PythonBootCamp

Pythontz: CIVE-Event

Anthony FAUSTINE

May 2017

python Tanzania

Presenter Bio

- PhD student at Nelson Mandela African Institution of Science and Technology,
- Conduct research in the field of applied machine learning and signal processing for computational sustainability with focus on building energy efficiency.
- Develop probabilistic-deep learning algorithm for energy dis-aggregation problem
- A Python enthusiast, loves to write blog posts and talk on applied machine learning with python.

python Tanzania

Outline

1 Session 1. Python-Basics

2 Session 2. Python for Scientific Computing and Data Visulization

3 Session 2. Python for Data Analysis and Machine Learning

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,

python Tanzania

What can you use Python for ?

- Web development (Django, Flask)
- Web Scraping (Beautiful Soup)
- Scripting Language.
- Scientific programming and Numeric Computing (Scipy).
- Machine Learning and Artificial Intelligence (scikit-learn)
- Automation and Embedded Sytstem.
- Desktop GUIs and 3D modelling.

But Why Python ?

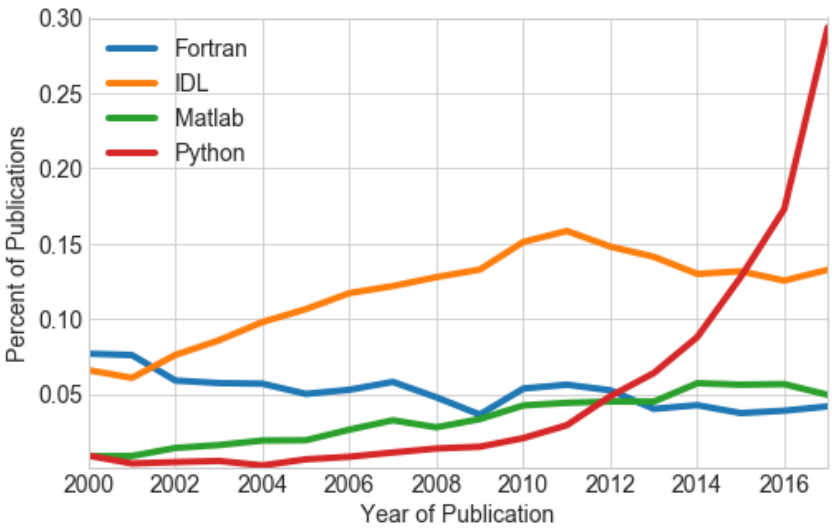


FIGURE – Jake VanderPlas PyCon 2017

- 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
5. Python is the future of AI and Machine Learning

python Tanzania

Why learn Python ?

Want a higher salary ? Become a Python programmer

2017 Average Developer Salary in the U.S.

indeed.com estimations (USD)	Language
#1 117,147	Ruby/Ruby on Rails
#2 116,027	Python
#3 115,597	C++

Resource to learn Python

10 Resources to Get Started Learning Python

Practical Session

Practical Session 1

Outline

1 Session 1. Python-Basics

2 Session 2. Python for Scientific Computing and Data Visulization

3 Session 2. Python for Data Analysis and Machine Learning

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.

List vs Numpy

- Numpy array is memory-efficient container that provides fast numerical operations.
- lists vs NumPy arrays $\Rightarrow$ functionality and speed.
  - Lists $\Rightarrow$ give you basic operation,
  - NumPy $\Rightarrow$ adds FFTs, convolutions, fast searching, basic statistics, linear algebra, etc.

Matplotlib for Data Visulization

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.
- It can also be used for animations.

Other Python Library for Visualization

- Seaborn : a Python visualization library based on matplotlib.
  - It provides a high-level interface for drawing attractive statistical graphics.
  - It offers various features for building complex visualizations.

Other Python Library for Visualization

- Bokeh : s a Python interactive visualization library that targets modern web browsers for presentation.
  - It provides a high-performance interactivity over very large or streaming datasets.
  - It offers various features for creating interactive plots, dashboards, and data applications.

Practical session

Practical Session 2

Outline

1 Session 1. Python-Basics

2 Session 2. Python for Scientific Computing and Data Visulization

3 Session 2. Python for Data Analysis and Machine Learning

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.

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.

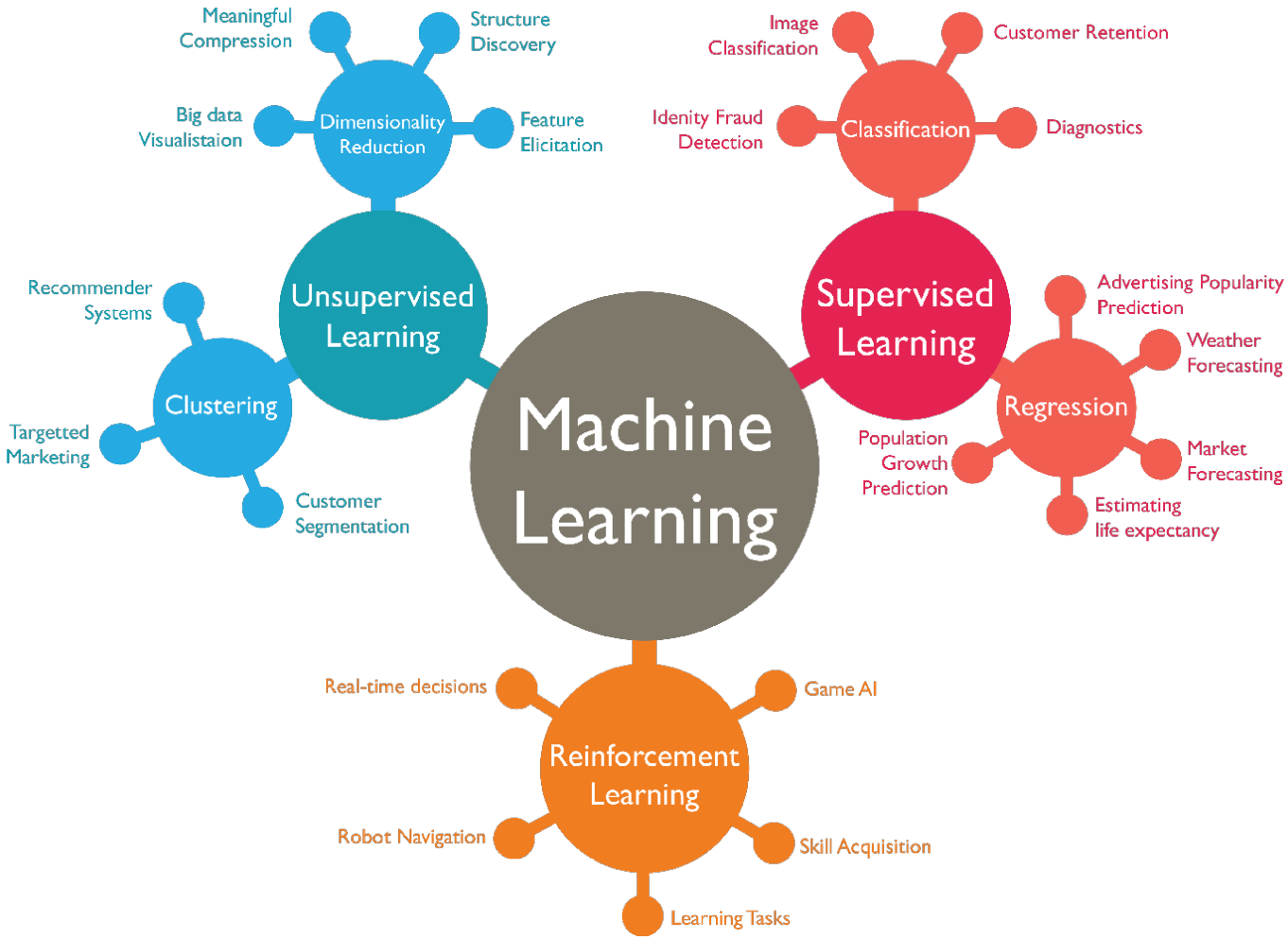
Why Machine Learning

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



Type of Machine Learning

Machine Learning Algorithms *(sample)*

	Unsupervised	Supervised

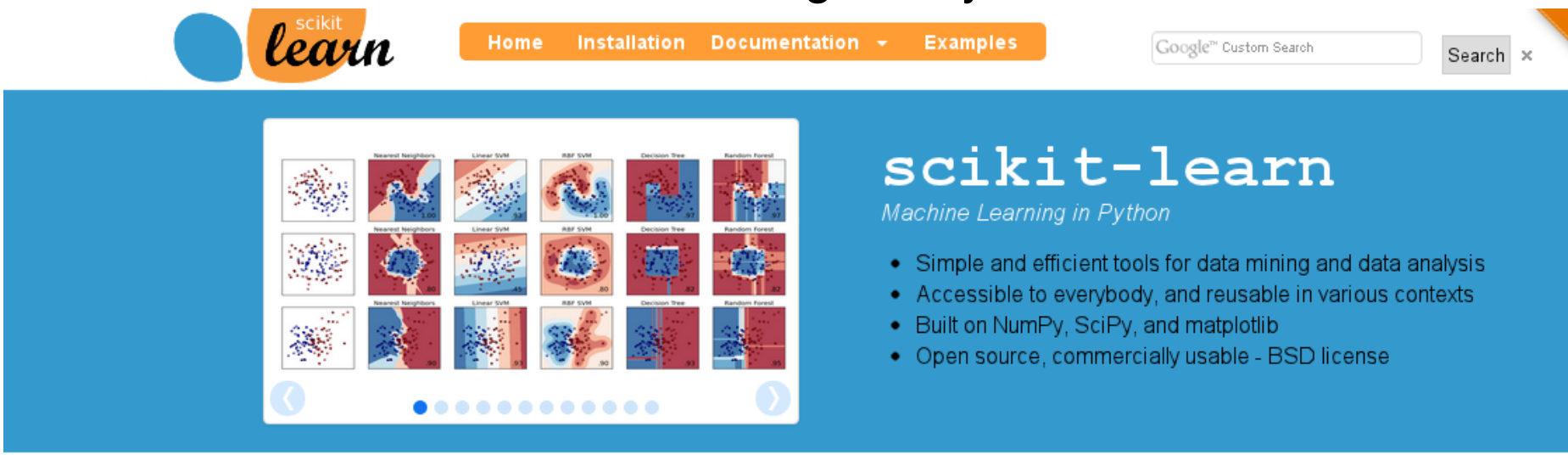
Continuous	• Clustering & Dimensionality Reduction
	  ○ SVD
	  ○ PCA
	  ○ K-means
		• Regression
		  ○ Linear
		  ○ Polynomial
		• Decision Trees
		• Random Forests

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

python Tanzania

Scikit-Learn for ML

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

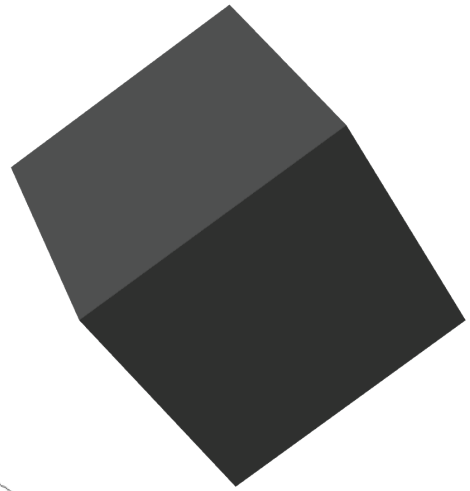


Python ML and AI libraries

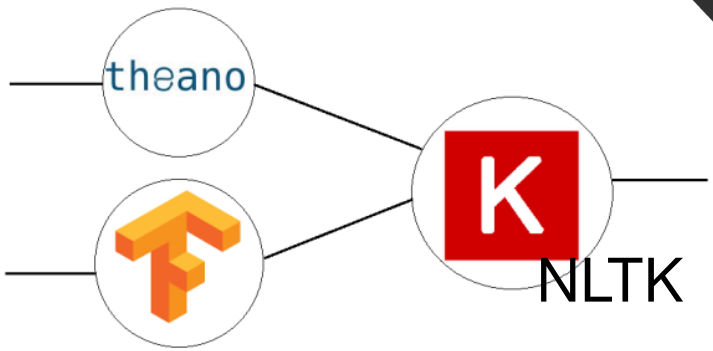
Tensorflow

PYTORCH

Keras



Theano



Pytorch

Edward

python Tanzania

Practical Session 4

THANK YOU