

1.The first package is NumPy which is

a math library to work with N-dimensional arrays in Python.

It enables you to do computation efficiently and effectively.

It is better than regular Python because of its amazing capabilities.

for working with arrays, dictionaries,

functions, datatypes and working with images you need to know NumPy.

2.SciPy is a collection of numerical algorithms and domain specific toolboxes,

including signal processing, optimization,

statistics and much more.

SciPy is a good library for scientific and high performance computation

3. Matplotlib is a very popular plotting package that provides 2D plotting,

as well as 3D plotting.

Basic knowledge about these three packages which are built on top of Python,

is a good asset for data scientists who want to work with real-world problems.

As we'll be using SciKit Learn quite a bit in the labs,

let me explain more about it and show you why it is so popular among data scientists.

SciKit Learn is a free Machine Learning Library for the Python programming language.

It has most of the **classification,**

**regression** and **clustering** algorithms,

and it's designed to work with

a Python numerical and scientific libraries: NumPy and SciPy.

Also, it includes very good documentation.

On top of that,

implementing machine learning models with SciKit Learn

is really easy with a few lines of Python code.

Most of the tasks that need to be done in a machine learning pipeline are

implemented already in Scikit Learn including pre-processing of data,

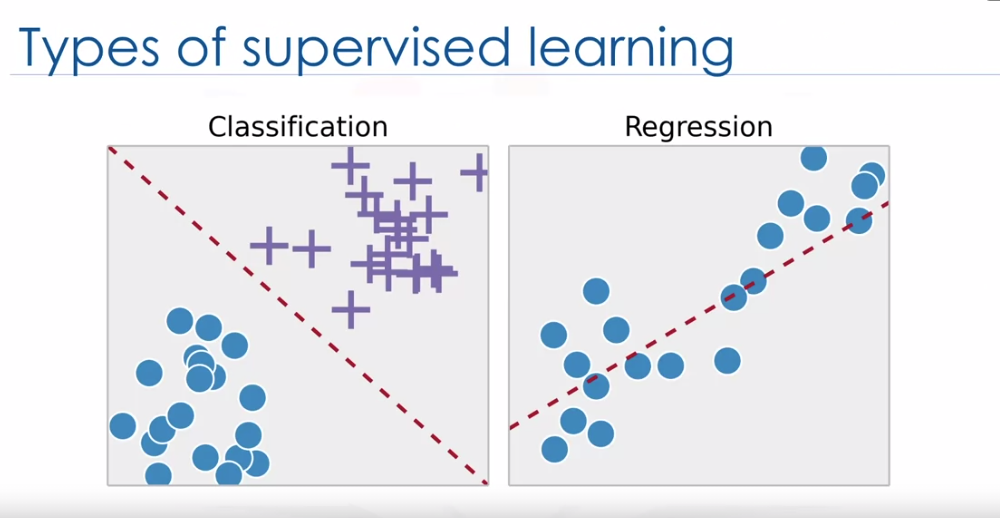
feature selection, feature extraction, train test splitting,

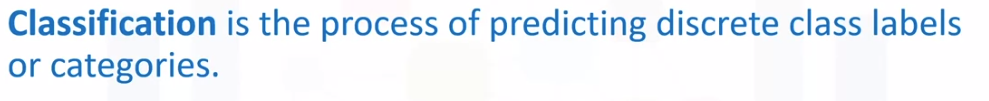
defining the algorithms, fitting models,

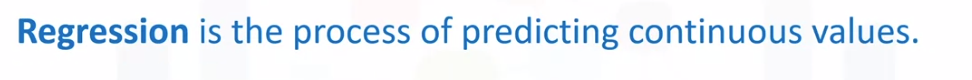
tuning parameters, prediction, evaluation, and exporting the model.

Supervised VS Unsupervised Learning

Supervised: With **supervised learning** you use labeled data, which is a data set that has been classified, to infer a **learning algorithm**. The data set is used as the basis for predicting the classification of other unlabeled data through the use of **machine learning algorithms**.

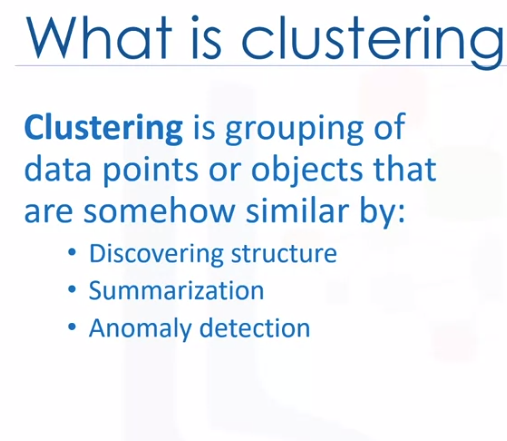






Unsupervised:

***Unsupervised learning****is a branch of*[*machine learning*](https://en.wikipedia.org/wiki/Machine_learning)*that learns from test data that has not been labeled, classified or categorized. Instead of responding to feedback, unsupervised learning identifies commonalities in the data and reacts based on the presence or absence of such commonalities in each new piece of data.*



NOW,

