IEEE Student Branch
University of Boumerdes

Machine Learning Bootcamp

Are you ready for this!

AGENDA

01 **Python**

Fundamentals Review

02 **ML Libraries**

How to use the tools (jupiter, libraries)

03 **ML Algorithms**

SVM/ Decision Tree/K-nearest

04 **Project 01**

Iric Flower Prediction

What is python ?

Python is an interpreted, high-level, general-purpose, programming language.
It is well known for its simple syntax, which brought it very close to normal human speech, and for its large community.

General format

Imports

Functions

A **function** is a block of code which only runs when it is called.

Main program

```
import os, time
from math import cos

def add(x,y):
    return x+y

def sub(x,y):
    return x-y

x = 2
y = 3
print(add(x,y))
```

Variables

- In python, variables are defined the moment that a value has been assigned to them
- The type of the variable is determined by its content
- Everything is an object

Data types

- Integers : var = 4
- Float : var = 4.5
- String : var = “This is a string” or var = ‘This is a string’
- Boolean : var = True or var = False
- List : var = [2, “string” , True, 5.4, [“a”]]
- Dictionary : var = {key : value}

What is difference between list and dictionary?

A list is an ordered sequence of objects, whereas dictionaries are unordered sets. However, the main difference is that items in dictionaries are accessed via keys and not via their position.

Loops

For loops

```
for i in range(start, end, step):
    # do some stuff here

# default start = 0
# default step = 1
# the condition is : i < end
```

While loop

```
while condition :
    # do things

i = 0
while i < 10 :
    print(i)
    i+=1
```

Conditional statement

If statement

```
if condition :
    # do stuff
elif second condition :
    # do other stuff
elif third condition :
    # do other other stuff
else :
    # do the left stuff
```

Conditional operators

- Greater than : x > y
- Lower than : x < y
- Equal : x == y
- Great or equal / lower or equal : x >= y / x <= y
- Different : x != y
- And : X and Y
- Or : X or Y
- Not : not X
- Include : x in y

Functions

```
def function_name(arg1, arg2, arg3):
    # do some actions here
    # return something if necessary

# to call a function
function_name(arg1, arg2, arg3)
```

Module

A file containing a set of functions you want to include in your application.

```python
#creating a module

def greeting(name):
  print("Hello, " + name)

#Save this code in a file named mymodule.py. Now use the Module

import mymodule

mymodule.greeting("Jonathan")
```

Library

The **library** is having a collection of related functionality of codes that allows you to perform many tasks without writing your code. It is a reusable chunk of code that we can use by importing it in our program, we can just use it by importing that library and calling the method of that library with period(.).

```
1  import pandas as pd
2
3  df = pd.read_csv("file_name.csv")
4
```

Scikit learn

Scikit-learn is focused on **data modelling.** It comes with the support of various algorithms such as:*Classification/Regression/Clustering...*

matplotlib

Its core task to visualize the data through graphs..

Used for Data cleaning and manipulation. NumPy offers speedy computation of complex functions, and support mathematical operations

Python data analysis library and is used primarily for **data manipulation and analysis**.Some features of Pandas: Merging and joining datasets/ Handling of missing data/Data filtration options

When you start using a library without reading the documentation first

Support Vector Machine (SVM)

What is SVM?

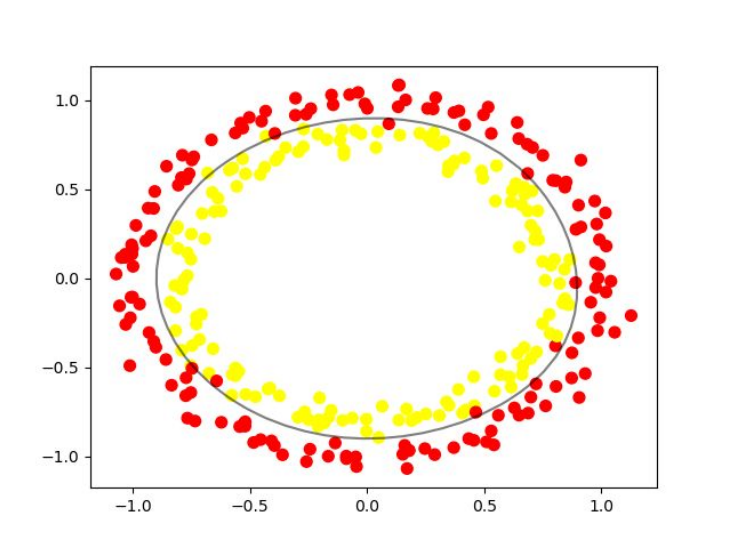
A support vector machine (SVM) is machine learning algorithm that analyzes data for classification and regression analysis.

SVM is a supervised learning method that looks at data and sorts it into one of two categories.

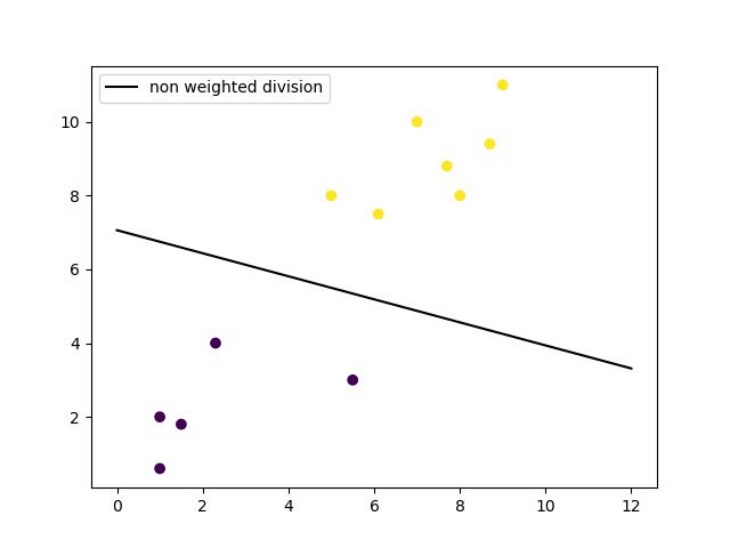
Why use SVM?

- Image classification
- Recognizing handwritten characters

Support Vector Machine (SVM)



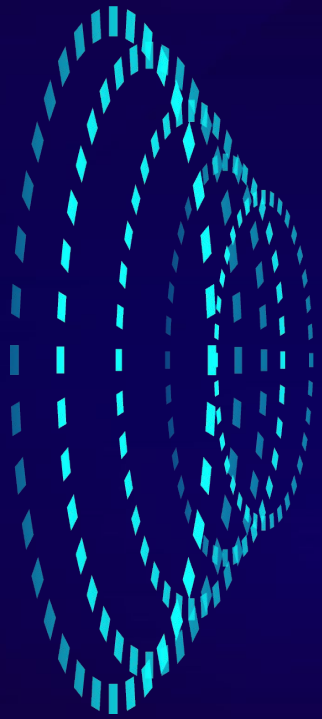
Non-linear



Linear

What is the difference between linear and nonlinear SVM?

When we can easily separate data with hyperplane by drawing a straight line is Linear SVM. When we cannot separate data with a straight line we use Non – Linear SVM.



Decision Tree

What is Decision Tree?

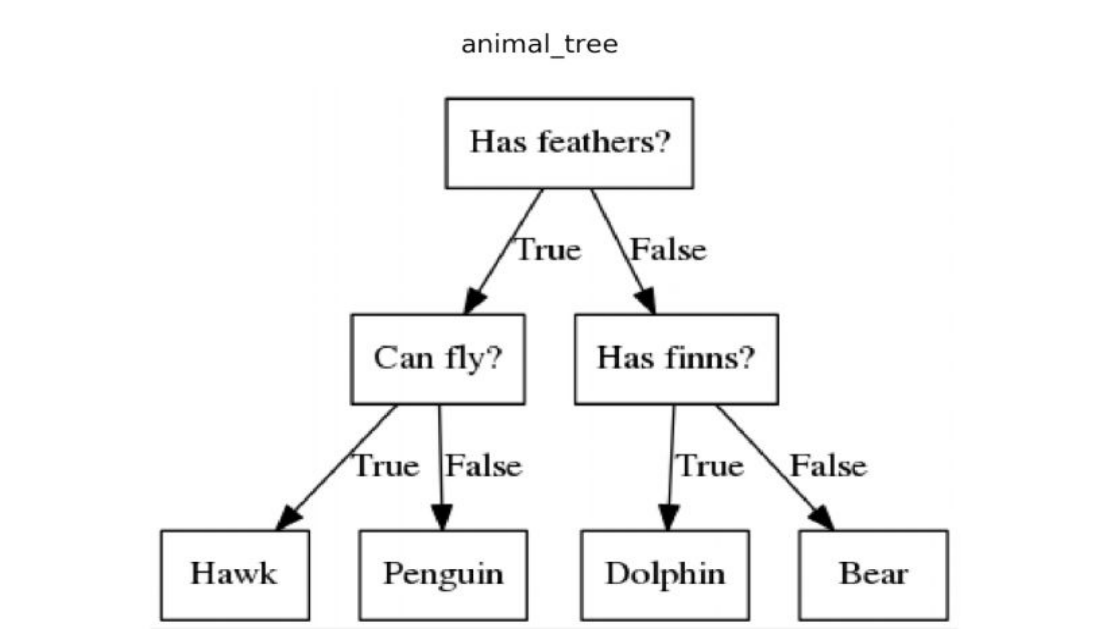
It is a Supervised learning technique that can be used for both classification and Regression, but mostly classification.

Why use it?

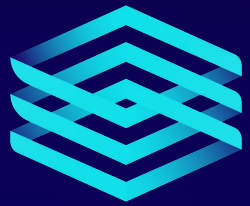
Logic easy to understand

A decision tree follows a set of if-else conditions to visualize the data and classify it according to the conditions

Internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.



K-NEAREST NEIGHBOR (K-NN)

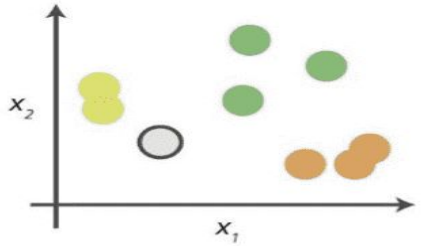


KNN tries to predict the correct class for the test data by calculating the distance between the test data and all the training points.

Then select the K number of points which is the closet to the test data.

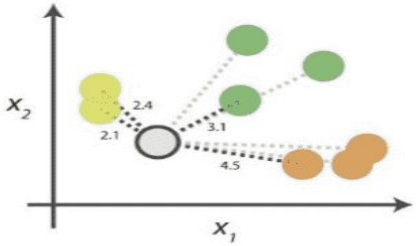
K-NN

0. Look at the data



Say you want to classify the grey point into a class. Here, there are three potential classes - lime green, green and orange.

1. Calculate distances



Start by calculating the distances between the grey point and all other points.

2. Find neighbours

Point	Distance	
○···●	2.1	→ 1st NN
○···●	2.4	→ 2nd NN
○···●	3.1	→ 3rd NN
○···●	4.5	→ 4th NN

Next, find the nearest neighbours by ranking points by increasing distance. The nearest neighbours (NNs) of the grey point are the ones closest in dataspace.

3. Vote on labels

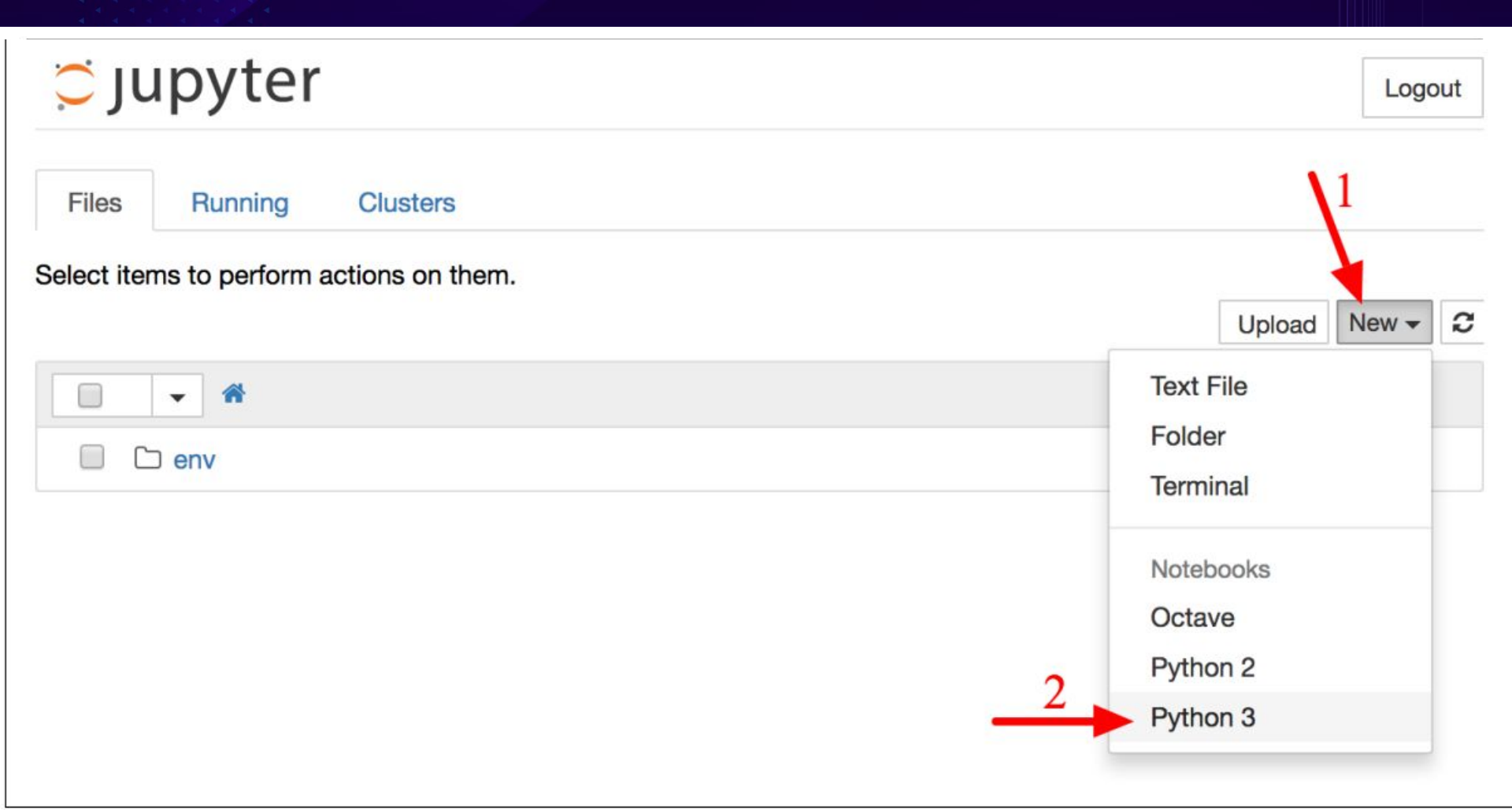
Class	# of votes
●	2
●	1
●	1

→ Class ● wins the vote!

Point ○ is therefore predicted to be of class ●.

Vote on the predicted class labels based on the classes of the k nearest neighbours. Here, the labels were predicted based on the k=3 nearest neighbours.

Let's code!


jupyter
Logout
Files
Running
Clusters
Select items to perform actions on them.
1
Upload
New
Text File
Folder
Terminal
Notebooks
Octave
Python 2
Python 3
2
env

Confusion Matrix:

The general idea is to count number of times instance of class A are classified as class B. Each raw is a class, each column is predicted class. This helps you determine accuracy of the positive predictions.
Precision=TP/ (TP+ FP)
Used with another Matric called recall (also sensitivity or the positive rate)
recall= TP/ (TP+ FN)

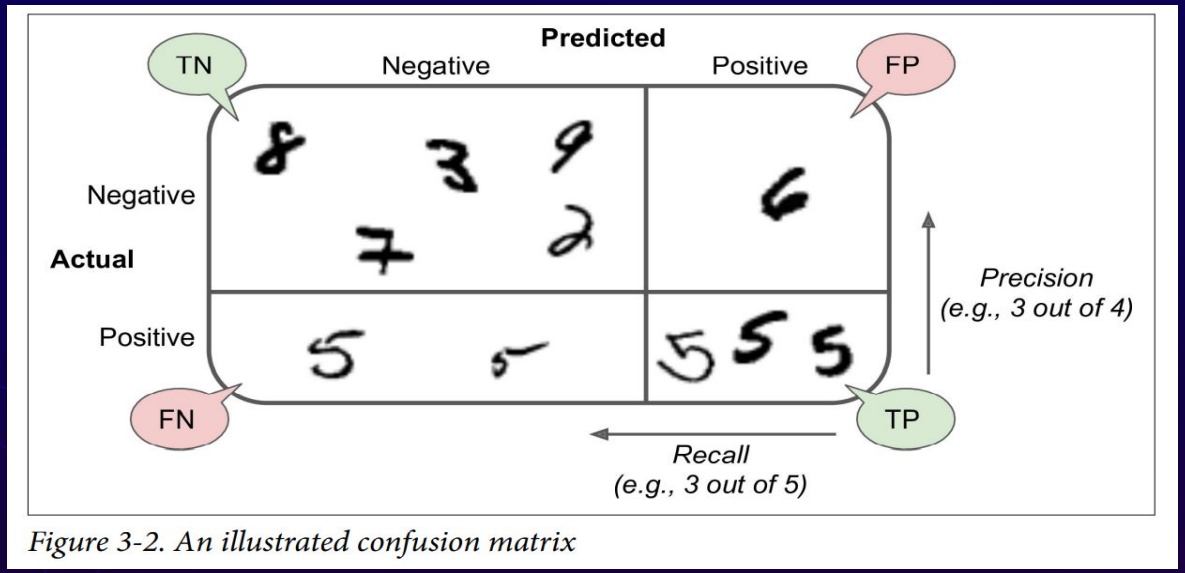


Figure 3-2. An illustrated confusion matrix

Thank You!

See you next time!

All learning resources, and challenges will be shared in our discord server

https://discord.gg/rftmD2Vd

MI BOOTCAMP