

## 5.1 Intro to Machine Learning

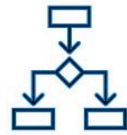
---

## 5.1 Introduction to Machine Learning

### 5.1.1 What is Machine Learning

Machine learning can be described as developing a system that can extract knowledge from data to perform a specific task.

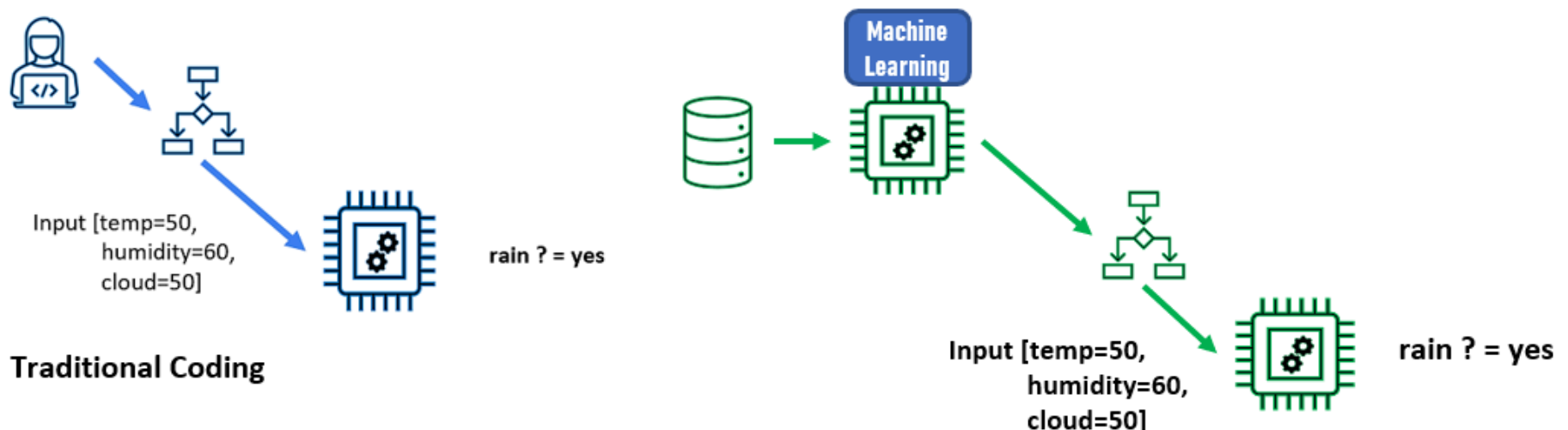
We know how to write code to complete a task. Suppose we want to write a program to tell if it is going to rain today. We may know that if the temperature is high, humidity is high, cloud cover high, it will rain. So we can design an algorithm and write an "if else clause" in Python to check the condition and tell us if it is going to rain.



```
def going_to_rain(temp, humidity, cloud):
    rain=0
    if (humidity > 60):
        rain=1
    else:
        if temp > 60 and cloud > 50:
            rain=1
    return rain
```

Can we be sure we got the condition right and the values correct? Is it possible for us to get the computer system to look at some data about situations when there was rain and no rain to write its own algorithm for us? So we only have to give the data, and the computer can learn to do the task for us?

The traditional way of writing code is for us to write the code and ask the computer to execute it. If we go back to our problem of finding out if it is going to rain or not, we can design an algorithm, code it, and execute it. When we give input with the temperature, humidity, and cloud cover, the program will tell us if it is going to rain.



With Machine learning, we can give the data to a Machine Learning algorithm and ask the Machine Learning algorithm to develop a system or model that can find out if it is going to rain or not using the data that we have given. Once we have the model or the system, we can give input with the temperature, humidity, and cloud cover and find out if it will rain.

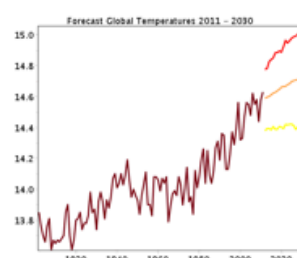
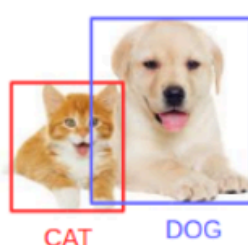
Machine Learning is like giving you a large collection of questions and corresponding answers and asking you to find out how you get the answer for similar questions. When you are given a question without an answer, you are able to answer the question based on what you have learned.

### 5.1.2 Machine Learning Applications

There are lots of applications for Machine Learning. Some of the major application areas are in Object Classification, which would help us separate Cats from Dogs or find out good or faulty products in a manufacturing plant.

We can also make predictions of temperature, stock prices, oil prices, or exchange rates using regression-type approaches.

Facebook is using ML to figure out which post out of all the posts available for you should be shown to you so that Facebook can keep you interested in continuing using Facebook without getting back to your homework.



There are many more applications in transport, banking, retail, entertainment, etc., where Machine Learning is extensively used.



### 5.1.3 Why do we need Machine Learning

If we look at some of the presented applications, it is very hard and tedious to come out with rules that can be used to do those tasks. But in most cases, we can collect large amounts of data when those tasks occur. Even if we build a set of rules and write a program with those rules, if we come across a slightly different scenario, the program will fail to work properly or give a wrong answer.

So Machine Learning allows us to ask computers to build systems by looking at relevant example data.

Whatever area you look at, there are Machine Learning applications. With the availability of a lot of data, computational power, and sophisticated Machine Learning Algorithms we see loads of applications in Character Recognition, Recommender Systems, Health Care, Medicine, Autonomous Robots, Self Driving Cars, Finance, etc.

### 5.1.4 Key Definitions (Terms) in Machine Learning

There are some key definitions that we use in Machine Learning.

**Label** is the object or class that we want to predict or classify. For example, for rain, it is Yes or No

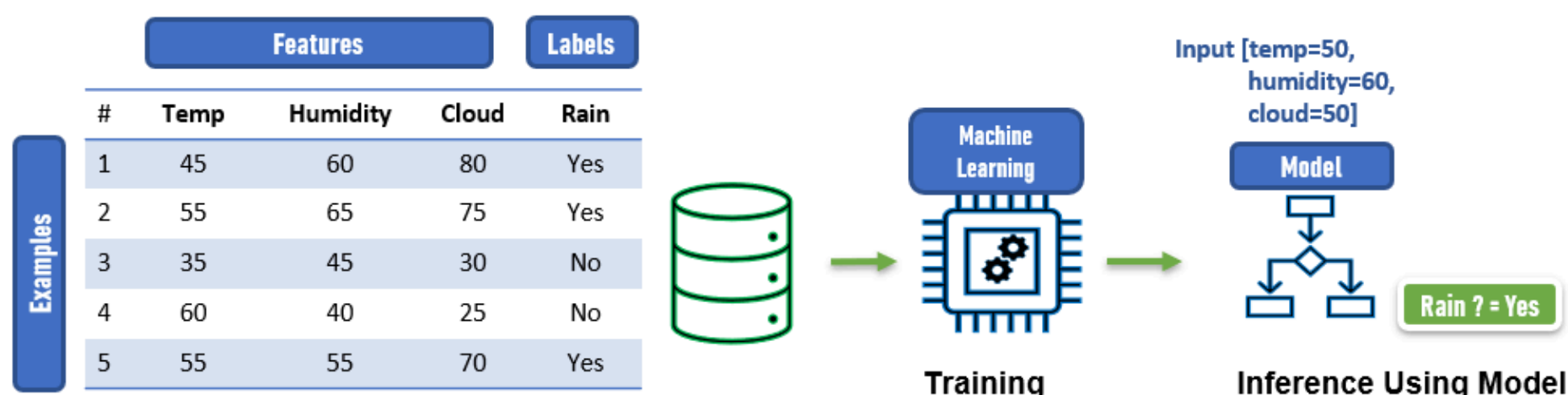
**Features** are the characteristics of the data objects we have, such as the temperature, humidity, and Cloud cover.

**Examples** are the labeled data samples that we have captured that can be used to build the model.

**Model** is the system that captures the relationship between the features of an object and the corresponding label.

**Training a model** is the activity of building the model using the labeled examples.

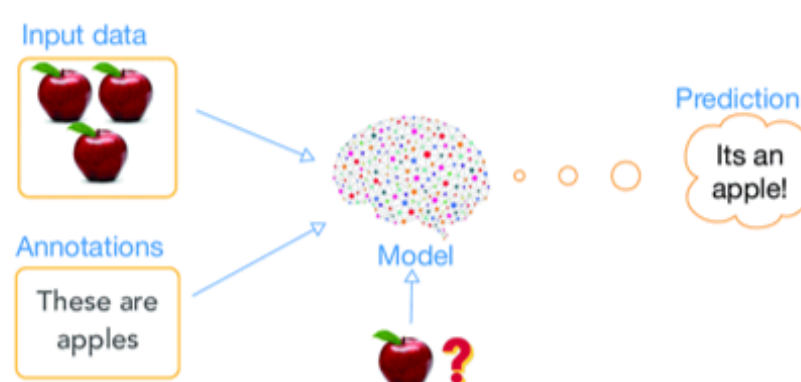
**Inference using the model** is when we attempt to use the trained model to forecast if it is going to rain or not by giving it a sample input with temperature, humidity, and cloud cover in this example.



### 5.1.5 Types of Machine Learning

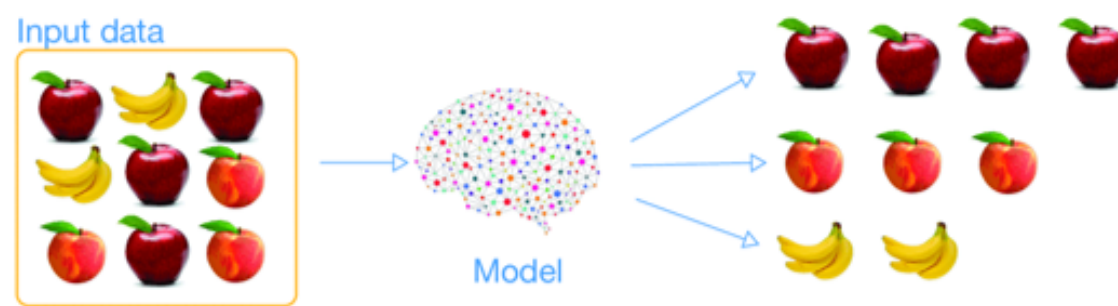
There are three major types of Learning.

**Supervised Learning** we give labelled data and asks the system to learn a model to tell us the label when we give an item without a label. In this example, if we give an unlabelled Apple to the system, it will be able to tell us that it is an Apple.



**Unsupervised Learning** is where we give a collection of data without labels and ask the Machine Learning algorithm to learn a possible way to separate them into groups based on the characteristics of the data in this example, separating them into Apples,

Oranges, and Bananas.



**Reinforcement Learning** is where we get an agent or system to continuously learn based on the rewards and punishments that it gets from the environment it is learning from. It is similar to how we would train a Dog to listen to our commands by rewarding it when it behaves the way we want it to behave and giving small punishments when it goes astray!

### 5.1.6 Machine Learning with Python

Python can be used to explore and build applications in Machine Learning because it has the power of a general-purpose language such as Java or C++. It can also be used for scripting similar to using domain-specific scripting languages such as R or MATLAB.

There are many libraries available for Python, so you do not need to write code for most of the basic Machine Learning functionalities. You can make API function calls from those libraries to do most of the Machine Learning activities.

Python also allows us to directly interact with the data using a command terminal or using notebooks that are easy to use and keep track of what you have done while building your Machine Learning Model. This helps us to quickly develop Machine Learning Models and try them multiple times with multiple variations.

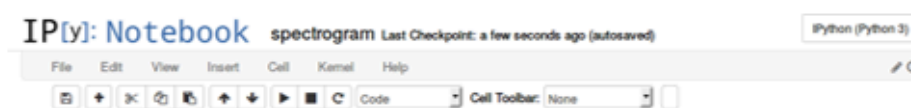
### 5.1.7 Python Libraries and Tools

Here are some of the major Python libraries and tools available for Machine Learning. In the following few slides, we will give you a very brief introduction to these libraries so that you get to know what you can do with them.

1. Jupyter notebooks
2. Pandas
3. Numpy
4. Scipy
5. Scikit-learn
6. Matplotlib

#### 5.1.7.1 Jupyter Notebooks

Jupyter Notebooks is an interactive programming environment that allows us to execute code in a browser. This is very useful for us to explore data quickly and also helps us to build Machine Learning models and test them quickly.



Text  
Description

#### Simple spectral analysis

An illustration of the [Discrete Fourier Transform](#) using windowing, to reveal the frequency content of a sound signal.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi}{N}kn} \quad k = 0, \dots, N-1$$

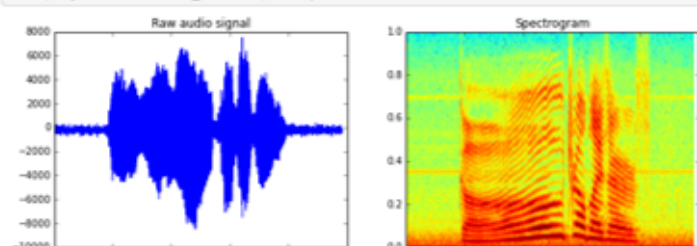
We begin by loading a datafile using SciPy's audio file support:

```
In [1]: from scipy.io import wavfile
rate, x = wavfile.read('test_mono.wav')
```

And we can easily view its spectral structure using matplotlib's builtin spectrogram routine:

```
In [2]: %matplotlib inline
from matplotlib import pyplot as plt
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 4))
ax1.plot(x); ax1.set_title('Raw audio signal')
ax2.spectrogram(x); ax2.set_title('Spectrogram');
```

Output  
Graphs



It also allows us to give text descriptions of what we have done, including giving equations and other diagrams, etc., to keep these descriptions in an easy-to-understand format. This is like commenting in Python code, but on Jupyter notebook, comments can be formatted and presented in a much nicer way making it easy to understand.

Obliviously we can have the code the most important part. And also, we can have outputs such as graphs and plots as part of the notebook, making it easy to document visualized observations that we have made about the data and maybe about the outcomes of the Machine Learning algorithms we have attempted in the code.

We can consider notebooks as a one-stop-shop for quick and easy development of Machine Learning solutions. It is like a notebook or old diary you use to write down notes about a solution to a problem and observations and code and the ability to execute the code!

You can get more information about Jupyter Notebooks from <https://jupyter.org/>

#### 5.1.7.2 Scikit-Learn Python library

Scikit-Learn is a Python library with a large number of widely used state of the art Machine Learning algorithms with some very good documentation that will help you use the library. So with this library and some relevant data, you should be able to develop a Machine Learning model in a few minutes!



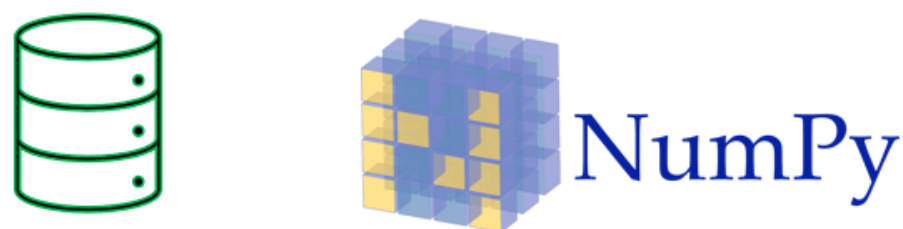
[You can learn about the API:](#)

[How to use the API:](#)

#### 5.1.7.3 NumPy Python library

NumPy is also a Python library that has a large number of facilities to help us do scientific computing or, in other words, allows us to solve some complex mathematical computations required to solve Science and Engineering problems. It provides functionality for multi-dimensional arrays and a collection of advanced mathematical operations that can be executed, such as Linear Algebra operations, Fourier Transforms, Pseudorandom number generation, etc. You can use these functionalities in Machine Learning to process the input and output data and sometimes modify some of the machine learning models.

The name NumPy comes from Numerical Python, indicating we use it to manipulate numbers in numerical computing.



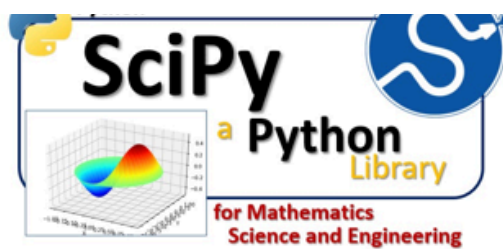
[Learn more about NumPy](#)

#### 5.1.7.4 SciPy Python library

SciPy is also another Python library with many facilities to help us do scientific computing. It has more advanced functions than NumPy, including advanced linear algebra, mathematical optimization, signal processing, mathematical, statistical distributions, etc. Once again, this library is also mainly used for data processing or, in other words, to manipulate or change your data the way you want. SciKit-learn library uses both NumPy and SciPy for building Machine Learning Models.

The name SciPy comes from Scientific Python, indicating we use it for Scientific Computing.



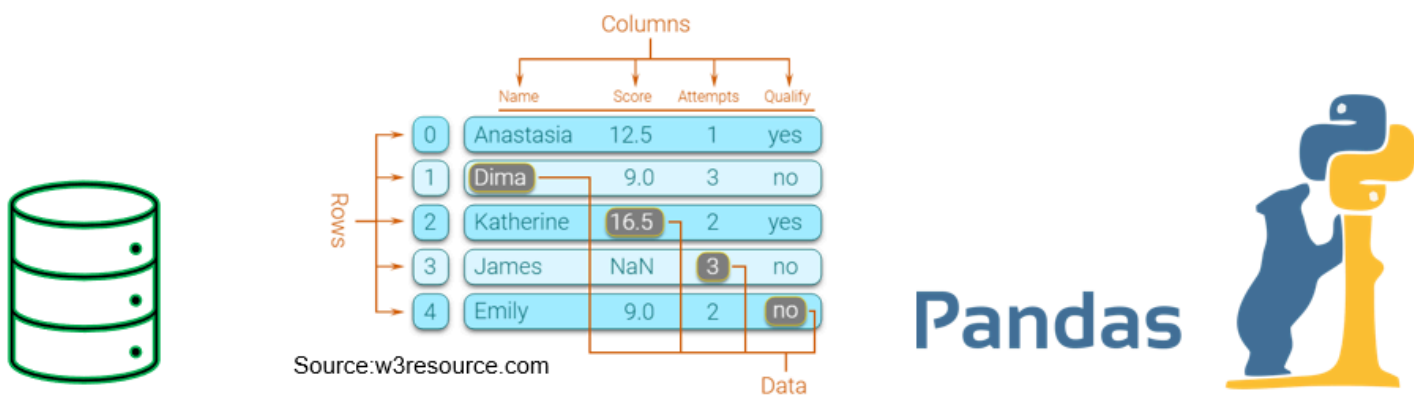


[Learn more about SciPy.](#)

5.1.7.5 Pandas Python library

Pandas is a Python library that allows you to manipulate data as if you are using a spreadsheet. You can use rows and columns to identify and write formulas to manipulate the data. Also, you can use SQL-like queries on the data. We call the place where we put the data the DataFrame. The DataFrame allows you to have multiple different data types in a single row, which is not possible in arrays that NumPy provides. We can also load data from any format such as SQL, CSV, xls to the DataFrame. In most Machine Learning projects, the first step would be to load the data to a Pandas DataFrame.

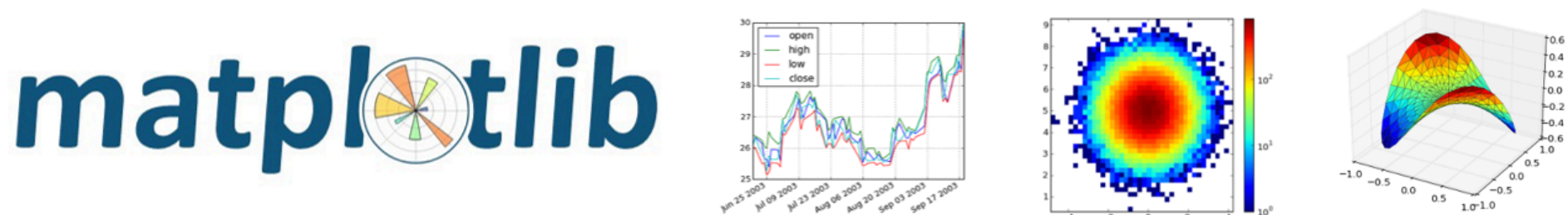
The name Pandas comes from Panel Data, indicating that the data is presented on a panel.



[Learn more about Pandas](#)

5.1.7.6 Matplotlib Python Library

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes it easy to produce very high-quality graphs very easily. With a few commands, you can draw graphs and plots. You can also have the graphs and plots in your Jupyter Notebook.



The following example code shows how you can draw a graph using the Matplotlib library. If you run the code it will show the plot. Try to see if you can add a few more data points to the graph by changing the code where it is taking the data points to plot.

Python3

Code

Run




Share

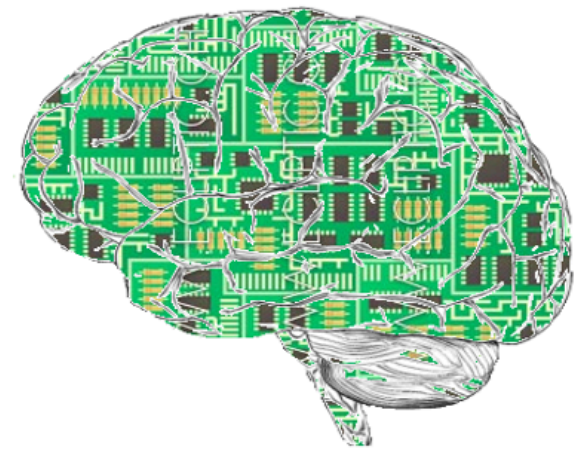
Choose File

No file chosen

5.1.7.6 Python Libraries for Deep Neural Networks

There are also a few other advanced Python libraries that allow you to use Deep Neural Networks for Machine Learning. In Neural Networks, we try to simulate how the human brain works to learn and get the computers also to learn similarly. Deep Neural networks are an advanced, more sophisticated version of Neural Networks and have been used very successfully in multiple applications, including Self Driving Cars, Natural Language Understanding, etc.

 <p>PyTorch</p>	 <p>TensorFlow</p>	 <p>Keras</p>
<p>Deep learning library from <b>Meta AI</b> (previously Facebook AI).</p>	<p>Symbolic math library for dataflow programming and deep <b>neural networks</b>, by Google.</p>	<p>High-level <b>neural network API</b>, that runs on top of <u>Tensorflow</u>.</p>
<p>Known for <b>efficient memory usage, dynamic computation graphs and flexibility</b>.</p>	<p>Known for <b>scalable production and deployment on multiple platforms</b>.</p>	<p>Good for <b>rapid prototyping</b>.</p>



### 5.1.8 Summary

In this section, we covered what is Machine Learning, Why we need Machine Learning, and what we can do with Machine Learning. Also, we got a brief introduction to some of the ML libraries and available tools. You may not need to know Machine Learning to be a full stack developer, but the purpose of introducing this topic to you is to give you the skill of using libraries in Python so that you have the confidence to make API calls using Python or any other language when required to build applications in the future as a full stack developer.

You may also be thinking, if we can get Machine Learning to write code for us, how are we going to find jobs!. Don't worry; there is a long way to go..... for this type of code development for all our requirements to become a reality.

So if you work hard and become a full stack developer, there will be plenty of jobs for you to do.

### 5.1.9 Example code to label data based on distance

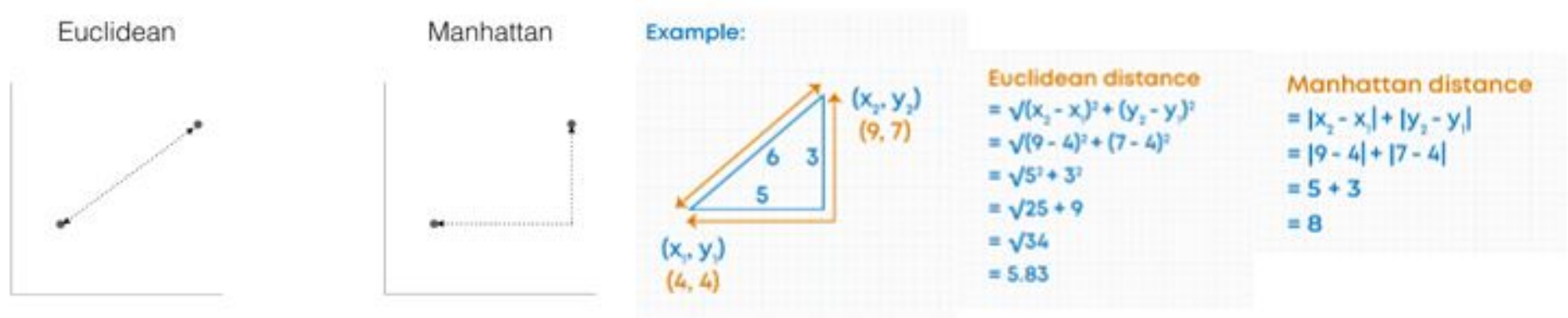
The following code is using a few data points given to it regarding rain and no rain to label a new unknown data point. Play with the code and see if you can understand how it works.

[illegible]

In this code the total distance from the unknown new data point to the members of each group is calculated to see how far it is from each group. The rain or no\_rain label is assigned to the closest group. The distance is calculated by adding the distance on the x-axis

each group. The item on the program is assigned to the closest group. The distance is calculated by adding the distances on the x-axis and the y-axis. Not the direct line distance between two points. This is called the Manhattan distance. The direct line distance is called the Euclidean distance as shown in the diagram given below. Try to see if you can change the code to use the Euclidean distance so that you can improve the labeling of the unknown data point. Also note that since we are comparing two distance values to make a decision, we can leave the Euclidean distance in the squared form rather than taking the square root.

Note when you use Python libraries for building ML models all this will be just an API call!



Jump to...

GET IN TOUCH

🏠 University of Moratuwa  
Centre for Open & Distance Learning  
CODL

☎ 011 308 2787/8  
☎ 011 265 0301 ext. 3850,3851  
✉ open@uom.lk

🌐 [University Website](#)  
🌐 [CODL Website](#)

[Data retention summary](#)