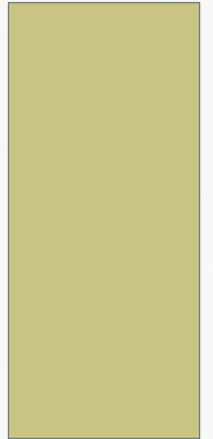


MINI-COURSE MACHINE LEARNING

BY: BOUZAGZAOU WALID



ABOUT ME



My name is Bouzaglou Walid, and I'm a Moroccan industrial engineer. Interested in the Data Science topic, I developed the skills and the knowledge to pursue my passion for Data Science.

ABOUT THIS COURSE

Hello, this mini-course is built for machine learning beginners, so if you just started looking for pieces of information about this subject this is probably the right spot.

First, I want to tell you that this course will not get you to the pro level but it will teach you the basics you are going to need on your journey. I spent so many time and I watched sooooo many videos but here you will find the juice of everything, and even the links of the best videos that I found pretty helpful, so welcome.

WHAT YOU SHOULD KNOW

- In this course, I'm going to study data science cases, and to do that, I'm going to use python, and the libraries:
 - Pandas
 - Numpy
 - Matplotlib
 - Seaborn
 - Sklearn
- When we talk about machine learning, we are talking about the Sklearn library, and it's the library this course is about.
- For the other libraries, I'm only going to explain the code I used, so I suggest to read about those libraries.

SO WHAT IS MACHINE LEARNING ?

Before explaining what does machine learning mean, I want you to know that it's not a scary subject at all, most people think that machine learning is hundreds of lines of code but this is not true, well mostly, let me explain.

There are two types of data scientists, the first type is the people who make the algorithms, those are the people who have to deal with complexed algorithms, and as I said hundreds or maybe thousand lines of code, and those people are less than 10% of the global data scientists.

Now the second type is what we want to achieve, and those are the data scientists who treat the machine learning algorithms like blocs, they know what's the job of this block of code, what kind of inputs should we give it, and what kind of outputs we are going to get.

SO WHAT IS MACHINE LEARNING ?

Now back to machine learning definition. I want you to imagine machine learning as you are teaching the machine things, you want it to do on its own later, let me explain.

Let the machine learning process be like a school course, where the programmer is the teacher, and the machine is the student. let's say the teacher wants to teach the student what a nine looks like, so he's going to show the student how the nine is written, and it looks like, and the programmer does the same thing with the machine, by giving it a bunch of photos (data) in which the machine can learn how the nine looks like from every angle, and with every colour, let's say the machine learns the pattern of the nines, and this is called training.

The second thing that the teacher does is to test his student, to know if he needs to train him more or not, by showing him numbers and letting him answer if the number is a nine or not, and the programmer does the same thing with the machine to score the performance and the accuracy of the machine learning model, and this is called training the model. when the answers of the student are not quite accurate, the teacher should teach him more or change the method of teaching, for the programmer he should decide either train the model more by feeding it more data (pictures of nines), or change the algorithm he used, and do everything all over again.

And this is what machine learning actually means.

WHY DO WE NEED MACHINE LEARNING? AND WHEN DOES IT BECOME HANDY?

Machine learning is usually used in the field of data science, and through this tool, the data scientist can answer many questions like:

- Is this A, B or C: For that, we use classification algorithms (EXP: Does this Spotify user like hip-hop, classic, jazz or pop ?).
- Is this weird or normal: Anomaly detection algorithms(credit card fraud: all client transactions are saved, and one day there is an unusual transaction amount, is it normal ?).
- How much or how many: Regression algorithms (if we have a house with 7 rooms, in a small town, near to the center of the town, how much would be the price ?).
- How is this organized: Clustering algorithms (we ask the machine to regroup the elements of the dataset that are similar).
- What should I do next: Reinforcement learning (when you play chess versus the computer, it takes a decision based on the situation, those decisions are taken using this type of algorithms).

MACHINE LEARNING MAJOR TYPES

- Supervised learning: and we call it supervised because it needs examples of inputs and outputs to be trained, let me break it down for you, and there is no better way for that than an example. Let's say you are a Spotify data scientist, and you want the clients to see their favourite styles in recommended field. What you should do is pretty simple, you already have data with the favourite music type based on the age of the clients (the dataset has inputs and outputs), and what you want to do, is to create a model that can predict the favourite music type of the client based on his age. To do that, you need to train the model with the data you have, and the model will learn the patterns (for example, the model will learn that most of the people under the age of 20 like hip-hop, and for those who are older than 50, they like classic. And that's because of in the dataset we trained the model with, most people younger than 20 like hip-hop, and most people older than 50 like classic, does it make sense now?). NB: "I must inform you that this is not how they do it in Spotify, the classification they do is based on so many factors, not just the age, so this is just an example". And that's why it's called supervised learning.
- Unsupervised learning: It is the opposite of supervised learning. This type of algorithms does a classification, but it doesn't use examples to learn the patterns. In this case, the machine finds the patterns on its own and then does the classification.
- Reinforcement learning: It is used when we want the machine to make decisions, it's not very useful in the data science field, especially as a beginner, so I'm not going to cover this type of algorithms in this course.

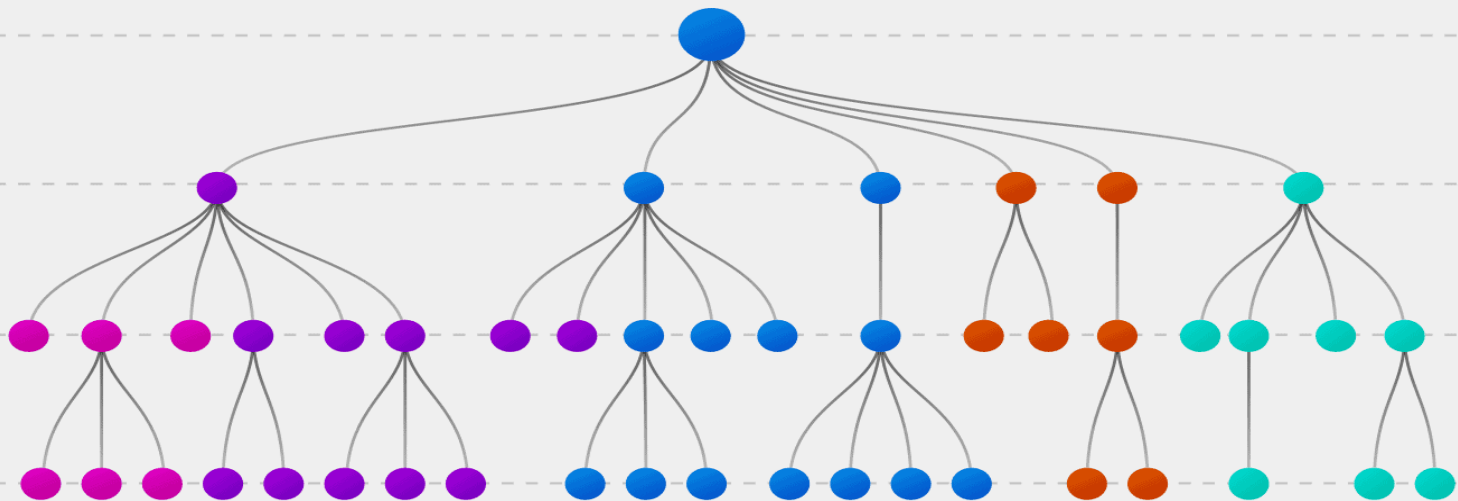
IN THIS COURSE WE WILL SEE

- Supervised learning:
 - Classification:
 - Decision tree (we use it to do a classification with many possibilities)
 - Logistic regression (it is a regression algorithm, but it is used to perform a classification)
 - Simple Linear regression
 - Multiple Linear regression
- Unsupervised learning:
 - K-means

Every algorithm will be explained with a use case, isn't it exciting? Let's Go !!!

As I told you, in this course, you will understand the basics of machine learning, and sometimes I'll be linking some videos that I found pretty interesting.

DECISION TREE



DECISION TREE

The logic behind the decision tree is pretty simple. As you saw in the precedent slide, the decision tree works in levels, every level is for one column of the dataset. So let's get back to the Spotify example, you want to predict the favourite music style for a customer based on his age and his gender. When we use a decision tree to do the classification, and the first level will be about gender, two branches will be coming out, one for male and another for female. The second level is about age, one way to do it is to use four branches, the first is for the people under 20 yo, the second for the people between 20 yo and 40 yo and so on. The last level is the output we are looking for, we should calculate the probability of each branch liking each music type (exp: the first branch is, the first level gives us male and the second level gives us under 20 yo, so now we are working on the people with the characteristics male and under 20 yo, now we should answer the questions like, calculate the probability of a male under 20 yo liking hip-hop, and do it for every music type, then write the type with the highest probability in the output field).

This is the logic behind the decision tree algorithm, and this the job it does, so all we have to do is give this algorithm a dataset to help it learn the patterns and make the decision tree (train the model), and then the algorithm will do everything on its own.

Hope you are now ready to see the code, in the example I worked on I used the same dataset for the same purpose I mentioned in the example before.

Go see the decision tree python file.

SIMPLE LINEAR REGRESSION

- The regression is a mathematical topic, so to understand the code you should first of all understand the mathematical aspect of this subject.
- That's why I will link a video through which you can understand the basics of linear regression. And by the end of that video, you will be able to understand the code.
- The video link:
<https://www.youtube.com/watch?v=iAgYLRy7e20>
- This video will be very useful, but I suggest you watch the whole playlist, it will take a while but it's worth it.

SIMPLE LINEAR REGRESSION

- After understanding the mathematics in this topic, now it's time to do some code.
- For that you can go to the linear regression Python file.

MULTIPLE LINEAR REGRESSION

- When you understand simple linear regression, the multiple linear regression becomes an easy topic, because it's nothing but a linear regression based on multiple independent variables.
- Same as simple linear regression, to understand the basics I invite you to watch this video, and watching the whole playlist will be very beneficial.
- The video link:
https://www.youtube.com/watch?v=dQNpSa-bq4M&list=PLleGtxpvyG-lqjoU8liF0Yu1WtxNq_4z-
- When you finish the video you will be ready to do some coding, for that you can go to the linear regression python file.

LOGISTIC REGRESSION

- The logistic regression is not only used in binary classification (when we have to classify elements into two categories), but in this course I will cover only binary classification.
- To understand the basics of logistic regression, I suggest you watch these videos, the first is mathematics, and the second is code, but at the end of the second video there is an exercise, for what you will find the solution in the logistic regression python file.
- Basics video:
https://www.youtube.com/watch?v=NmjT1_nClzg
- Code video:
<https://www.youtube.com/watch?v=zM4VZR0px8E&t=189s>

K-MEANS

- To understand this algorithm, you won't find a better place than this video:
<https://www.youtube.com/watch?v=EItIUEPClzM>
- In the end of this video, there is an exercise, you can find the solution in the K-means python file.

CONCLUSION

- Machine learning is a journey, and you just started learning the basics and the tools to take this trip.
- I hope you found this course interesting and useful.

GOOD LUCK GUYS