Machine Learning From scratch

Mojtaba Maleki

ISBN:

DEDICATION

I will never be able to find the words to thank you for all you have done for me as a sister, and for that, I must apologize, and thank you each day from now until forever for having you in my life. I want to dedicate this book to you to thank you for being an amazing sister.

CONTENTS

1 INTRODUCTION

“Machine Learning from scratch” illustrates the pathway of becoming a professional and expounds on how machine learning works at the same time. In this book, the author covered most of the important subjects such as predictive data analytics, models, big data, cardinality, ID3 algorithm, etc. In a beginner-friendly way.

This book contains fourteen chapters which are gathered like a great pathway that each chapter has a great summary of Machine Learning fundamental topics in the way of teaching the main idea of them. This combination makes a visible difference to other similar books.

The other positive element is all the chapters have taught as simple as possible for a beginner to understand. After reading this book, the audients can follow the pathway and read other mentioned topics to become professional in Machine Learning.

2 WHAT IS MACHINE LEARNING?

Machine learning is a part of artificial intelligence. If we expect Artificial intelligence a big circle, machine learning will be a smaller circle into that.

Machine learning is used to make computers seem intelligent. It is a bunch of codes with a high amount of data related to the exact subject which needs to be used. The duty of machine learning (ML) is to predict the future based on the data that we get.

The reason which makes Machine Learning popular is we have a lot of available data since social-network become common and useful. For example, when a person buys food for his dog online, machine learning will understand this person has a dog. So, for the next time, it will show him some other pet-related products. In this case, customer action is data, and it is having dog and machine learning use that data and shows him other related products.

There are different types of machine learning, but all of them do the same things which are predicting the future. As you know, by predicting the future we do not mean that predicting what the world will look like in 2050 or some other things like that. In Machine Learning future means the near future.

Machine learning is based on predictive data analytics and predictive data analytics is based on data analytics. Data analytics is data contain charts, graphs, and other mathematics tools for presenting datasets. On the other word, machine learning uses dataset by mathematical knowledge and make the prediction.

This theory makes sense if look how we (as humans) making our decisions. For example, when we want to detect whether a photo is a horse or not, we have data which is our experience, and by seeing those pictures, we measure all of the visible options. For example, if it is red, it is not a horse (we know there is no red horse) or, for example, we see the shapes, and bypassing these steps, we decide if it is a horse or not.

We tell machine learning to goes into our data that even can be terabytes of data and tell that to come up with something useful.

Many companies use machine learning to inform their products that might be useful for their customers.

As a user of social media, you can easily understand that the media always show you the things which you might interested in and they do so by machine learning.

Your data is valuable for those companies. Usually, companies share their datasets (which your data is in that) to develop their database.

3 PREDICTIVE DATA ANALYTICS

As stated in the last chapter, Machine learning is, teaching a machine to learn from data. Data analytics is applied from the dataset to measure the dataset much better than before and use it in the way of our needs.

Data set is a set of historical data that has been given from people's activities on the social network. This dataset is not reality, it just represents reality. For example, one person has a cat, he did not ever search about a cat on the internet, so the internet has no idea if he had the cat.

The dataset has been complete from your data and other people who have high similarities to you in those specific topics.

It is the reason which opening social media shows you what you like. Machine learning tries to guess your taste and give you the best result.

For example, when a person searches for a movie on the internet, he can see the pictures of them on his social media. The reason for that is the machine learning knows the people who searched that movie before you do. They have searched for the pictures too.

The impact of this type of Machine learning is showing suggestion products, pictures that you might like, and this makes that application, website, or other things like that much more efficient.

For example, one data set is 95 percent represent reality, the machine learning job is, predicting those 5 percent missed data based on the data set. That 5 percent of data that has been predicted is predictive data analytics.

We, as humans, use predictive data analytics for making our daily decisions. As an example, when a person goes shopping, he knows what he has and what he needed so this is his historical dataset, and by seeing products, he chooses what he should buy to use in the future. As you can see, there is no data about the future, that person by looking at the past (historical dataset) and seeing other data about himself will predict the future.

Here may ask why we use machine learning when we can guess the future? The answer is machine learning has thousands of eyes on our data and it would not miss any little detail. On the other hand, it is impossible to hire a lot of employees to analyze people and get good results to compare to Machine learning.

Other usages of predictive data analytics are in airports, hotels, malls, and other areas like that. In the airport, they use past data for the number of travelers to measuring the prices. In hotels, they use that to measuring prices as well and even they use that to hiring more employees at crowded times of the year. In malls and shops, they use this to guess people's attraction to their different products at different times, for example, they bring a lot of charismas trees in charismas because based on their dataset they will sell a lot of them in that time compared to other time of the year.

All in all, predictive data analytics is transforming data into future insights to make our suggestions nearer to people's attraction.

4 INTRODUCTION TO MODEL

The last two chapters covered why and where we use machine learning. This chapter contains an introduction to models. The previous chapter illustrated that a dataset is a high amount of data that represents reality. The more data the better result we get. Dataset is modeling reality, so Model is just a representation of reality.

Sometimes, one data is not enough to make a good model. In this case, we need to use several data set together to get the expected result.

In the dataset, we split our dataset into different small sections. Each one of these sections can represent a part of reality.

For example, we have a dataset that represents people's cancer history. In this dataset, there are some data about people's sex, age, their family cancer history, and the result if they have cancer (this is a super simplified example because the point is the teaching model in this chapter).

Here is our dataset:

	A	B	C	D
1	Sex	older than 50	Family history of cancer	Cancer
2	M	Y	Y	N
3	M	Y	N	Y
4	M	N	Y	N
5	M	N	N	Y
6	F	Y	Y	Y
7	F	Y	N	Y
8	F	N	Y	Y
9	F	N	N	Y

The first person, is older than 50, he has a family history of cancer, but he does not have cancer. Or in row six, she is older than 50, she has no family history, but she has cancer.

We can easily make this dataset into binary code. Since the answer to whether a person has cancer is a yes or no.

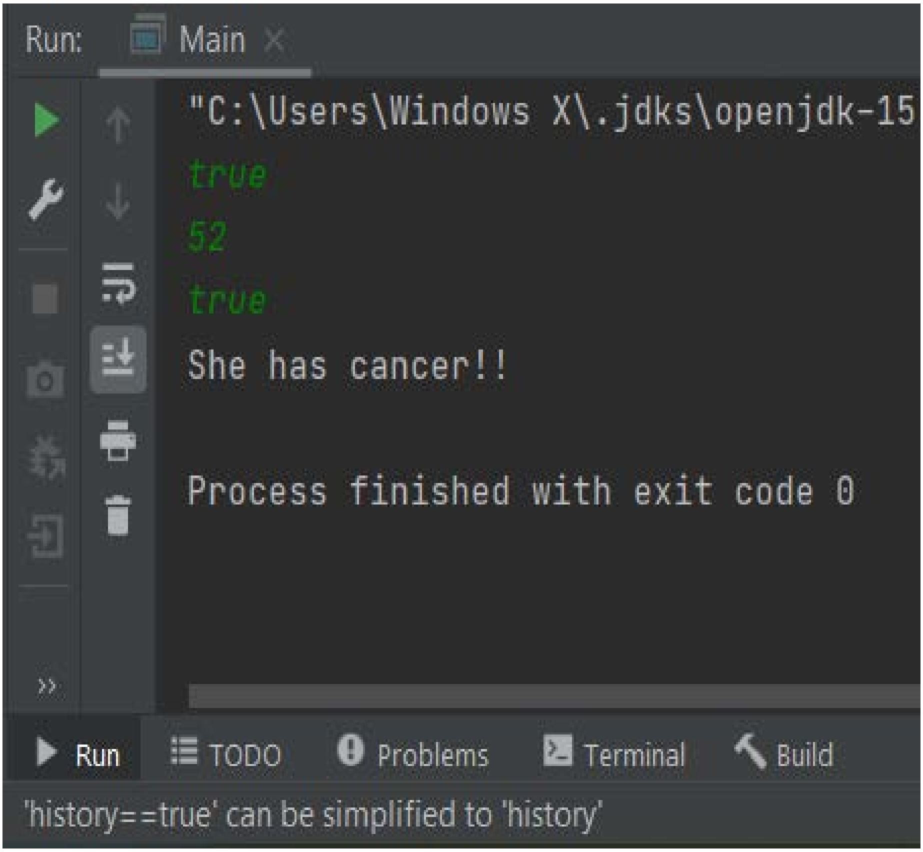
So the point is, we can simply write them as some if statements.

The example of this code in java looks like this (this is not a complete code and does not work if we get other inputs):

```java
package com.company;
import java.util.Scanner;

public class Main {

    public static void main(String[] args) {
        // Cancer Example

        Scanner scanner = new Scanner(System.in);
        boolean sex = scanner.nextBoolean();
        int age = scanner.nextInt();
        boolean history=scanner.nextBoolean();

        //we say if sex was female this is true (it is just initialisation) :)
        //and also we say if history is true, means the person have cancer.

        if(sex==true && age>=50 && history==true){
            System.out.println("She has cancer!!");
        }

    }
}
```

As you can see, this is exactly the data of the row number six in our data set, so we expect that if we run this code get “She has cancer!!”

Here is the output of our code:



As you can see, we can simply illustrate our data with an "if statement". To completing this code, you need to write other databases with the "else if" statement.

Now, based on our dataset, if a person has an exact situation, we can say whether he or she has cancer or not. The answer to this question (here "She has cancer") is a model.

That was an easy part in which we had all of the possible results. Imagine that there is no data about a person who is female, over 50, and no family history of cancer. In this case, Machine Learning becomes useful and it explores all possible data and predicts the model.

In this case, there will be only four possible situations for each sex which are:

1. Female - not less than 50 - have family history
2. Female – not less than 50 – no family history
3. Female _ less than 50 – have family history
4. Female – not less than 50 – no family history

Because they are binary, you can easily illustrate them by yes or no, true or false. The reason why they are four is we already have the other answers, and we are looking for our missing data to guess them based on the mathematics there are four possible answers. Because it is so hard to choose which of the possible answers closer to reality are, we get them to the machine learning to do this for us.

5 ALGORITHM

Algorithm is a piece of code that is used to solve a specific problem. Primarily tests, data structure-related algorithms, graph algorithm (like Facebook), etc. are some examples of the algorithm.

They ate two frequently used algorithm types which are iterative and recursive.

The iterative algorithm is the simple code that all fresh programmers use. The attribute of the iterative algorithm is describing every single step exactly one by one after each other.

The recursive algorithm, on the other hand, is the algorithm that uses itself inside its code. It is exactly like using a code that has not been written before.

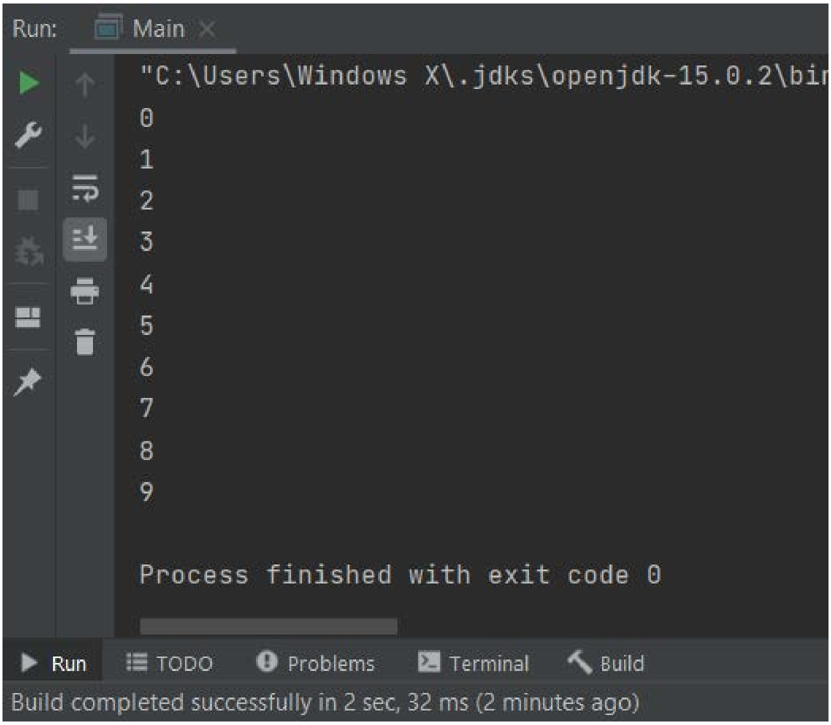
Example of iterative algorithm:

```java
package com.company;
import java.util.Scanner;

public class Main {

    public static void main(String[] args) {

        int a=10;
        if(a>10){
            System.out.println("X");
        }else{
            for(int i=0;i<10;i++){
                System.out.println(i);
            }
        }


    }
}
```

As you can see, in the first step we simply wrote “a” equal to 10. Then, in the next line, we wrote an if statement and asked it if a>10, print “X”. And in else statement, we bring a for loop and initialized i=0 and we asked for the loop to continue increasing “i ” by one until 10. In the next step, we asked to print the process.

In input, because “a” equals to 10, for loop will work. We consider that get numbers from 1 to 9 each in one line.

Here is the output:



As you can see, the code run without any problem.

Recursive algorithm:

A recursive algorithm calls itself which usually passes the return value as a parameter to the algorithm again. This parameter is the input while the return value is the output.

For better understanding look at picture below:

Recursive algorithm return to the first step after finishing whole process.

Here is Fibonacci sequence in both iterative and recursive algorithms ways.

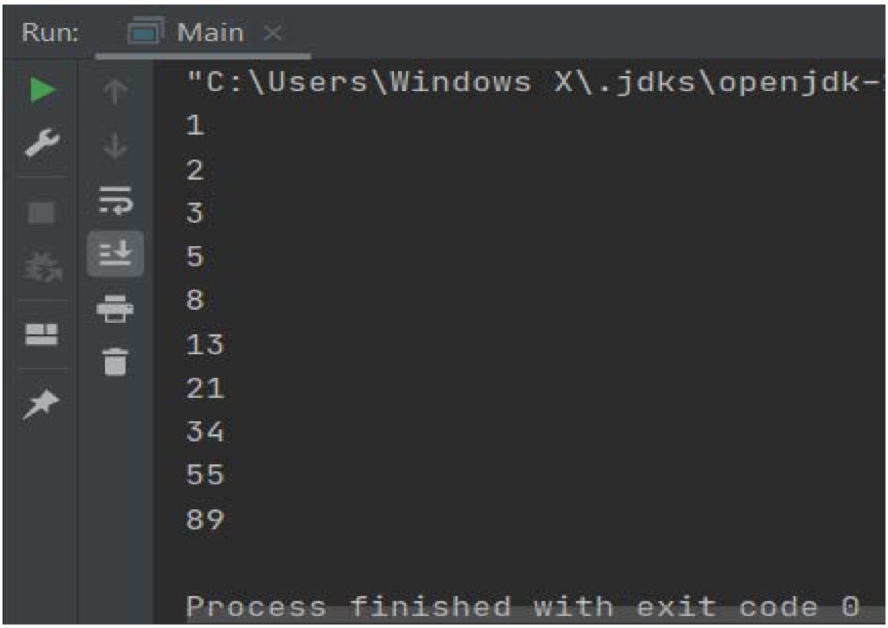
Iterative algorithm:

```java
public class Main {

    public static void main(String[] args) {

        int a=0;
        int b=1;
        int c;
        int n=0;
        while(n<10){
            c=a+b;

            System.out.println(c);
        a=b;
        b=c;
        n++;
        }
    }
}
```

In line 6 and 7 we initialized “a” and “b” as zero and one respectively, in next line we initialized “c”.

Then by while loop, we wrote our code to be run for first 10 Fibonacci sequence.

In output, we consider to get Fibonacci sequence. Here is our output:

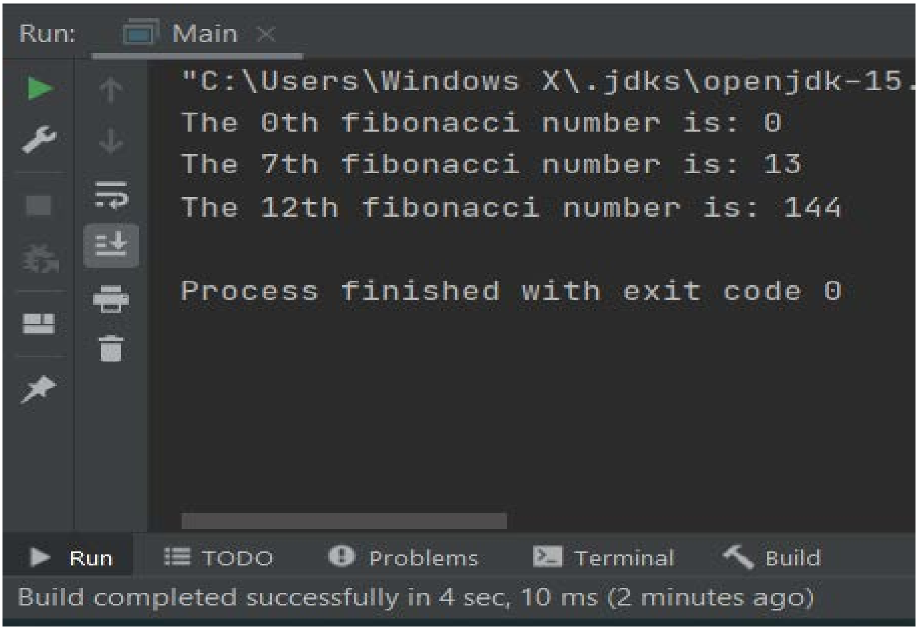


Here is the recursive way writing Fibonacci sequence:

```java
package com.company;
import java.util.Scanner;

public class Main {

    public static long fib(long n) {

        if ((n == 0) || (n == 1))
            return n;
        else
            return fib(n - 1) + fib(n - 2);
    }
    public static void main(String[] args) {
        System.out.println("The 0th fibonacci number is: " + fib(0));
        System.out.println("The 7th fibonacci number is: " + fib(7));
        System.out.println("The 12th fibonacci number is: " + fib(12));
        }

    }
```

We expect that to get the 0^{th} and 7^{th} and 12^{th} of the Fibonacci sequence in our output.

The output is:



Advantages and disadvantages of iterative and recursive algorithm:

Iterative algorithm advantages:

- Faster than their recursive counter parts, since for recursion we usually perform computations more than once.

- Beginner-friendly, since we tell every step exactly how it should be performed.

Iterative algorithm disadvantages:

- Code is usually much longer than for a recursive one.
- Some algorithms are very hard to “translate” to iterative version.

Recursive algorithm advantages:

- Codes usually very short.
- After some practice it is much easier to understand a recursive algorithm than an iterative one.

Recursive algorithm disadvantages:

- Usually slower than the iterative counterpart.
- Without practice hard to understand.

6 ALGORITHM IN MACHINE LEARNING

Last chapter, include some basic information about algorithm. This chapter contain the relationship between algorithm and machine learning, and we want to know how the machine learning decide which is the nearest answer compare to other possible answers.

Machine learning use algorithm to learn how to predict better and better.

Let us get back to our excel table:

	A	B	C	D
1	Sex	older than 50	Family history of cancer	Cancer
2	M	Y	Y	Y
3	M	Y	N	N
4	M	N	Y	Y
5	M	N	N	N
6	F	Y	Y	Y
7	F	Y	N	N
8	F	N	Y	?
9	F	N	N	?

There are a few changes in the table for better understanding.

Machine learning uses an algorithm to choose the closest prediction.

In this table, there is no data about women who are under 50 and have a family history of cancer, and another woman again under 50 and no family history of cancer.

The best way to choose the nearest prediction is to find the pattern between our data. By going through our dataset, we can see a pattern that People who have a family history of cancer, have gotten cancer in their life. It is the best pattern in this dataset.

Based on this pattern, we can say that the women who are under 50 and has a family history of cancer, will get cancer either. On the other hand, the next woman who is also under 50 but does not have a family history of cancer, will not get cancer.

Finding the best pattern is what machine learning algorithms do. For this example, in the real world, we cannot say a person will get cancer or not because our data set is not complete. For getting the closest result, we need much more data to feed our dataset.

Something which may be a question is, it could be exactly one person with the same attribute as row 9 (for example), but she did get cancer. That is true, the machine learning algorithm will analyze the dataset and see, for example, 95 percent of people with the same attribute got cancer, so it round the number to 100 percent because the most majority of people in that situation got cancer.

For other fields like predicting whether an email is spam or not, machine learning algorithm checks if it contains file, does it include any name, is that from out of the country, etc, it can predict that if it has all of those options, it is one hundred percent spam.

But for cancer situations, Machine Learning cannot guess for hundred percent. In this case, machine learning tries to guess the closest prediction, and by getting more and more data in the future, it would develop itself, and that is the bright side of it.

7 TECHNOLOGIES BEHIND MACHINE LEARNING

This chapter includes tools and technologies related to machine learning like Big data and Federation.

Big data:

Big data is a massive dataset that combines whole data from all datasets, spreadsheets, etc in one location that you can run analytics on it (this is equal to the dataset which was in previous chapters).

Usually, in the way of working with big data, you are going to use a data warehouse as well. Data warehouse is a tool that is like a database, but it has been designed for analytical processing.

There are two types of processing, transactional and analytical. Transactional is suitable for places like a web store, where you are going to have a lot of deleting and inserting, etc. But analytical is more design for select statements.

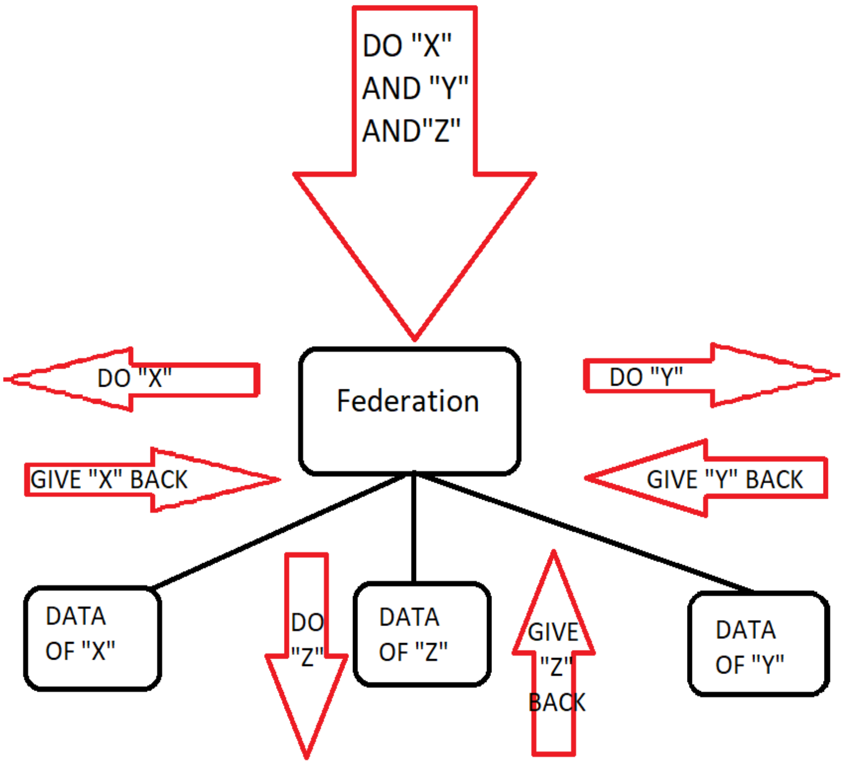
Data warehouse is good for analytical processing, but some data system has both analytical and transitional together as well.

When you have a database powering a website that is going to be optimized for transactional, and then extracts that data and put it into a data warehouse which is optimized for analytical. That’s why we are going to use processing to run analytics on that data.

The process of transforming data in Big data from different types of format to one special format is ETL (extract, transfer, load) process. First of all, we extract the data from the dataset, then transforming it to the right format then load it to the data warehouse.

Federation:

Federation is not a tool or equipment. It is mostly like a trick. For example, many databases have to be in Machine Learning, the problem would be the misconnection between different types of databases. There should be a situation that acts like instead of a huge majority of databases, there is only one.



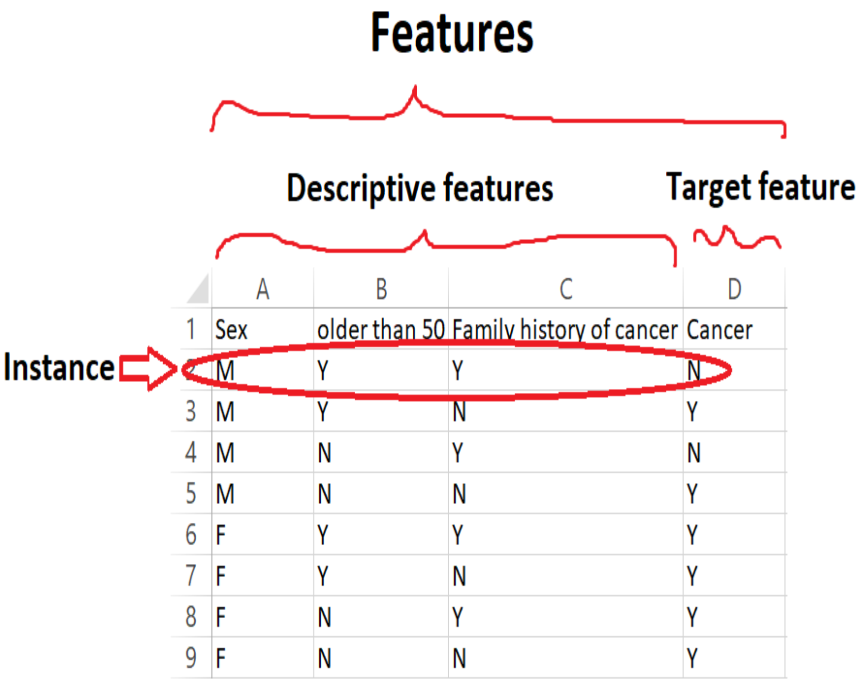
As you can see from the sample, three different inputs are in the three different datasets. Federation is kind of matching them together and those three databases now acting like one dataset.

8 ANALYTICAL BASE TABLE

Analytical Base Table or ABT means all of the data in one big table with a lot of calms and rows.

There is some special name which has been given to the calm, row, etc to clarify the meaning better and better.

The table below is a data set about cancer and also it is a great example for this chapter.



	A	B	C	D
1	Sex	older than 50	Family history of cancer	Cancer
2	M	Y	Y	N
3	M	Y	N	Y
4	M	N	Y	N
5	M	N	N	Y
6	F	Y	Y	Y
7	F	Y	N	Y
8	F	N	Y	Y
9	F	N	N	Y

The headings which carry different type of data (in this case, sex, older than 50, family history and cancer) are known as features. In some Science books, it has been named as an attribute either. There are two types of features, Descriptive features, and Target features. There is only one Target feature but there is no limit for Descriptive features.

Descriptive Features:

They are the features that all have different types of categories for our dataset.

Target feature:

There is only one target feature which is the data that we want to be predicted by our model. This is the main part of our dataset and the data set would not have value if the Target feature has not been available. So, the Target feature is the major part of our dataset.

The other element is instance. Instance is the predictive subject on which we want to predict it. In this case, person or patient is our prediction subject (the examples are rows two to nine).

All of these names have been invented to make talking about technical Machine Learning problems shorter and easier. As an example, instead of saying the headings that contain the categories, it is easier to say the Descriptive features.

9 CARDINALITY

Cardinality is a total option for a particular feature. There are two types of options, one of them is all of the options in the dataset and the other one is all of the possible options.

Sex is binary and there are two possible options for that (male and female), but when there is only one sex exist in the dataset, the feature would have the cardinality of one so it is no longer binary.

Let’s get a harder example. Imagine that there is 20 city in one country. When there is no data about 5 of them, the total possible option is 20, but the features would have a cardinality of 15. The cardinality data is something that we care about (because this is the only thing that we have).

On the other hand, there are two types of features, categorical features, and continuous features. What was included until this chapter was sex, age, and it was categorical features because sex is a category.

Predicting a person's annual income from their education, age and location is an example of a continuous feature. The other example is predicting the yield of corn on the farm based on the weather, water, etc. In these two examples, there is no category, the feature continues.

In the continuous feature, the prediction is a range of numbers. For example, when it predicts someone's salary is 10000$, it is not important if it would be 9999$ or 10001$. Machine Learning tries to predict the nearest prediction and 10001$ and 9999$ are near enough to the 10000$ to say that it is doing well.

As you can see, getting a riddle with the continuous feature is very hard. One of the solutions of making this easier is converting continuous features to categorical features. The process of converting continuous data to categorical data known as binning.

10 BINNING

The process of converting continues feature to categorical feature known as binning.

The amount of salary of company employees continues to feature. It is so hard and even impossible to predict a model based on those continuous features.

The person with a 10001$ salary and the other person with a 9999$ salary is in the different features but there are highly similar to each other (we just discuss salary-related subjects right now).

The best thing to do is binning data. Transferring data in different kinds of imaginary bins bypassing some specific filters is known as binning.

For example, people who earn fewer than 5000$ are in bin A, people who earn between 5000$ to 7500$ are in bin B, people who earn between 7500$ to 10000$ are in bin C and people who earn more than 10000$ are in bin D. By doing so, the efficiency of the dataset will be increased and Machine Learning gives us the possible nearest result.

Binning is not something unfamiliar to humans. When we discuss people's level of the economy, for example, we pass binning process in our mind without notice. Let’s write down the process of binning in that example. We say that people who earn more than 10000$ are rich people and people between 5000$ to 10000$ are not rich but they have a good lifestyle and between 1000$ to 5000$ are normal people and fewer than 1000$ is not normal (this is just an example).

This is a kind of binning if a person asked me about the other person's lifestyle who earns 20000$ we say “WOW, He is rich” because, in our binning in our mind, people who earn more than 10000$ are rich.

Most of the Machine Learning attitudes is like a human because we want it to be like humans. So there is no wonder if it has similarities.

11 DATA CLEANING

Cleaning data or normalization is one of the important parts of creating a Machine Learning model. Machine learning needs clean data to guess the nearest prediction.
It is better to start this chapter with an example. Imagine that we want to teach little kids to detect cat pictures. The best thing to do is show them some **Regular** cat photos. So, first thing first, we show them some normal cat photos. Irregular cats are cats with three feet, a cat with one eye or ear, etc. These are irregular photos and since we show them to the kids, they know cats as animals that have three-foot, etc.
We will do the same thing for teaching Machine Learning. Before giving the data to Machine Learning, we have to pass the data cleaning process. Bad data, is the data that has a problem and has to be fixed or removed.
There are two types of bad data, one type of bad data is the data which have some problems in programing or have gotten issue from ETL process and it must be fixed by checking previous steps. Finding and cleaning these bad data are not as challenging as the other type is.
The other type of issue is when there are some problems with data. This is a huge topic and now we are going to enter it.

Here are some types of dirty data:

- Missing data: missing data is one type of dirty data which happened when referring to data which do not exist there. For example, the algorithm is referring to one specific data but it is empty. The name of missing data is NULL. There are some ideas about if more than 60 percent of the data is missing, that feature or instance should be deleted. There is one more solution available for NULLs which is the imputation process. The imputation process means guessing the NULLs by our self. This is a very hard process because you must guess the model which represents reality. It is better to change it a bit not too much.

- Malicious data: there might be some people who want to gain or cheat your system and you don’t know about it.

They are going to make fabricated behavior data to make their product look more attractive.

- Erroneous data: the data which runs without any problem but the results do not make any sense. When this happened it means that something is wrong in coding and it has to be checked and fixed.

- Irrelevant data: data that are relevant to one group of users. For example, data that contain information about one specific place such as Hungary is irrelevant to other groups of people.

- Inconsistent data: this is a huge issue. People in different places have their style of writing address, phone numbers, date, etc. And even one book could have a different name in different countries. You have to match them all.

Having clean data is much important than having a lot of algorithms. Even with a simple algorithm but with clean data, you can gain the best result which you expected. Cleaning data, most of the time, is time-consuming but it has a huge impact on the final result.

12 SUPERVISED AND UNSUPERVISED LEARNING

There are two types of machine learning, supervised and unsupervised. There are some fundamental differences between these two types.

Here is more explanation:

Supervised learning:

Supervised learning is much easier than the other type. In supervised learning, we teach Machine Learning how to detect the result.

An example of supervised learning is detecting spam emails. We teach machine learning to explore email and teach it if it contains pictures, some special words like “Free” or “OFF” and whether the email is from out of the country, so it is spam.

The other example is teaching how to recognize the photos. We give it some example and it will predict the model based on what we gave (data).

Unsupervised learning:

Unsupervised learning is commonly used for transactional data and it is harder than supervised learning. In unsupervised learning, we teach Machine Learning how to teach itself. For example, we gave tons of photos to it and it must figure out the differences and give it back to different groups. Imagine that we gave it a lot of photos of cats and dogs, it should give the photos back as two types.

The other use of unsupervised learning is the recommendation of shopping sites. Imagine that customer A bought bread, milk, and apple, and customer B bought bread, milk, and chips. So for the next customer who clicks on the site, machine learning must recommend bread and milk, because all of the previous customers have been bought them. This is how unsupervised Machine Learning works.

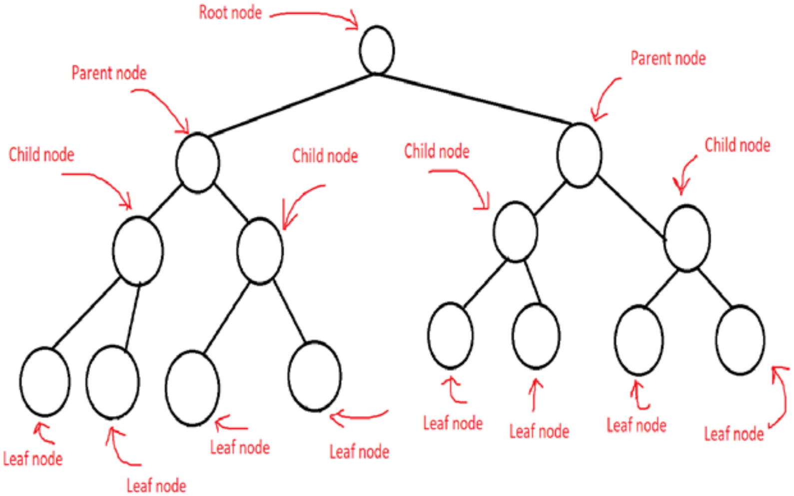
Sometimes, both supervised and unsupervised learning is being used in one model and they call this semi-supervised learning.

13 DATA STRUCTURE TREES

Tree is very useful in the field of computer science. In this chapter, we are going to introduce them and give some examples.

In each tree, there are some circles connected by lines. The name of circles is "node".

Here is a binary tree to explain it for more:

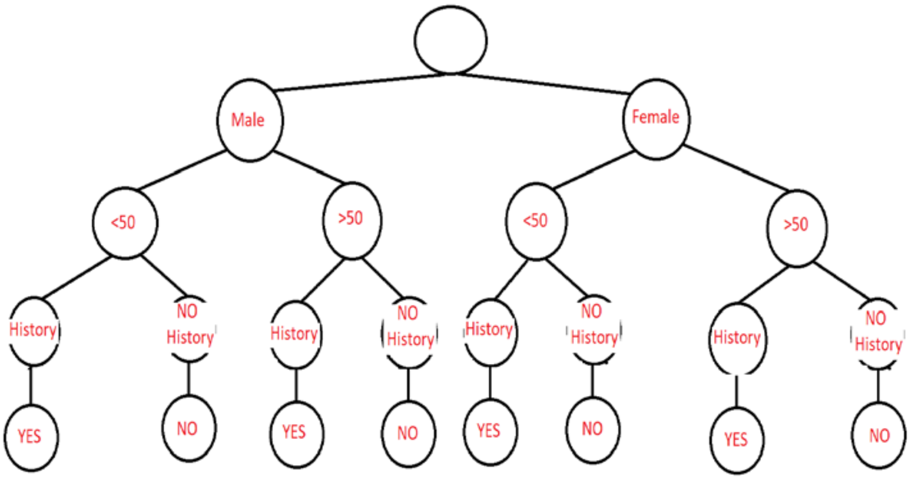


The first node name is “Root node”. And the other nodes which are connected to the root node are the child node, but if the tree is bigger than finishing after one step, they call the next node “parent node” and the node behind the parent node is a child node.

The last nodes are our target features and known as a leaf node.

It is better to move to a working example tree and see the results.

Here is a tree which is related to the cancer that we explained before:



Here you can see that whether there is more cardinality, the lines connected to the circle would increase.

In this case, for each sex, we have just two cardinalities that are under fifty or more than fifty. After age, the other cardinality is similar for all and it is family history. It is not necessary to be all the same and it could be the other features for each age group.

It is nothing to be unfamiliar with until now because we use it every day in our life. For example, imagine that you went to the doctor's office and he will see you and realize your sex, after that, he asks you your age and then ask you whether you have a family history of cancer or not. Overall, he would be able to guess if you get cancer or not.

As you can see, all of the leaf nodes are the target features. So we have

to continue our tree until getting to the target feature. If we don’t, that exact node and other nodes connected to it are not valuable.

So, here is a question, how we got those leaf nodes from?

Based on the historical data from other people, Machine Learning predicts the future result. In this case, most of the people who were male, under fifty, and had no family history of cancer, did not get cancer.

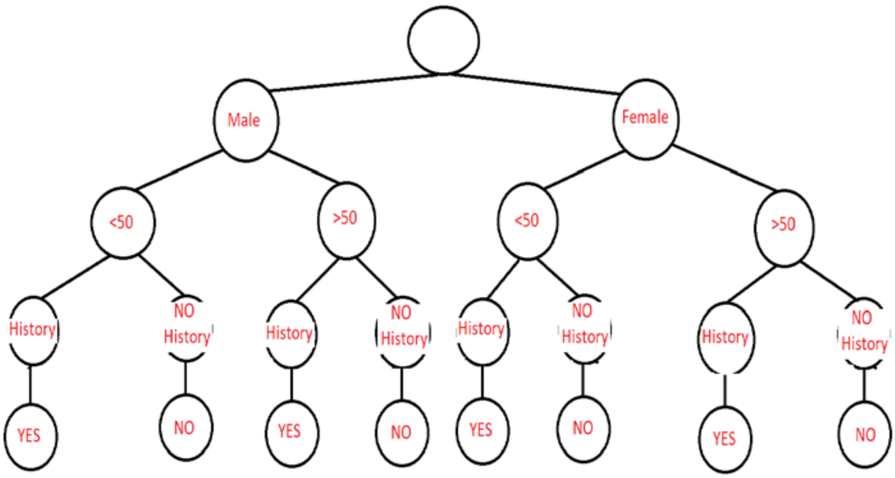
Of course, there were some few people with that exact data got have got cancer but most of the people had not. This is how Machine Learning predicts the model.

14 DISCRIMINATORY POWER

There always been a question which is why we choose sex as our first filter in the model?

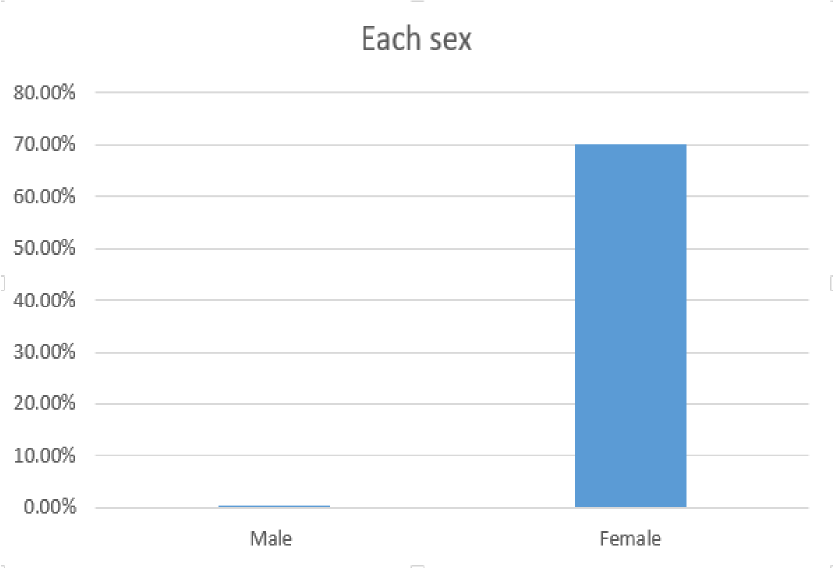
It is not accidental and it has a big reason.

This tree belongs to the breath cancer dataset.



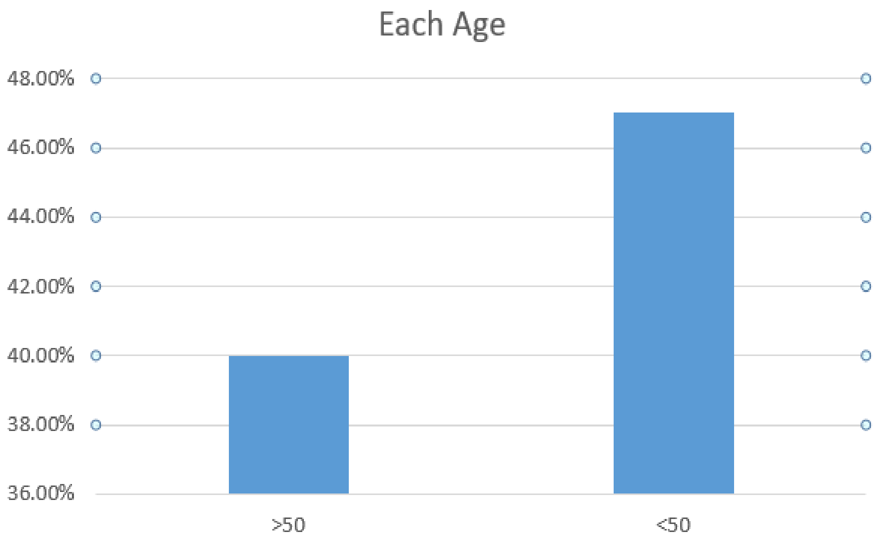
Based on the data which we have, the majority of women who get breath cancer in much more than men.

The chart below is included the percentage of each sex who get cancer.



As you can see, almost 0 percent of men got breath cancer, but women around 70% of them got breath cancer. This is a huge difference that makes it to be the first filter. The higher the discriminatory power the higher effect.

But let us analyze the other chart which is related to the age groups.



As you can see, compare to the last chart, the numbers in this chart are more close to each other.

So, the discriminatory power of this chart is fewer than the last chart. So, the sex will be the beginner filter because it has the highest discriminatory power and the other features will be the next respectively.

15 ID3 ALGORITHM

ID3 Algorithm is a specific Machine Learning Algorithm. It is a recursion algorithm that has three parts base cases, recursive, and other processing parts.

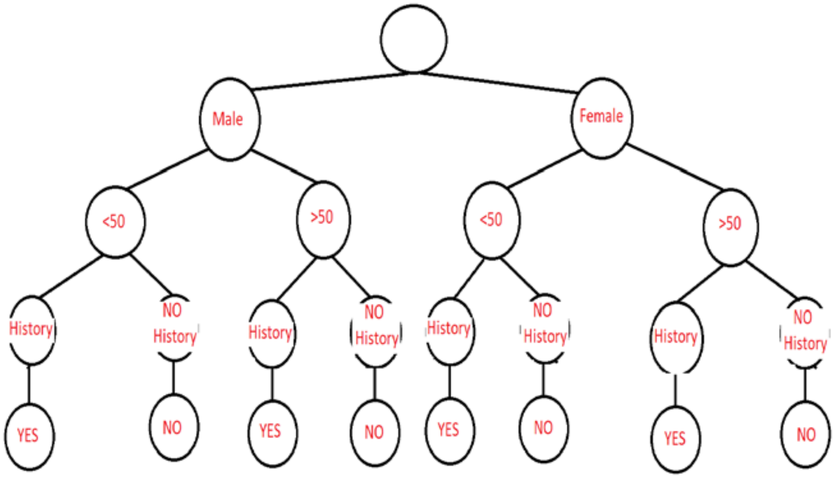
We want to discuss the recursive part first.

The first thing the recursive algorithm would do is choosing the descriptive features based on the discriminatory power.

Based on the example in our last chapter, this would be the sex. After that, it would separate our data into two sections which are male and female. Then it going to remove the sex from the descriptive features list. And it will choose the next most descriptive feature for male or female. And after that, it would do this again and again until arriving at the base case.

Then, it would come back and goes to the places which it had not to enter before. For example, if it had chosen the male at first, after going down the trough to all of them, now it has to into the female and do the same thing.

For the better understanding, I will make an example on the next tree:



First, it chose the most discriminatory power feature. In our example, these are male and female.

Let’s imagine that it chose male and after that, it moves to the next most discriminatory power feature which is age. In this step, imagine I choose under fifty.

Then it would check for the next discriminatory power which is family history, and imagine it choose it has a family history. In this step, it has arrived at the base case.

After this, it would come one step back and goes to the no history, then it will come back and goes to the over fifty, and it would go down and again comes back and do this for all features.

Here are the steps: Male, under fifty, history, no history, over fifty, female, etc.

Now we move to the next topic which is base cases.

The base case is the place which we reach the end and go back up to the next node.

There are three base cases for the ID3 algorithm.

The first of them is when we go through to our features and reach the end, where we arrive at the leaf node and all of the people or whatever we want to predict are the same. For our breath cancer example, the “History” and “no History” are our base cases.

The other type is when it reaches the end but not all of the attitudes have the same value. What happened if we reach our “history” descriptive feature but about 3percent of them have not cancer? In this case, we calculate the majority of the people and select the result based on the average.

The last type of base case happens when it goes to the end but it found nothing in that place. For example, when it reaches the “History” there is no data to shows whether people with this situation are going to have cancer or not. In this case, it goes one step back and makes a prediction based on the average of the last descriptive feature. In our example, when there is no data about the people with family history, it will go one step back to the age, and then based on the average, it makes a prediction.

ABOUT THE AUTHOR

He is a computer science student at the University of Debrecen in Hungary. He is interested in continuing his study on Machine Learning for his master. He decided to share his knowledge by publishing several books related to computer science and machine learning. Meanwhile, he is in a professional area and have a good connection with his professors.

Contact him: mojtabamaleki02.github.io