
Machine Learning in Python using Scikit- Learn

Submitted by Amol Agrawal as proposal for
Workshop in BangPypers

Machine Learning

“Statistics on Steroids”

Exploring and recognizing patterns using data analysis through computer science and statistics.

Lots of computer science ! Even more statistics!



Scikit-Learn

“The Black Box”

A Python module for all your Machine Learning needs.

Easy to understand, easier to implement.



—

0. Beginning -

What? Where? Why? How?



- What is Machine Learning?
- What are the aims of this workshop?
- What this workshop is not!
- Setup and basic config.

—

1. Scikit - Learn -

The module, our savior.



- About the module.
- But why Scikit - Learn ??
- Dependencies.
- Some basic terminologies.
- Let's begin !

—

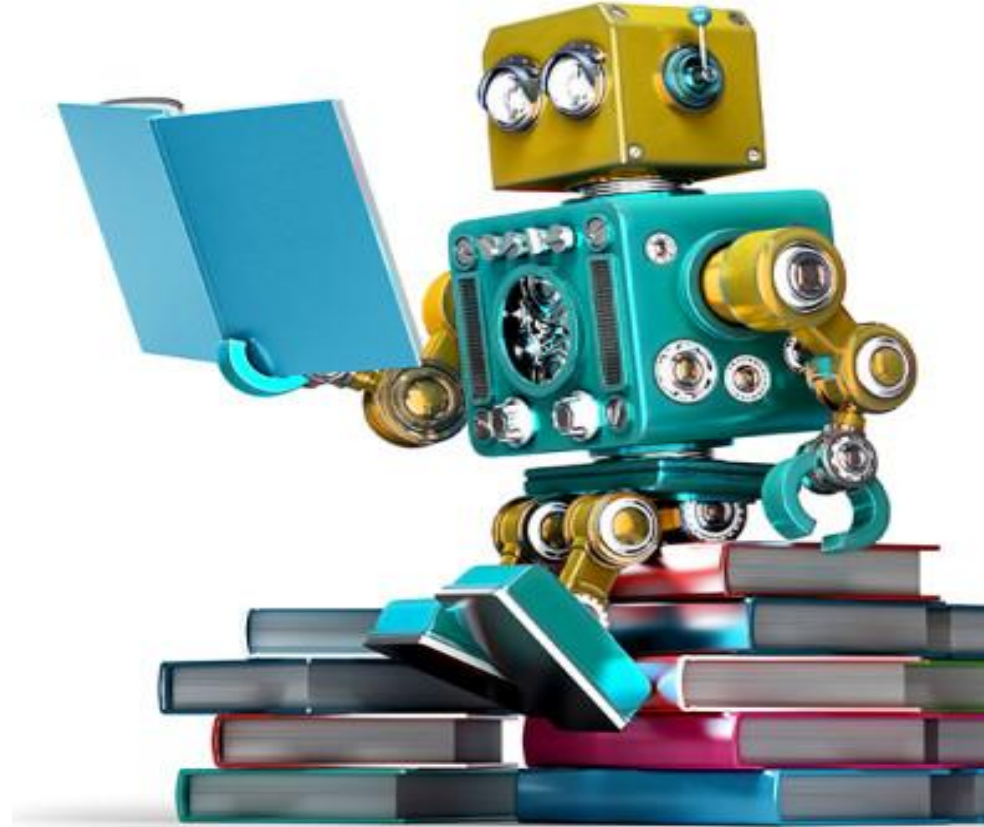
2. Machine Learning - Or How I Learned To Stop Worrying and Love the DATA

- 
- Introduction
 - Core Terminologies
 - Definitions.
 - Usage.

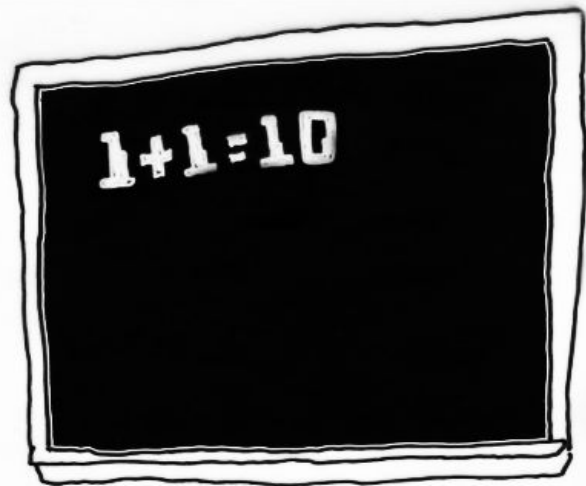
Machine Learning can be broadly categorized into

TWO CATEGORIES.

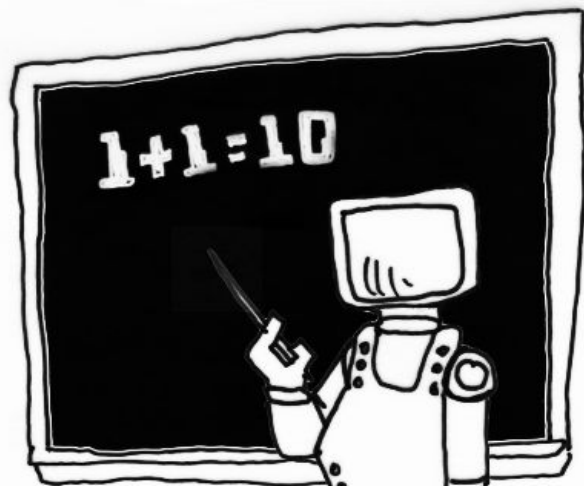
Let's look at them.



UNSUPERVISED MACHINE LEARNING



SUPERVISED MACHINE LEARNING



3. Supervised Learning - Classification and Regression



Definition

Herein, we have labeled data, i.e, for each X we know what the value of Y will be !

So for new unlabelled data, we can use the labelled data to make some predictions. .

3.1 Classification

Nearest Neighbors and Support Vector Machines



Definition

The labels have discrete values.

Eg:

Is Mail Spam or Not

Bug or Not

3.2 Regression -

Linear Regression and

Random Forests



Definition

The labels lie in a continuous range:

Eg.

Stock Market Prediction

Housing Price Prediction

4. Unsupervised Learning - Dimensionality Reduction and Clustering



Definition

We have NO LABELS !

Patterns must be found
and labels must be
generated on their own.

Time to freak out? NO !

4.1 Dimensionality Reduction - Principal Component Analysis



Definition

When the dataset has a large number of features, out of which some may be redundant, to save computational power... Reduce the Dimensions !

Eg. Working on Images or Text

4.2 Clustering -

KMeans



Definition:

Divide the dataset into clusters based on the feature set.

Eg.

Find prominent colors in an Image.

Cluster Documents

5. Model Selection and Validation -

What model to use and Hypertuning



Definition

What model to use when? Which algorithm to use when?

How to tune the parameters?

How to get the best results?

6. Real Word Examples-

Showcase of working examples of how ML is being used.



Definition:

Some real world examples of ML implemented and showcased.

7. Where to go from here ?-

How to incorporate ML in your work

Words of Caution



Definition:

How to integrate ML into
your work.

Do's and Don't's



8. Workflow

How the workshop will be conducted.:

→ **Ipython-Notebooks**

Code will be made available for each aforementioned section through Ipython-Notebooks which makes it easy to follow along, tweak and run it quickly.

Scikit - Learn library will be utilized throughout to show how it's inherently implemented algorithms can be used.

→ **Exercises**

The notebooks will be littered with incomplete code blocks for the participants to attempt and solve.



9. Requirements

→ Basic knowledge of Python



10. Takeaway

- Machine Learning Concepts
- How to use Scikit - Learn
- How to create ML models quickly using Scikit - Learn.

—

Thank You.