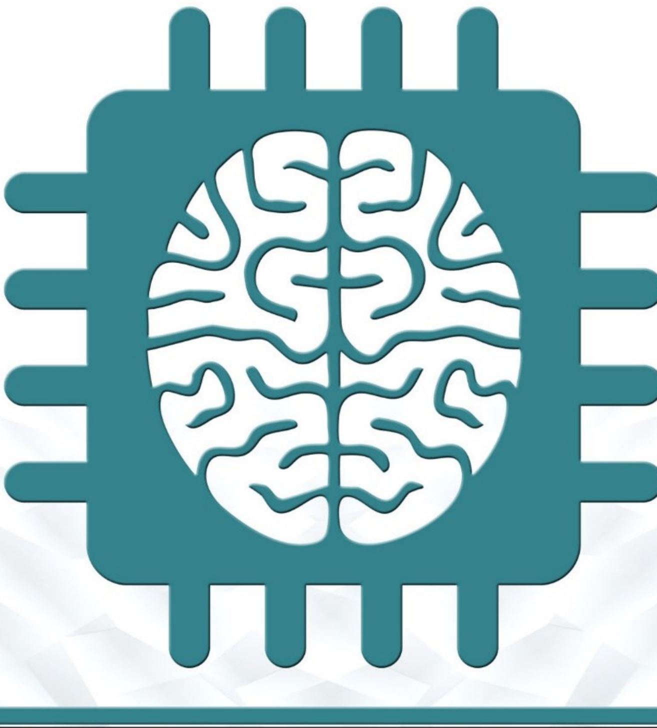
MACHINE LEARNING FOR ABSOLUTE BEGINNERS

INTRODUCTION TO MACHINE LEARNING



ANTHONY S. WILLIAMS

Machine Learning for Absolute Beginners

Introduction to Machine Learning

By Anthony S. Williams

Table of Contents

Introduction

In the recent past there has been renewed interest in the field of machine learning, this is due to the release of huge amounts of data from sensors globally; however, it is not everyone that develops the interest of machine learning understands its basic skill. In the simplest terms, machine learning is a method by which computers deal with large data or information by developing algorithms to derive actionable insights.

Looking at the above definition critically, you will realize that it is based on a few things that will help you get a clue of what the aspect is all about. The main terms that have been included are large data, and algorithms.

This leads us to ask and finding an answer to what an algorithm is. An algorithm is basically a formula or a procedure for solving a particular problem in conducting a sequence of actions towards solving a recurrent problem.

A good example of an algorithm is the computer program that works on a sequence of specified actions. It is now clear that machine learning focuses on the changes that do happen on data patterns, unlike the normal way in which data mining applications do. Machine learning will study the direction of its data and update accordingly.

For instance, machine learning is widely used on social media by updating feeds on what the user has been following in the recent past, the question is, how does the computer detect that you stopped following a certain field and you have resorted to something else?

This is about machine learning algorithms. Basically, machine learning uses statistical analysis and predictive analytics to identify patterns and update news feeds accordingly.

On the other hand, this should be understood in a broad term, assuming that the news feed is data and the updates are the patterns of changing data, machine learning software will detect the relevant information and produce results according to the inputs of data.

Before we deal with the real basic tips for a beginners understanding, here is a question that will simplify the understanding of machine learning, have you ever thought of how Google handles a more than one request at the same time with the greatest accuracy?

There are no people behind Google computers to be attending to all the searches in the world but the systems are able to provide and deliver answers according to the searches and different commands.

So, how is this possible? It is possible because of machine learning where artificial intelligence is applied.

Technically speaking, machine learning is a set of techniques, which helps in handling huge data in an intelligent manner in the development of algorithms or set logical rules towards deliverance of results for users like in the case of Google searches.

There is something else that has been confusing beginners in the field of machine learning it is worth noting that it has been mentioned not only here but in other papers defining the subject as an AI of this generation but how exactly does Artificial Intelligence relate to machine learning, in this context artificial intelligence in regard to computers cannot be mentioned without inclusion of ML.

The two go hand in hand in the understanding of the subject in question. Yes, indeed the two depend on each other for its understanding and yet the working of the machine learning, the logics behind its operations, the technicalities of the calculations involved and the algorithms and the language is all about Artificial Intelligence.

Artificial intelligence can be referred as a type of intelligence that a computer is programmed to learn from actions in the past. A machine (Computer) is tuned to work on preset rules and instructions based on Artificial Intelligence in regard to the past. In essence, ML operates on a subset of Artificial intelligence.

It is obvious, to ask how the computer will sense and learn from the past data experience, this is possible with the use of algorithms such as Naive Bayes (NB), Support Vector Machine(SVM) and the combination of data collected for the delivery of final results.

How do technicians teach machines to use preset algorithms? This is a question that probably you may not be the only one trying to find an answer, many have been trying to figure out what exactly happens to a machine or a computer so as to learn how to group and update data according to user preference.

The most imperative thing to get to know is the subject of data and machine learning, but data in this context has been mentioned vastly and it is time to look at it in the perception of how is it important to machine learning? What does the machine learning concept do to data?

Basically, data and machine learning are two components that are studied together and that is the reason why you will find that the term statistics comes in together with data analysis. Machine learning does the following to data that is input into a computer.

There are five steps that a computer will do so as to manage the data in it and produce the recommended and expected final results. It should be understood that this is also a way in which the data is handled and also performs the tasks or problems. In the following chapters, we are going to discuss what the term "Problem" and "Task" means when it is included in the study of machine learning.

Steps Machine Learning Uses to Perform Tasks/Problems

1. Gathering of data

This is the foundation of future learning and it forms the basis of all the operations on how every type of data is going to be analyzed in the future. The learning prospects of the machine will depend on how the data is presented, in this case, the better the data in terms of volume, variety, and density, regardless whether it is presented in the form of excel sheets, word documents, and text files.

2. Data preparation

The quality of data determines the end results, in machine learning it should be clear and with no data missing so that the best and qualified results can be obtained at the end of the process. Machine learning can be understood as an analytical process and as such, quality is dependent on the data used.

3. Model Training

This level chooses the required model and the algorithms for the purpose of presenting the data in the form of a model. There will be the split of data into two categories, trained part of the data and the second is test data, where the first one is majorly for the model development and the later will act as a reference. In this context we are left to find out what a model is, this should not be a worry because the importance of modeling to machine learning will be discussed in the following chapters.

Improving the data and evaluation of the data are the fourth and the fifth steps, testing of the accuracy of the data are the key activities that are performed at this stage and the most important tools and concepts that are applied here are the test and the categories that were applicable at the stage of training the model.

The choice of algorithm is based on the type of outcome expected. The improvement will also be dealt with when there is a need for changing the models and this is after the performance is verified and be able to evaluate the performance. You could be wondering how the above steps play a vital role in the outcome of data analysis; it is the basis of machine learning.

Chapter 1 Clustering

This is one of the most important problems in unsupervised learning. To explain it, let's start simple. Imagine a night view of the US West Coast taken from a satellite in the sky, suitably placed, so it sees the distribution of lights over the entire West Coast. You've probably seen something like this on a plane about to land late evening. This was probably at the scale of a city. Just scale up your imagination to the entire West Coast. From this view, i.e. the distribution of the "points of light" we would like to discover well-lit clusters. One would hope to discover (the geographic locations of) cities this way. The Figure below is a good example of how a cluster is formed.

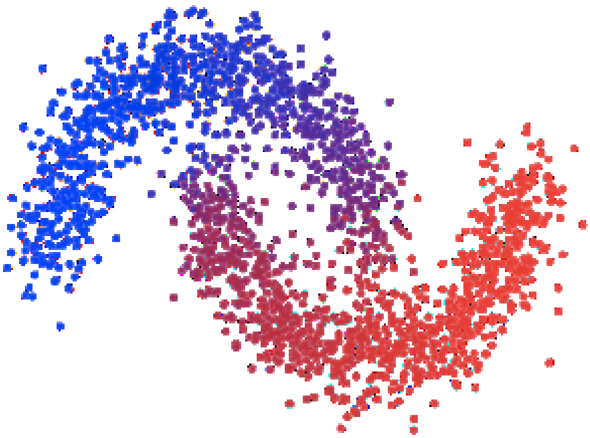


Figure: A Cluster

Let's now define clustering a little more formally. We are given data set D. The aim is to partition D into sets, called clusters, such that items in the same cluster are more similar to each other than items in different clusters.

In the example of the previous paragraph, D is the set of (x,y) coordinates of the "points of light." The similarity measure would be Euclidean distance.

Consider another example. This one is concrete, with numbers. Let $D = \{1,100,2,8,9,7,10,102\}$. Use

$|x-y|$ as the distance measure. A reasonable clustering of *D* using this measure is {{1,2}, {7,8,9,10}, {100,102}}. That is, we have three clusters. Each cluster contains numbers close to each other. Numbers in different clusters are relatively farther apart.

K-means Clustering

In this method, the number of clusters is fixed in advance, to *k*. This method is best explained in the setting of clustering a set of points *D* in *n*-dimensional space (*n* could be 1, as in our example of the previous paragraph, or 2, or 3, or larger). Each cluster is represented by the mean of the points in it. The algorithm is below.

Initialization: Assign each point in *D* to exactly one of the clusters, chosen at random.

Loop:

1. (Re)compute the means of all the clusters, from the sets of points currently assigned to them.

2 For each point p in D, (re)assign p to the cluster whose mean is closest to p. **Until** assignments stop changing

Let's illustrate this method on clustering D = {1,100,2,102}. Assume we somehow know to choose the correct k, here k=2. Denote the clusters 0 and 1. The table below depicts the algorithm's evolution.

Cluster 0	Cluster 1	Mean 0	Mean 1
{1,102}	{2,100}	51.5	51
{100,102}	{1,2}	101.5	1.5

Initially, the random cluster assignment is {1,102} → 0 and {2,100} → 1. From this assignment, the cluster means 51.5 and 51 respectively are computed. Next, 1 and 2 are assigned to cluster 1 because they are closer to 51 than to 51.5. The same reasoning explains why 100 and 102 are assigned to cluster 0. As one sees, this is the optimal clustering. The algorithm converged in one iteration in this case!

Hierarchical Clustering

Reconsider the "points of light" example. Let your imagination run wilder. We wonder whether we could find not only clusters that represent cities, but also cluster neighboring cities into metropolitan areas (where appropriate). We'd like to be able to discover, for instance, the Seattle-Tacoma area, the San Francisco Bay area, the Southern California metro area, and so on. Hierarchical clustering aims to do exactly this.

Bottom-Up Hierarchical Clustering

In this approach, also called *agglomerative clustering*, we start with each data point in its cluster. We then iteratively cluster the clusters and keep going until we end up in a single cluster encompassing the entire data set.

Consider the "points of light" example. Each point of light would first be assigned to its cluster. Then neighboring clusters - currently individual points of light near each other - would be merged into a single cluster. Then these clusters would be merged into bigger ones. During this process, one hopes that some clusters come to represent cities, and neighboring cities get merged into clusters representing their metropolitan areas.

Let's now explain "iteratively cluster the clusters" in more detail. In any iteration, we have a current set F of "frontier" clusters. "Frontier" means that none of these clusters is part of a bigger cluster. In this iteration, we find the two clusters A and B in F that are the nearest to each other, breaking ties arbitrarily. Next, we merge A and B into their “parent” C. We then add C to F and delete A and B from it. And then do the next iteration.

Let’s illustrate this process with a numeric example. Let $D = \{1,100,13,10,110,2,115,104\}$. Let’s define the distance between two clusters as the absolute value of the difference of their means. So the distance between {1,3} and {11,13} is |2-12|=10. Initially, we assign each data point to its own cluster. This initial set of clusters also becomes the current frontier. Next, we find the two nearest clusters {1} and {2} and merge them into {1,2}. We add {1,2} to F and delete {1} and {2} from it. The two nearest clusters in this new F are now {10} and {13}. We add {10,13} to F and delete {10} and {13} from it. The two nearest clusters in the new F are now {100} and {104}. We add

{100,104} to F and delete {100} and {104} from it. Therefore, it goes.

The final "tree" of clusters formed after the process terminates is

The Two Children	Parent
{1} and {2}	{1,2}
{10} and {13}	{10,13}
{100} and {104}	{100,104}
{110} and {115}	{110,115}
{1,2} and {10,13}	{1,2,10,13}
{100,104} and {110,115}	{100,104,110,115}
{1,2,10,13} and {100,104,110,115}	*D*

Decision Trees

A decision tree is a "tree of decisions". We will focus on decision trees in which each decision is based on the value of a single attribute of the input. Below is a decision tree for the Boolean operator *x1* AND *x2*. (The reader is invited to turn this pseudo-code into the picture of a tree.)

if not *x1* return false

else if not *x2* return false
else return true

The logic is this way because of our constraint that a decision be based only on the value of a single attribute.

In the supervised learning setting, we are given a training set of (input, output) pairs. The aim is to learn a decision tree that tries to produce the correct output on any suitable input. To illustrate the issues, a different example will work better. Consider the Boolean function *y* = *f*(*x1*,*x2*) = not *x1*. Note that it ignores *x2*. Consider the problem of learning a decision tree that emulates this function, from a training set of (input, output) pairs from the graph of this function. Specifically, we have:

x1	*x2*	*y*
F	F	T
F	T	T
T	F	F
T	T	F

Which of the two attributes - *x1* or *x2* - predicts the output *y* the best? To answer this question, you may only look at the training set. (You do not know the function; it is what you are trying to learn.)

Consider *x2*. *x2* is F in two examples in the training set. *y* is T in one and F in the other. Also, *x2* is T in two examples in the training set. Here also, *y* is T in one and F in the other. So clearly the value of *x2* does not predict the value of *y* at all. Now consider *x1*. *x1* is F in two training set examples. In both of them, *y* is T. *x1* is T in the remaining two training set examples and both of them, *y* is F. So the value of *x1* predicts the value of *y* perfectly.

Now that we have learned that *x1* predicts *y* much better than does *x2*, our first test in the decision tree is the value of *x1*. Moreover, since we have already noted that *x1*'s value predicts *y* perfectly, this is the only test. Therefore, our decision tree is

if *x1* is true return false

else return true

Before we end this example, one loose end remains. How do we formalize the process by which we found that *x1* predicts *y* much better than does $x2$? (We need to do this to have a formal algorithm, i.e. one that can be programmed.) One sensible approach uses conditional probabilities. $P(y|xi=\mathrm{a})$ is the distribution of the values of y for value a of xi. If $P(y|xi=\mathrm{a})$ is concentrated on a certain value b of y, then xi = a predicts y = b strongly. The entropy I(P) of a probability distribution is a measure of its (lack of) concentration. So $\mathrm{I}(P(y|xi=\mathrm{a}))$ is a good measure of the (lack of) predictive power of xi = a. To turn this into a measure of the predictive power of xi - that is, across all values xi can take - we can use

$$\mathrm{PP}(x) = \mathrm{Ex}(-\mathrm{I}(P(y|x)))\ldots\ldots\ldots\ldots\ldots\ldots(\mathrm{DT1})$$

Equation (DT1) is the expected value of $-\mathrm{I}(P(y|x))$, where the expectation is taken over x, i.e. the probability distribution over the values x can take. Consider the case when all values of x are equally likely. In this case, (DT1) is just the average value of $-\mathrm{I}(P(y|x))$ over the different values x can take.

Okay, we have some intuition about what entropy of a distribution is, but what exactly is its formula? We present this below.

The entropy I(P) of a discrete probability distribution P is

$I(P)$ = -sum_v Pv log2 Pv.........(ET)

Here the sum is taken over all values v in P's sample space and Pv is the probability of value v.

To understand (DT1) better, let's apply it to our example.

PP($x2$) = -0.5 I($P(y|x2$=T)) - 0.5 I($P(y|x2$=F)) = -1

This is because $P(y|x2$=T) = $P(y|x2$=F) = {T => 0.5, F => 0.5}. Also I({T => 0.5, F => 0.5}) = 1.

PP($x1$) = -0.5 I(P($y|x1$=T)) - 0.5 I(P($y|x1$=F)) = 0

This is because $P(y|x1$=T) = {F => 1.0} and $P(y|x1$=F) = { T=> 1.0}. Each of these distributions has an entropy of 0. Check this out yourself by evaluating (ET) for each of these P's. (Note that 0 log2 0 → 0.)

Since PP($x1$) > PP($x2$), $x1$ predicts y better than does $x2$.

Linear Regression

This is the problem of learning the linear function y = **mx** + **c** that is the best fit to a sample S = {(**xi**,*yi*), *i*=1,...,*m*}. Here **xi** is the input, and *yi* the desired output. **xi** is a scalar or a vector. In the latter case, **m** is also a vector and **mx** the inner product of two vectors. Note that the learnable parameters here are **m** and **c**.

Let *f***mc**(**x**) = **mx** + **c** denotes a particular instance of a learned function. How good is this function? That is, how well does *f***mc** fit the data set *S*? We check, for each example (*xi*, *yi*) in *S*, how close *f***mc**(**xi**) is to *yi*. We then aggregate the results of all these checks into a single number. While there are many ways to define closeness and also to aggregate the results, the method that is both the most sensible and the most widely used is below

E**mc**(*S*) = sum_{*i*=1,...,*m*} (*f***mc**(*xi*)-*yi*)^2.........(ER1)

Notice that E**mc**(*S*) is always non-negative, with the best possible fit happening when it is 0.

The aim of learning is to find a pair (**m**,**c**) that minimizes E**mc**(S).

Probabilistic Classifiers

These algorithms build probabilistic models that predict the desired output Y from the input X. The most obvious way to start is to frame this as modeling the conditional probability distribution $P(Y|X)$. Think of this as follows. $P(Y|X=\mathrm{x})$ gives a probability distribution over the various values the output variable Y can take for the particular input $X = \mathrm{x}$.

If Y is a categorical variable, i.e. it has a discrete set of values with no ordering on them (e.g. 'male' or 'female' for Y = gender), then

y* = argmax_*y* $P(Y=\mathrm{y}|X=\mathrm{x})$ can be interpreted as the most likely output (class) of x. $P(Y=\mathrm{y}^*|X=\mathrm{x})$ may be interpreted as the confidence with which we can say that y* is indeed *the* class of x. We can also choose to avoid making a decision when $P(Y=\mathrm{y}^*|X=\mathrm{x})$ is not sufficiently large.

We have already learned some important lessons from the previous paragraph.

1. A probabilistic classifier outputs probabilities of the various possible outputs (classes) for the given input.

2. If one needs to output a single class, choose the most probable one for the given input.

3. The probability of the most probable class for the given input may be interpreted as the classifier's confidence in its decision in 2. above.

4. The classifier’s decision-making can be made more or less conservative by choosing a suitable threshold on the minimum confidence a decision must have. (‘confidence’ is as defined in 3. above.)

We explicitly call out these points because most non-probabilistic classifiers do not naturally support these features. In some cases, the "arms" of some such classifiers can be "twisted" to support some of these features heuristically. In many cases, even this is not possible.

Kernel Functions

Consider the following supervised learning problem. We have an unknown function *f* which maps R2 to a discrete set *C*. Think of a point (*x*,*y*) in R2 as a (latitude, longitude) pair denoting a particular location on Earth. Think of *C* as a set of cities, plus a symbol *b* denoting background. Here *f*(*x*,*y*)=*b* means that (*x*,*y*) is not in any city in *C*.

The Basics of Machine Learning and how it works

"The Lazy Way to Mastering the Concepts of Machine Learning"

In the recent past there has been renewed interest in the field of machine learning, this is due to the release of huge amounts of data from sensors globally; however, it is not everyone that develops the interest of machine learning understands its basic skill. In the simplest terms, machine learning is a method by which computers deal with large data or information by developing algorithms to derive actionable insights.

Looking at the above definition critically, you will realize that it is based on a few things that will help you get a clue of what the aspect is all about. The main terms that have been included are large data, and algorithms.

This leads us to ask and finding an answer to what an algorithm is. An algorithm is a formula or a procedure for solving a particular problem in conducting a sequence of actions towards solving a recurrent problem.

A good example of an algorithm is the computer program that works on a sequence of specified actions. It is now clear that machine learning focuses on the changes that do happen on data patterns, unlike the normal way in which data mining applications do. Machine learning will study the direction of its data and update accordingly.

For instance, machine learning is widely used on social media by updating feeds on what the user has been following in the recent past, the question is, how does the computer detect that you stopped following a certain field and you have resorted to something else?

This is about machine learning algorithms. Machine learning uses statistical analysis and predictive analytics to identify patterns, and update news feeds accordingly.

On the other hand, this should be understood in a broad term, assuming that the news feed is data and the updates are the patterns of changing data, machine learning software will detect the relevant information and produce results according to the inputs of data.

Before we deal with the real basic tips for a beginners understanding, here is a question that will simplify the understanding of machine learning, have you ever thought of how Google handles a more than one request at the same time with the greatest accuracy?

There are no people behind Google computers to be attending to all the searches in the world, but the systems can provide and deliver answers according to the searches and different commands.

So, how is this possible? It is possible because of machine learning where artificial intelligence is applied.

Technically speaking, machine learning is a set of techniques that helps in handling huge data in an intelligent manner in the development of algorithms or set logical rules towards deliverance of results for users like in the case of Google searches.

There is something else that has been confusing beginners in the field of machine learning it is worth noting that it has been mentioned not only here but in other papers defining the subject as an AI of this generation but how exactly does Artificial Intelligence relate to machine learning, in this context artificial intelligence in regard to computers cannot be mentioned without inclusion of ML.

The two go hand in hand in the understanding of the subject in question. Yes, indeed the two depend on each other for its understanding and yet the working of the machine learning, the logics behind its operations, the technicalities of the calculations involved and the algorithms and the language is all about Artificial Intelligence.

Artificial intelligence can be referred as a type of intelligence that a computer is programmed to learn from actions in the past. A machine (Computer) is tuned to work on preset rules and instructions based on Artificial Intelligence regarding the past. In essence, ML operates on a subset of Artificial intelligence.

It is obvious, to ask how the computer will sense and learn from the past data experience, this is possible with the use of algorithms such as Naive Bayes (NB), Support Vector Machine(SVM) and the combination of data collected for the delivery of final results.

How do technicians teach machines to use preset algorithms? This is a question that probably you may not be the only one trying to find an answer; many have been trying to figure out what exactly happens to a machine or a computer so as to learn how to group and update data according to user preference.

The most imperative thing to get to know is the subject of data and machine learning, but data in this context has been mentioned vastly, and it is time to look at it in the perception of how is it important to machine learning? What does the machine learning concept do to data?

Data and machine learning are two components that are studied together, and that is the reason why you will find that the term statistics comes in together with data analysis. Machine learning does the following to data that is input into a computer.

There are five steps that a computer will do so as to manage the data in it and produce the recommended and expected final results. It should be understood that this is also a way in which the data is handled and also performs the tasks or problems.

Steps Machine Learning Uses to Perform Tasks/Problems

1. Gathering of data

This is the foundation of future learning, and it forms the basis of all the operations on how every type of data is going to be analyzed in the future. The learning prospects of the machine will depend on how the data is presented, in this case, the better the data regarding volume, variety, and density, regardless whether it is presented in the form of excel sheets, word documents, and text files.

2. Data preparation

The quality of data determines the end results, in machine learning it should be clear and with no data missing so that the best and qualified results can be obtained at the end of the process. Machine learning can be understood as an analytical process, and as such, quality is dependent on the data used.

3. Model Training

This level chooses the required model and the algorithms for the purpose of presenting the data in the form of a model. There will be the split of data into two categories, trained part of the data and the second is test data, where the first one is majorly for the model development, and the later will act as a reference. In this context we are left to find out what a model is, this should not be a worry because the importance of modeling to machine learning will be discussed in the following chapters.

Improving the data and evaluation of the data are the fourth and the fifth steps, testing of the accuracy of the data are the key activities that are performed at this stage and the most important tools and concepts that are applied here are the test and the categories that were applicable at the stage of training the model.

The choice of algorithm is based on the type of outcome expected. The improvement will also be dealt with when there is a need for changing the models, and this is after the performance is verified and be able to evaluate the performance. You could be wondering how the above steps play a vital role in the outcome of data analysis; it is the basis of machine learning.

Chapter 2 Working with Machine Learning

We have spent just a little bit of time talking about machine learning, and it is considered a subfield that is found in artificial intelligence and data science. Machine learning is going to apply to statistical methods to improve performance based on the detected patterns or previous experience. One aspect of machine learning that you should keep in mind is the usage of the self-improving algorithms. Just as humans can learn from some of their previous experiences, or by going through a process of trial and error to make their decisions, machine learning is going to go through some of the same processes.

However, not only will your machine be able to think and learn as humans do, but also they are much more effective. Humans are not predisposed to be proficient and reliable in tasks that are repetitive like computers, especially when it comes to going through large data amounts. Also, the speed, complexity, and size at which big data generated is often going to exceed the capabilities of humans.

Let’s take a look at how this will work by going through the following dataset:

1; [0, 0]

2: [3, 6]

3: {6, 12]

4: [9, 18]

5: [12, ?]

Now this one is a pretty simple example and many times you will be able to figure out the answer on your own. Since the second number in each of these groupings is going to be twice as larger as the first number, you can guess that the answer in group 5 is going to be 24. You most likely do not need the help of a computer to come up with this number. However, what if your rows had much bigger numbers or ones that had decimal points for a few digits. This can make it harder to figure out and process the information to come up with a prediction quickly.

You would be able to do the same process with more complex numbers with the help of a machine. The machine that you work with will take on some of the mundane tasks of working with numerous possibilities to isolate the large segments of data to solve the problem you are working on, while also storing and visualizing the data. Machine learning can be great because it will end up freeing up your time so you can focus on improving the results or working on other parts of the business.

So the next question is how you will program a computer to calculate the answer if you don't even know how to do the calculation yourself? This is something that is really important for machine learning. If you configure these properly, the algorithms for machine learning can learn and recognize the new patterns all on their own. However, the process of machine learning won't start to happen all on its own. As with any machine, a human will need to come into the process to program and then also to supervise the automated process. This is where the machine learning engineers, as well as data scientists, are going to enter.

The role of a data professional is to configure the equipment, which could include the databases, operating systems, and servers, and the architecture, which is how the equipment is going to interact with each other and then add in the programming algorithms using different operations. The programming is so important because it helps you to train the machine on how it should behave in different situations.

Now, working with machine learning means that you will need it to be able to work on some complex tasks. Rather than working to program your machine to respond to something that is a fixed possibility, such as responding to a red light, the data scientist needs to go through and work on the method a bit differently. The engineer will not be able to program the computer to see animals in a picture, for example, based on a human description because this would induce a pretty high rate of failure. For example, if you said that all animals had fur and four legs, options like kangaroos would be missed out on.

Instead, the data scientist needs to work to program their computer to identify animals based on socializing examples, such as you would teach a child. A young child would not be able to recognize a goat based on just the description of the key features. You could say short neck, white fur, and four legs, but this does describe a few different types of animals. It is more effective to show the child what the goat will look like by showing images of the toy goats, real life goats, and images of a goat. Image recognition of your machine learning is going to be the same, but teaching is going to be managed through images and programming language of your choice.

Whether the computer is working to recognize human faces, illicit material animals or something else, the machine can apply examples to help write its programs to provide it with the capability to identify and recognize subjects. This is going to help eliminate the need for the programmer to explain, in lots of detail, the characteristics of each subject and it will help to reduce the amount of failure that happens.

Once the algorithms and the architecture have been configured properly, it is possible for machine learning to take place. The computer will then star to implement the algorithms and the models to classify, predict, and even to cluster data to form their future predictions and to later draw up new insights based on whatever patterns it can see.

Chapter 3 Data Mining

Data mining is another thing that you can learn from when working with machine learning. Data mining is a discipline of data science that will aim to unearth some of the unknown regularities, patterns, and relationships from really large data sets and this can be seen in Figure below. Given that data mining doesn't have to start with a hypothesis, there are many different techniques for data sorting that are applied, including association analysis, sequence analysis, clustering, and text retrieval. A big question that a lot of people are going to have then when it comes to all of this is what is the big difference between machine learning and data mining.

Figure: Data Mining

There is a pretty strong correlation that comes with these two. At the abstract level, both are going to be working with analyzing data and then extracting the valuable insights. There are many times when data mining is going to utilize the same types of algorithms for self-learning that are used in machine learning. Both of these are also going to spend their time focusing on forming future predictions by looking at some of the historical data, and often they are going to use the same kinds of algorithms to form these new predictions. However, while machine learning is going to use these algorithms to improve with experience at the task it is doing, data mining is going to focus on going over the data to see some unknown properties and patterns.

Therefore, data mining is exploratory in nature because it is looking for the knowledge that you don't know, which machine learning is going to spend time concentrating on studying and reproducing specifically known knowledge to form its predictions. Data mining will take large pieces of data sets, cleans up the data, and then will spit out its analysis based on the patterns and the relationships that are found inside of the data, and it likes to work with a lot of big data. It is often done for just one project, or at a set interval rather than doing it all of the time.

On the other hand, working with machine learning is going to focus on incremental as well as ongoing problem solving and it can work on all kinds of different problem solving, no matter what size. When Google is comparing the first and the second search query that you look at, and then makes an inference based on what you are searching for it is going through a process of self-learning to help figure out what you want to work with.

That is one of the nice things about machine learning. It is not going to turn itself off once you finish your task. It is always working to learn, and it will adapt some new inputs of your data, whether that data is big or small, and then it works to store these values to use at a different time. The ability to learn from these experiences over time is what will make machine learning so important and so useful in your work.

Chapter 4 Relationship between Machine Learning and Artificial Intelligence

This is an important aspect in the study of machine learning, especially for the beginners. This part is meant to make you relate to the computer intelligence and the technicalities of the artificial intelligence as it is applied to the computers to understand its environment and how it can be fed with information in the form of data to give results.

Results are to be delivered in the form of data from its analysis of relevant actions.

The two terms are sometimes used interchangeably to mean the same thing, of which in the study of machine learning and the general knowledge pertaining to computer science, the two are different topics altogether. Artificial intelligence and machine learning are far much apart regarding how they work and how they relate to real life situation.

Artificial intelligence has been there before the coming of the machine learning technology, and this can mean that it is more of highly appreciated than artificial intelligence, AI some time with the aim of trying to prove if computers can think like humans and that is how machine learning came into its existence.

In other words, we are trying to see how artificial intelligence can perform tasks just the same as human beings, the question that is being answered in this subject is whether machines can identify images like those of faces and distinguish between objects.

This is not exactly what AI involves but what we want to bring to our attention is how machine learning is closely related to artificial intelligence, as we have concluded that AI was there before ML, it can be concluded that machine learning came into existence as a result of artificial intelligence.

The supporting pieces of evidence of this statement are clear because, some years back, computer scientists started coming up with the machine that can perform tasks just like human beings, the robots, that were being used by autonomous to get information from the universe is a good example of where artificial intelligence has come from.

Artificial intelligence involves creation of computer systems that can do the things that humans can do, in other words, think like human beings, in this context and in regard to the word "think" in the description of artificial intelligence, is to analyze situation in the form of data and make decisions in the form of actions.

To help you understand better the concept of AI and machine learning, how Google self-driven cars work, how do they identify routes and other objects on the road, this is a combination of both machine learning and artificial intelligence. It has been proven that artificial intelligence is no longer a science fiction but a scientific fact.

There has been a wave where leading technological companies are investing hugely in projects related to AI or AI specifically and it is clear, although most of the user do not know that they are interacting with artificial intelligence software almost on a daily basis, this is how the interaction is made real, the Smartphone that we use almost every hour of the day, social media sites, E-commerce sites and search engines.

We cannot rule out that one of the Artificial intelligence types that human interact with frequently is Machine learning. Now you have known that the relationship between AI and ML is the fact that machine learning is one of the types of artificial intelligence.

In giving the difference between the two, we can start from the concept of data mining where the machine learns the patterns and then give information sets in.

For instance, the common application of ML as a type of artificial intelligence is the image recognition where the machine has to be exposed to a lot of images of human beings, plants, animals, nature, and others and then after learning to distinguish various types of data through technical interaction then, it will use the learnt procedure to deliver the results.

Some prefer to use the term "machine learning" for the reason that they think it sounds more technical and a little professional than "artificial intelligence."

From the statement itself, it can be deduced that there is a slight difference between artificial intelligence and machine learning. In the recent past, machine learning has been closely discussed within the concept of artificial intelligence, most of the applications and software are using machine learning and artificial intelligence thereby making it clear that the two have a very small difference come up with the conclusion that one depends on the other.

Having stated the relationship of the two, and made a conclusion that the difference is insignificant, there is another issue that has been a thorn in the flesh of the beginners in the discipline of machine learning and artificial intelligence. The terms that have been used by the experts and scholars during the study of the same include "Deep Learning" and "Neural nets" or artificial neural networks.

The two terms have come up due to research development and it every field of study there should terms that do come up whenever there is an advancement in the study, Machine learning and the general artificial intelligence, is not left out as a field of study in computer science.

Most of the starters in the field of the artificial intelligence have heard difficulties when they come across the terms, this leads to contradictions and confusions; neural nets are systems that are designed to work like brains, and it forms the basis of the deep learning while the deep learning concept specifically represents analysis of data in multiple layers by the use of certain algorithms.

It should be understood for the sake of understanding the Artificial intelligence that AI has been in existence, and due to technological growth in the use of systems and computers, several concepts are invented, and to some point, the basics can bring in other terms that can contradict. It is important to study the field right from the beginning so as to update yourself and avoid being a victim of concepts confusion.

Types of Artificial Intelligence

Since the time artificial intelligence was introduced by researchers, and the fact that it is a dynamic field, there has been technological advancements and more especially in the invention of several algorithms to be used in machine learning. Artificial intelligence leads to the development of machine learning and in the same trend the functions and algorithms used in the field, computer science researchers and scholars have also categorized artificial intelligence into different classes.

The major accelerator of the categorization was driven by the ability of the machine to understand what human can understand, research has also played a major role, and most importantly the thinking that machines can even work among human beings. Besides that, machine learning algorithms have been in the forefront in the recent developments.

The division or classification of artificial intelligence is due to the rhetoric posed by the boundaries between humans and machine.

There are four main types of artificial intelligence as we are going to discuss in details including their characteristics.

Reactive Machines

This is the most basic and common artificial intelligence systems known so far. One trait that has been put across by the experts in the field is that it does not have the ability to learn from the past experience in making its decisions. Blue is a perfect example of this type of machines used in playing chess, one thing to be noted is the ability for it to master the pieces and identify them on the chess board but at the end of the day it cannot remember what happened in the past, meaning that it has mastered each piece move the, but it cannot remember the past moves.

This type of machine relates to its environment and decides the next move or what to do to the world/environment, generally what it perceives. In summary, the computer systems have no ability to master the concepts of a wider world they keep on stagnating when exposed to a different environment.

They only work on a specific task that they are only trained to thereby they can easily be misled. However much, this sounds disadvantageous, it is a way to prove that Artificial intelligence can be trusted in the environment that it is used to, or has ever been exposed to.

Theory of the Mind

This type of artificial intelligence is based on the theory that the mind is a computation that comes from the brain which acts as a machine, in essence it perceives the brain as a computer and the mind as the program that is run by the brain, where the program and in this case the mind is an algorithm which provides and determines the next cause of action for the machine or the computer and specifically, the brain.

The theory of the mind is a computation of a machine, which is the brain. The theory of the mind artificial intelligence is based on the knowledge of human behavior, emotions, agents in the environment and the relationship between the machines with the latter. This theory of the mind that sets the basis of the artificial intelligence tend to bring in that machines have to work like the human brain and understand the feeling and emotions of the agents in the environment.

This is important in the study of machine learning as a component of artificial intelligence since the machines will be able to learn the behavior and the changes in its immediate environment in the improvement of the specified intelligence if after some time we believe that machine will one day behave like the human brain.

Self-Awareness AI

This could be the cream of Artificial Intelligence steps of development building systems that can structure representations about themselves. Eventually, researchers will have to bring in the understanding of consciousness regarding machine learning algorithms. This type of intelligence brings forth the inclusion of instincts and consciousness.

It can be termed as an extension of the "theory of mind." Since consciousness is also called self-awareness" in support of this type of AI. Conscious organisms do understand themselves, and they know about their status and can predict feelings of others. We assume someone shooting behind us in traffic is irritated or impatient because that's how we feel when we hoot at others.

Without the theory of mind, we could not make those sorts of inferences, and in this case, machine learning should be equipped with the algorithms to decipher the same feelings as a human. The focus should be made on efforts toward understanding memory, learning and the ability to base decisions on past experiences.

This is an important step to understanding human intelligence on its own about artificial intelligence.

Limited Memory

This is a representation of a type artificial intelligence, where it can look into the past. For instance, the example that has been stated earlier of the Google self-driven cars that can be able to detect the movement of the other cars, their direction, and possible turn; this takes time to master the trend of each object over a given period.

In machine learning, the algorithms are added as programs in the representation of the world, like the traffic lights and the road lane markings as seen in the Figure below. The past experience is not inclusive of the memory of the machine, but it can learn it with time just like the human drivers who learn to base on the experience they gain from driving cars for a long time.

Figure: Artificial Intelligence

This type of artificial intelligence emphasizes on the possibility of the machine be able to build their memories of the past and use it to make decisions of the present and come up a representation of their own.

Chapter 5 What is a Neural Network?

A neural network is a kind of network in which the nodes are seen as “artificial neurons,” this is well explained by the Figure below. The concept of the neural networks began in the 1980s. The neural network of the human being is made up of a network of interconnected neurons for maintaining a high level of coordination to receive and then transmit messages to the spinal cord and the brain. In machine learning, such types of networks are referred to as “Artificial Neural Networks (ANNs).”

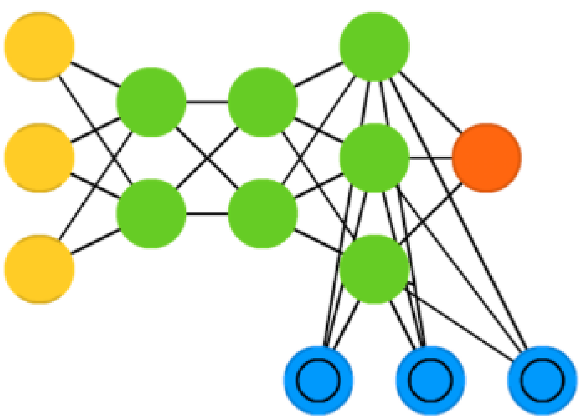


Figure: Neural Networks

The Artificial Neural Networks are made up of “neurons” which have been artificially created.

These are then taught so that they can adapt to the cognitive skills of human beings. Some of the applications of ANNs are image recognition, soft sensors, voice recognition, time series predictions, and anomaly detection.

Learning in Neural Networks

In artificial neural networks, learning refers to the process of modifying the bias and the weight. A perception works by computing the binary function for its input. The perception works by learning concepts. Learning in neural networks is facilitated by training it, whereby a certain set of inputs is fed to the network while expecting a particular output, which is the target output.

This calls for adjusting the values for the weights and the biases so that they can give us the target output. The training technique which is used in this case is referred to as the "perception learning rule." The process of learning in the artificial neural networks is similar to the concept of learning in human beings.

The Learning Rule

The training of the perception is done so that it can respond to the input vector with a target output, which can be either 0 or 1. For you to achieve a solution with the learning rule, a number of trials are needed, and there must be a solution.

Since learning is very central in artificial neural networks, it is good for one to choose the learning algorithm which is to be used. The activation rule in the neural network is fixed, and the same case applies to the target output since it's what we are targeting. Since these two cannot be changed, we have to change the values of the input weights. We, therefore, need a method which we can use to modify the weights in response to the changing inputs/outputs.

Note that once the input vectors are presented to the network, and the answer is found to be correct, then no adjustment will be needed. If the output you get is not equal to your target output, then the weights should be adjusted. In this, an epoch refers to a complete pass of some input vectors.

You can present an input vector to the network at any time, and you will be given a set of output vectors from this.

Chapter 6 Various Models in Machine Learning

Several models have offered solutions in the application of machine learning, there are different types of machine learning models for solving different problems, and it is crucial to mention that each model comes from a background of unique algorithm approach.

The fact that they have different origins means that their performance differs from one data set to another. The question that a beginner will be left with is how to choose from the various models the one that best suits a particular task. This is very simple; you do not have to worry.

The usage of cross-validation is the common method used to determine which is best for a particular test of data. This was important to mention because it will provide a basis for the understanding of the models and their nature of the algorithm.

Several models depend on the type of problem that they are meant to offer a solution for. The models for machine learning include tree based models like the decision tree models, and others are the random forest and gradient boosting.

What are Tree Based Models?

They are also known as tree-based learning algorithms, and they are commonly used in the supervised learning type of machine learning. Tree-based models are of different varieties which include decision trees, random forest, and gradient boosting. To start with, we are going to give the simplest words that can translate to what decision trees are and also how they are important in the discipline of machine learning.

A decision tree is a type of algorithm under supervised learning. It is mostly applicable to classification problems for both categorical and continuous inputs. If this method is to be utilized, the data, and sometimes known as population or sample is split into two or more similar sets known as sub-population by the use of significant differentiators. The following is an example of a decision tree.

The way a decision tree works is based on the ability to identify the most significant variable together with the values that deliver the best similar sets of the population. In this context, it is professional to replace the word alike or similar with "Homogeneous" to refer to characters of similar or closely related traits. It uses various algorithms to identify variables. Now that you have known how a decision tree work and what it is, let us look at the types of decision trees as used in machine learning and artificial intelligence.

So far it is assumed that you understand why ML and AI are used interchangeably; this is because their difference is insignificant.

The commonly used decision tree type are; categorical variable decision tree and continuous variable decision tree. The first one has categorical variables hence the name, and on the other hand, the continuous variable decision tree has continuous variables.

For instance, where an institution can wish to predict if its clients will remit their monthly payment yet they do not have the knowledge of the client's income, the income of the company is the significant variable, and the particulars pertaining to the client's income is not known, then the decision tree will predict the client's income basing on other factors such as the profession and the type of product specifically for continuous variable.

Common Decision Tree Terms

Root node - Representing the whole population of the data forming a task or input to the machine which will be classified according to homogeneous sets.

Decision Node- This is the result of a sub-node divided which becomes a decision node.

Splitting-Division of a node into small sub-nodes

Pruning- This is the removal of sub-nodes of a decision node, and it is the opposite of the splitting.

Leaf or Terminal node- nodes do not split, and they are known as nodes.

Branch- This is a sub-section of an entire tree

Parent and Child sub-node-The nodes are split from the parent node, and the resultant nodes are called in decision trees sub- nodes, the sub-nodes are the children of the original parent nodes.

The merits of Decision Tree

Based algorithms in the machine learning applications

- A decision tree is not parametric since they do not have any assumptions concerning the distribution of space and classifier structure.

- It is capable of handling both numeric and categorical data hence elimination of constraint.

- The influence of outliers and missing values is minimal or impossible, and data cleaning is less required.

- With the help of decision trees, new variables can be identified and can easily be created thereby making it possible for data exploration. For instance, a decision tree can help the most significant variable in a huge data that has a lot of variables. The sense is that it is a good model for data exploration.

- Provides an easier understanding of data analysis and the concept is not even new to the individuals who have never had a clue on the analytical knowledge. They can easily be interpreted, and their hypothesis can easily be related to its users.

Demerits of Decision Tree

It not recommended for continuous variables since it loses information more especially when working on numerical data in the process of categorizing of data to different classes. Overfitting is common in decision trees which its solution is vested on constraint setting in model parameters and also pruning is another solution.

Random Forest

Random forest is another way of designing models for training in machine learning for the purpose of classification regarding the application of variables and correlation. Random forest classifier inputs into several trees of classification and each tree represent a classification and the tree vote for the classification that has the highest number of votes.

One characteristic about the random forest is that pruning is not the art of classification meaning Tress are grown to the largest extent possible. So how does random forest work with the data input about machine learning? When the data is fed to the system, data runs down the tree as proximities are figured out for every variable.

At times a terminal node is occupied by more than one variable or a pair of two different pairs of data, in this case, it means that they are similar, but one good thing about the random forest it will extend the proximities by one branch for distinguishing purposes. You could be wondering what proximities are, they are used for data replacements and locating outliers.

The presence of bootstraps makes random forests not to use the technique of cross-validation or a separate test set, and each tree has a bootstrap from the original data. The proximities are the most important tools for random forest since they form (NxN) matrix.

Characteristics of Random Forests

- It is excellence in accuracy compared to current algorithms

- It works well on efficiently on large databases.

- It has the capability of handling thousands of input variables without any variable deletion.

- The random forest can estimate the variables that are important in classification.

- The random forest can generate an internal unbiased estimate of the generalization error as the forest building progresses.

- Random forest is an effective method for estimating missing data and maintains accuracy when a large proportion of the data are missing it can also balance errors in populations of unbalanced datasets.

- Prototypes are computed that give information about the relation between the variables and the classification.

Chapter 7 Machine Learning Techniques

There are a lot of different techniques and algorithms that you can use when it comes to machine learning. Some of the popular ones that you can use include neural networks, association analysis, and Bayesian inference. We are going to look at a few of these as we go through this chapter, but we will start with some of the higher-level options, such as reinforcement, unsupervised, and supervised to help you get started.

Supervised Learning

The first technique that we are going to work with for machine learning is supervised learning. Supervised algorithms are going to refer to learning that is guided by observations from humans and feedback from the known outcomes. This one will work by showing the data to the machine and then the correct value output of this data as well. The machine will then be able to apply the algorithm to decipher the patterns that exist in the data, and then it will work to develop a model that can produce the same results, even with a new entry data.

Let's look at an example of this. Let's say that you want to teach the machine how to separate out emails into non-spam and spam messages. With supervised learning, you already have some information that you can provide to the machine to describe what type of emails you would like to go into each category. The machine knows that it is working with two different categories and how to sort the data that comes in, and if you did it properly, the machine would know the characteristics of both the non-spam email and the spam email.

These algorithms are going to work well by going backward based on historical data, and they are often used to solve various predictions, such as when you want to be able to predict the prices for used homes and cars. The challenge that comes with these is that you have to have enough data to catch all of the variations, including the anomalies and the outliers, or the machine is going to make the wrong prediction. You also want to make sure that the data is relevant and if you are taking that data from a larger dataset, it needs to be picked at random for the best results.

Unsupervised Learning

It is also possible to work with what is known as unsupervised learning. With this type of learning, there isn't going to be integrated feedback or any data tags. Instead, the machine that is learning the algorithm needs to rely only on clustering data and then modifying the algorithm to respond to the initial findings. All of this needs to be completed without the external feedback of humans.

Clustering of algorithms is a popular example of this type of learning. For example, if you cluster some data points together based on the height and the weight of young high school students, you are likely to find that two clusters come from this data. One of the clusters, usually the larger one, will be the males, and then the other one will be the females. This is because boys and girls are going to have separate averages or commonalities when it comes to their physical measurements.

The advantage of using the unsupervised algorithms is that it is going to make it easy for you to discover the patterns that are there that you may not have seen before inside of the data, such as the presence of different genders in the data. You will be able to use clustering to provide the start for conducting more analysis on a particular group, once you discover what those groups are.

Reinforcement Learning

The next learning that you can work with on machine learning is reinforcement learning, and it is considered one of the most advanced categories of this. Reinforcement learning is often explained through analogies to Pac-Man. As the player that you are working with progress through the space of the game, it is going to learn how each of the actions work in various conditions of the game. These values that they have learned will directly influence how they behave later in the game. After you have learned and explored the space of the game, the player will naturally do better in the game and will get better results based on their prior experience.

Reinforcement is similar to working on a game because you are feeding in algorithms to the machine to help it progress through data that is unknown. The biggest difference in this learning type is that no direct oversight happens with the learning. Instead, the machine is already equipped with the algorithms, and then the machine will need to fend and go through the learning process on its own.

One of the best-known examples of this kind of learning is known as Q-learning. With Q-learning, you are going to start inside of a set environment. Let's look at the example of Pac-Man, and your environment is going to include the pathways, obstacles, and challenges that are near you. There could be a wall on your right, a ghost to the left, and so on. The set of actions that the machine can use to respond to this environment will be referred to as A. With Pac-Man, your actions will be limited to up, down, left, and right movements. And then the other letter that you need is Q, or the starting value, which is going to be 0.

As you go through and explore the space that is in front of you, there are going to be two things that you can have to happen. The first is that Q will drop as negative things occur after a specific action. Or the Q can increase as rewards happen after a specific action. In this type of learning, the machine is going to learn how to match the cation for a specific state that maintains or generates the highest level of Q while avoiding the actions that make it lose out. In the beginning, these are going to be random in the beginning while the computer learns what to do, it is going to start learning as it gets the penalties and rewards for the actions, and the actions will become more planned.

This sounds pretty simple to work with, but it does take a bit of time to implement and make work the way that you would like. But all three of these techniques are going to help you to gain the control that you need with machine learning so that your particular machine can take the information that you provide it, whether you are working with a large amount of data or a small amount, and give you some of the results that you want. It can do this much faster and more efficiently than humans can do, which is one of the reasons that machine learning is so important to use!

Conclusion

Machine learning is a huge subject to be learned in a day or even a month, to be specific in all that it takes to be a machine learning expert, one has to be a fast understanding individual and always to be on the look for the inventions and advancements. The algorithms are dynamic in nature, what we are emphasizing on it the fact that it is a field of computer science that is much angled towards mathematics.

The algorithms are purely based on formulas, and its mastery depends on how you are good in mathematics. As a matter of fact, in the next few years, machine learning and artificial intelligence will be the next investments for those authorities that have tested its benefits. Imagine from today you have vehicles with no human drivers; you get into a store and you are served by a computer. It will be amazing, and this is the next level of implementation of artificial intelligence.

Machine learning being a field in computer science can be and is, as we have seen in this book that it is worth studying although, it is not something that you can learn in a day, probably we are talking about two to three weeks. Yes! That period is enough for a person that wants to learn the concepts of machine learning. Machine learning can be a good aspect of the management data and delivery of the correct results if applied by the right people or rather those who have understood.

One last thing about this subject and this goes to beginners; Machine Learning is a topic and an area of study with diverse technological advancements. These calls for patience and soft understand if you are really in a serious understanding of machine learning. It will give machines an opportunity to interact with the environment and as such be a helpful aspect in solving today's technological and social problem.