's difcult to program the correct behavior by hand $\Rightarrow$ with machine learning these taks are easier.

Other reasons why use machine learning :

- A system might need to adapt to a changing environment.
- A learning algorithm might be able to perform better than its human programmers.
- We may want an algorithm to behave autonomously for privacy or fairness reasons.

Type of Machine Learning

Types of Machine Learning
- At a glance

Supervised Learning

- Makes machine learn explicitly
- Data with clearly defined output is given
- Direct feedback is given
- Predicts outcome/ future
- Resolves classification & regression problems

Training

Inputs → 💻 → Outputs

Unsupervised Learning

- Machine understands the data (Identifies patterns/ structures)
- Evaluation is qualitative or indirect
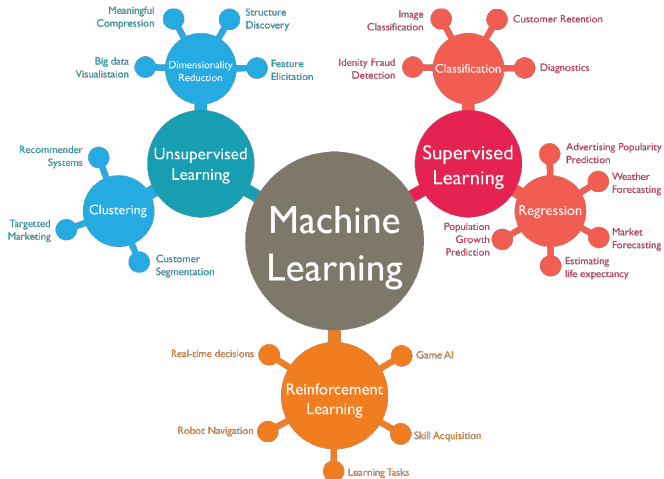- Does not predict / find anything specific

Inputs → 💻 → Outputs

Reinforcement Learning

- An approach to AI
- Reward based learning
- Learning from +ve & -ve reinforcement
- Machine learns how to act in a certain environment
- To maximize rewards

Rewards ←

Inputs → 💻 → Outputs

Type of Machine Learning

Type of Machine Learning

Machine Learning Algorithms *(sample)*

	Unsupervised	**Supervised**

Continuous	• Clustering & Dimensionality Reduction
	  ○ SVD
	  ○ PCA
	  ○ K-means
		• Regression
		  ○ Linear
		  ○ Polynomial
		• Decision Trees
		• Random Forests

Categorical	• Association Analysis
	  ○ Apriori
	  ○ FP-Growth
	• Hidden Markov Model
		• Classification
		  ○ KNN
		  ○ Trees
		  ○ Logistic Regression
		  ○ Naive-Bayes
		  ○ SVM

Scikit-Learn for ML

Scikit-Learn (sklearn) is Python's premier general-purpose machine learning library.

THANK YOU