AI Sciences

# Machine Learning for Layman

AI SCIENCES

# MACHINE LEARNING FOR LAYMAN

## Concept, Techniques and Tools

AI Sciences Publishing

## How to contact us

Please address comments and questions concerning this book to our customer service by email at:

**contact@aisciences.net**

*Our goal is to provide high-quality books for your technical learning in Data Science and Artificial Intelligence subjects.*

*Thank you so much for buying this book.*

**If you noticed any problem, please let us know by sending us an email at review@aisciences.net before writing any review online. It will be very helpful for us to improve the quality of our books.**

AI SCIENCES

# Table of Contents

- Do you want to discover, learn and understand the methods and techniques of artificial intelligence, data science, computer science, machine learning, deep learning or statistics?
- Would you like to have books that you can read very fast and understand very easily?
- Would you like to practice AI techniques?

If the answers are yes, you are in the right place. The AI Sciences book series is perfectly suited to your expectations!

Our books are the best on the market for beginners, newcomers, students and anyone who wants to learn more about these subjects without going into too much theoretical and mathematical detail. Our books are among the best sellers on Amazon in the field.

### *About Us*

We are a group of experts, PhD students and young practitioners of Artificial Intelligence, Computer Science, Machine Learning and Statistics. Some of us work in big companies like Google, Facebook, Microsoft, KPMG, BCG and Mazars.

We decided to produce a series of books mainly dedicated to beginners and newcomers on the techniques and methods of Machine Learning, Statistics, Artificial Intelligence and Data Science. Initially, our objective was to help only those who wish to understand these techniques more easily and to be able to start without too much theory and without a long reading.

Today we also publish more complete books on some topics for a wider audience.

### *About our Books*

Our books have had phenomenal success and they are today among the best sellers on Amazon. Our books have helped many people to progress and especially to understand these techniques, which are sometimes considered to be complicated rightly or wrongly.

The books we produce are short, very pleasant to read. These books focus on the essentials so that beginners can quickly understand and practice effectively. You will never regret having chosen one of our books.

We also offer you completely free books on our website: Visit our site and subscribe in our Email-List: www.aisciences.net

By subscribing to our mailing list, we also offer you all our new books for free and continuously.

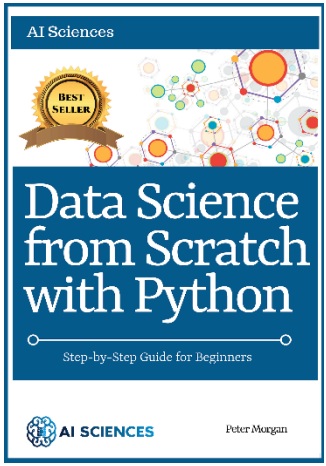
### *To Contact Us:*

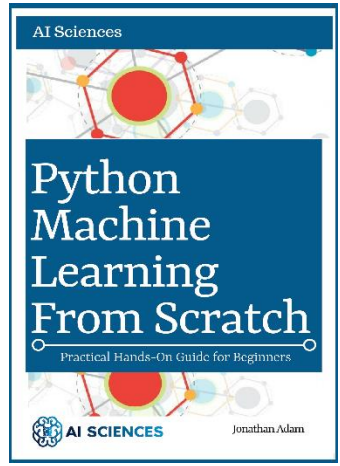
- Website: www.aisciences.net
- Email: contact@aisciences.net

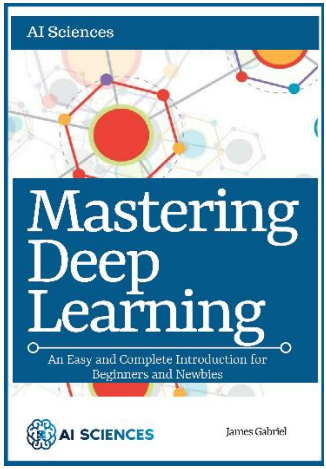
Follow us on social media and share our publications

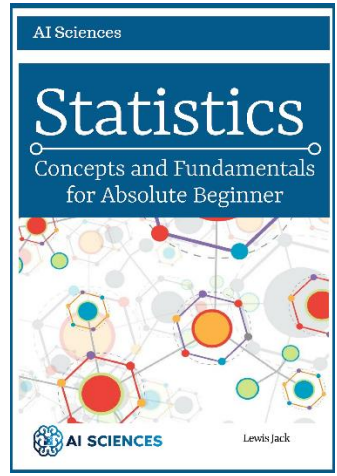
- Facebook: **@aisciencesllc**
- LinkedIn: **AI Sciences**

# From AI Sciences Publishing



**WWW.AISCIENCES.NET**

EBooks, free offers of eBooks and online learning courses.

Did you know that AI Sciences offers free eBooks versions of every books published? Please subscribe to our email list to be aware about our free eBook promotion. Get in touch with us at contact@aisciences.net for more details.


At www.aisciences.net , you can also read a collection of free books and receive exclusive free ebooks.

**WWW.AISCIENCES.NET**

Did you know that AI Sciences offers also online courses?

We want to help you in your career and take control of your future with powerful and easy to follow courses in Data Science, Machine Learning, Deep learning, Statistics and all Artificial Intelligence subjects.

Most courses in Data science and Artificial Intelligence simply bombard you with dense theory. Our course don't throw complex maths at you, but focus on building up your intuition for infinitely better results down the line.

AI SCIENCES
ACADEMY
AI SCIENCES LLC

Please visit our website and subscribe to our email list to be aware about our free courses and promotions. Get in touch with us at academy@aisciences.net for more details.

© Copyright 2017 by AI Sciences
All rights reserved.
First Printing, 2017

Edited by Davies Company
Ebook Converted and Cover by Pixel Studio
Publised by AI Sciences LLC

ISBN-13: 978-1984811671
ISBN-10: 1984811671

The contents of this book may not be reproduced, duplicated or transmitted without the direct written permission of the author.

Under no circumstances will any legal responsibility or blame be held against the publisher for any reparation, damages, or monetary loss due to the information herein, either directly or indirectly.

**Legal Notice:**

You cannot amend, distribute, sell, use, quote or paraphrase any part or the content within this book without the consent of the author.

**Disclaimer Notice:**

Please note the information contained within this document is for educational and entertainment purposes only. No warranties of any kind are expressed or implied. Readers acknowledge that the author is not engaging in the rendering of legal, financial, medical or professional advice. Please consult a licensed professional before attempting any techniques outlined in this book.

By reading this document, the reader agrees that under no circumstances is the author responsible for any losses, direct or indirect, which are incurred as a result of the use of information contained within this document, including, but not limited to, errors, omissions, or inaccuracies.

# Introduction

I want to thank you for choosing this book, 'Machine Learning for layman - Concept, techniques and Tools.'

Machines have come a long way since the Industrial Revolution. They have become a part of our lives. No matter where you go, you will always find a machine around you, but it is only over the last few years that we have understood and enhanced the capabilities of machines. Machines can now perform tasks that involve the simulation of cognition, which is a task that until recently only human beings could accomplish. Driving cars, judging competitions and beating professional chess players at their game are just a few examples of the compound tasks that machines are now capable of performing.

The ever-growing capabilities of machines are instilling a sense of fear among observers. The fear nestles on the insecurities that people have concerning survival, and this concern raises some 'what if' questions. What if machines became more intelligent than human beings and decided to eliminate any human being who was not as smart as them? What if they were able to procreate and produce offspring, like Rachael and Deckard from Blade Runner, who had capabilities that humans never imparted to machines? What if singularity, which is considered to be a myth, is true?

The other fear that people have is the threat to job security. BBC had conducted an online survey, "Will a robot take my job?" and it was concluded that job titles like waiters/waitresses, chartered accountants, taxi drivers, bar

workers, and receptionists would become automated by the year 2040.

When you come across any research on job automation, you should read it with a slight level of skepticism since the future of artificial intelligence and machines is unknown. Although technology is moving fast, its adoption is an unchartered path with unforeseen challenges. Machine learning is not as simple as turning a switch on and off. It is not an out-of-the-box solution either. Machines and statistical algorithms work in parallel with each other, which are the tasks that machine-learning engineers and data scientists oversee. Some industry experts believe that there will come a time when there will not be many people to operate machines.

Your path to understanding machine learning starts from this very moment. If you do not want to become an expert, you can also satiate your thirst by understanding the basics of machine learning for now. But, let us assume that you are looking at becoming a data scientist or machine-learning engineer in the future. This book will help you achieve the same.

The book covers all the information that one would need to gather on machine learning. We will first look at what machine learning entails and also the subjects that it deals with. We will deal with some of these issues that are linked to machine learning on a broader level at a later stage in the book. You will also gather knowledge on different machine learning techniques that have been developed over the years. All you need to remember is that every expert in machine learning started from here. I hope this book provides you with all the

information you need to kick-start your machine-learning career. Let's begin!

# Chapter One: What is Machine Learning?

Learning is an activity that is always difficult to define since it comprises of a wide range of processes. If you were to look for the definition of learning in a dictionary, you would come across the following phrases - "to gain knowledge, or understanding of, or skill in, by study, instruction, or experience," and "modification of a behavioral tendency by experience." In a manner that is similar to psychologists studying the human mind, this book will focus on the learning processes of machines. The two spheres of machine learning and animal learning are intertwined. There are a few techniques derived from the former that are used in the latter and vice versa. It has also been seen that breakthroughs in machine learning brought to light a few facets of animal or biological learning.

When it comes to machines, it can be said that a change made to the structure of the machine, or the data stored on the machine to help the machine perform better and efficiently, can be a sign of learning in the machine. However, when we go deeper into the field, only a few of these changes can be subsumed into the category of machine learning. For instance, consider a machine that is meant to predict weather forecasts in a particular area for a few weeks. If data about the weather in the area over the past year is added to the memory of the machine, the machine can learn from this data and predict the weather more accurately. This instance can be called machine learning.

To be precise, the field of machine learning applies to machines that are associated with artificial intelligence. Machines related to artificial intelligence are responsible for

tasks such as prediction, diagnosis, recognition, and others. These types of machines "learn" from data that is added to them. This data is called training data because it is used to train the machine. The machines analyze patterns in this data and use these models to perform their actions. Machines use different learning mechanisms to analyze the data depending on the actions they are required to perform. These mechanisms can be classified into two broad categories- supervised learning and unsupervised learning.

Skeptics of the field of machine learning might question why machines need to learn in the first place. People may wonder why machines aren't designed specifically for the tasks that they need to carry out. There are many reasons as to why machine learning is advantageous. As mentioned earlier, research into the field of machine learning can help us better understand certain aspects of human learning. Also, machine learning can also increase the accuracy and efficiency of machines. A few other reasons are:

- Even with the most significant efforts by engineers, some tasks cannot be defined explicitly. Some tasks need to be explained to the machine using examples. The idea is to train the machine with the input and teach it how to reach the output. This way, the machine will know how to deal with future contributions and process them to achieve appropriate outputs.
- The field of machine learning is also intertwined with the field of data mining. Data mining is essentially the process of looking through loads and loads of data to find important correlations and relationships. This is another advantage of machine learning in that it might lead to the finding of relevant information.

- On many occasions, it is possible that humans design machines without correctly estimating the conditions in which they will be functioning. Surrounding circumstances can play a significant role in the performance of the machine. In such cases, machine learning can help in the acclimatization of the machine to its environment so that the performance is not hindered. It is also possible that environmental changes might occur and machine learning will help the machine to adapt to these changes without losing out on performance.
- Another loophole in the process of human beings hardcoding the process into the machine is that the process might be incredibly elaborate. In such a case, the programmer might miss out on a few details since it would be a very tedious job to encode all the details. So, it is much more desirable to allow the machine to learn such processes.
- There are constant changes in technology in the world. Major changes occur in other aspects as well such as vocabulary. Redesigning systems to accommodate for every change is not practical. Instead, machine-learning methods can be used to train the machines to adapt to these changes.

## Uses of Machine Learning

Machine Learning is now a solution to complete manual tasks that are impossible to achieve over a short span of time for a large amount of data. In this decade, we are overcome with data and information and have no manual way of processing this information paving the way for automated processes and machines to do that job for us.

Useful information can be derived when the process of analysis and discovery becomes automated. This will help us drive our future actions in an automatic process. We have therefore come into the world of Big Data, Business Analytics, and Data Science. Predictive analytics and Business Intelligence are no longer just for the elite but also for small companies and businesses. This has given these small businesses a chance to participate in the process of collecting and utilizing information efficiently.

Let us now take a look at some technical uses of machine learning and see how these uses can be applied to real-world problems.

### Density Estimation

This use of machine learning allows the system to use the data that is provided to create a product that looks similar to it. For instance, if you were to pick up the novel War and Peace from the shelves of a bookstore and run it through a machine, you will be able to make the machine determine the density of the words in the book and provide you with work that is exactly similar to War and Peace.

### Latent Variables

When you work with latent variables, the machine uses the method of clustering to determine whether the variables are related to one another. This is a useful tool when you do not know what the cause of change in different variables is and also when you do not know the relationship between variables. Additionally, when the data set is large, it is better to look for

latent variables since that helps to comprehend the data obtained.

### Reduction of Dimensionality

Most often, data obtained has some variables and dimensions. If there are more than three dimensions, it is impossible for the human mind to visualize the data. It is in these instances that machine learning can help in reducing the data into a manageable number of dimensions so that the user understands the relationship between the variables easily.

### Visualization

There are times when the user would just like to visualize the relationship that exists between variables or would like to obtain the summary of the data in a visual form. Machine learning assists in both these processes by summarizing the data for the user using specified or non – specified parameters.

# Chapter Two: Subjects Involved in Machine Learning

Machine learning is an intersection of a variety of subjects. The processes involved in machine learning derive information, different terminologies, and methodologies, from each one of these subjects. These concepts together form the discipline of machine learning. This chapter covers a few subjects that are involved in machine learning.

## Statistics

Some problems are studied in statistics. One of the most common problems in prediction is the usage of some samples drawn from a population with an unknown probability distribution and then to predict the distribution of that sample. Another common problem studied in statistics is the estimation of the value of a function at a certain point in time based on the value of the same function at different points of time. The solutions to these problems are instances of machine learning since these issues involve the estimation of certain future events based on historical data. Statistics is an essential part of machine learning.

## Brain modeling

The concept of neural networks is the part of brain modeling that is closely related to machine learning. Some scientists have suggested that one possible model for the neural network is to use non-linear elements with weighted inputs. Extensive studies have been conducted on the use of non-linear elements in recent times. Scientists are trying to gather more information

on human learning by studying these neural networks. Connectionism, sub-symbolic processing, and brain style computation are a few spheres that are associated with these types of studies.

### Adaptive control theory

Adaptive control theory is associated with the control of systems, and one of the major problems faced by every system is the change in the surrounding environment. This subject deals with the different methods that a system can adapt to when these changes occur and to also perform efficiently. The main idea is that the systems should anticipate these changes and modify themselves accordingly.

### Artificial intelligence

As mentioned earlier, a large part of machine learning is concerned with the subject of artificial intelligence. Studies in artificial intelligence have focused on the use of analogies for learning purposes and also on how past experiences can help in anticipating and accommodating future events. In recent years, studies have focused on devising rules for systems that use the concepts of inductive logic programming and decision tree methods.

### Evolutionary models

It is a common idea in evolutionary studies that not only do animals learn to perform better in life, but they also learn to better adapt to their surroundings to enhance their performance. As far as machines are concerned, the concepts of learning and evolving can be considered to be synonymous

with each other. Therefore, models that have been used to explain evolution can be used to devise machine-learning techniques. The most prominent method that has been developed using evolutionary models is the genetic algorithm.

### Psychological modeling

For many years now, psychologists have tried to understand the learning processes of humans. One such example is the EPAM network. This network was used for storing and retrieving one or two words when provided with the other. Later on, the concepts of decision trees and semantic networks were conceived in this field. In recent times, the work in the field of psychology has been strongly influenced by the subject of artificial intelligence. Another aspect of psychology that has been studied in recent times is reinforcement learning. This concept has also been used in machine learning extensively.

### Varieties of Machine Learning

So far, this book has given an introduction to machine learning and has answered the question about the subjects that constitute it. Now, we come to the more critical question of what can be learned on the subject of machine learning. The following are a few topics on which knowledge can be gained through the study of machine learning:

- Programs and logic rule sets
- Terminology and grammars
- Finite state machines
- Problem - solving systems
- Functions
- Artificial Intelligence

- Statistics

Out of the above, the two most focused on topics are those of statistics and artificial intelligence. These two subjects are used extensively in machine learning. We now move on to chapters that describe the two broad categories of machine learning techniques: supervised machine learning and unsupervised machine learning.

# Chapter Three: Facts about Machine Learning

Machine learning was once relegated to sci-fi movies about killer robots and machines. Now, machine learning is permeating numerous aspects of our everyday lives, right from optimizing Netflix recommendations to Google searches. Machine learning has contributed to improving different facets of building mechanics in smart building space, and the experiences of the occupant. You do not have to have a Ph.D. to understand the different facets and functions of machine learning. This section covers some facts about machine learning, which are very basic and essential to know.

## Bifurcation of Machine Learning

Machine learning can be bifurcated into supervised and unsupervised machine learning. These have been covered in detail in the following chapters. Smart buildings would often incorporate both types. Here is a simple example of how these kinds of machine learning look like: Let us assume that you want to teach a computer to recognize an ant. When you use a supervised approach, you will tell the computer that an ant is an insect that could either be small or big. You will also need to tell the computer that the ant could either be red or black. When you use an unsupervised approach, you will need to show the computer different animal groups and then tell the computer what an ant looks like and then show the computer another set of pictures and ask it which one is the ant until the computer learns the features specific to an ant.

When it comes to smart building spaces, both unsupervised, and supervised machine learning is used. Occupant-facing apps are built more rapidly when the prices of building sensors drop to help the occupants provide feedback that would improve the efficiency of the building. The occupants can help the building provide them with optimal climatic conditions.

### Machines are not fully automatic

Machine learning helps computers automate, anticipate and evolve but that does not mean that they can take over the world. Machine learning will still need human beings to operate to provide context, to set parameters for operation and also to continue to improve the algorithms being used.

Machine learning helps a computer discover patterns that are not possible for human beings to see. The computer will then adjust the system. However, it is not right to identify and understand why those patterns exist. For instance, most smart buildings have functions that have been created with the intent to ensure that the people inside the building are more productive. This does not mean that the building can be told that it needs to make people more productive. A human would need to set up the definitions and rules that the building will need to follow.

It is important to note that the data cannot always explain why any anomalies or outliers occur. For instance, an algorithm will always take note, and people in a certain area of work continually request the temperature in that area to be reduced by 1o degrees when compared to any other area in the building. The algorithm will not be able to tell the operator this since it will not be able to identify why the temperature is higher in

that area of the building. Machine learning did help the operator determine why. Therefore, it is important for skilled people to operate machines to ensure that the conclusions obtained are accurate.

### Anyone can use machine Learning

Writing a machine-learning algorithm is very different from learning how to use that algorithm. After all, you do not need to learn how to program when you use an app on your phone. The best platforms always create an abstract of the program to present the users with an interface, which would need minimal training to use. If you do know the basic concepts of machine learning, you are ready to go! It is left to the data scientists to fine-tune the algorithms that can be used for a particular case. Users do not need to understand math; they will just need to use the business domain.

Machine learning has come of this age and is proliferating. Buildings are using machine learning in different ways to make the existing infrastructure efficient and also help to enhance the experience of the occupants residing in the building. Right from an energy usage standpoint, buildings are always learning and analyzing the needs of the occupants.

How does this affect us going forward? This advance in machine learning goes to say that most things will happen without the need for us to ask. Machine learning engineering could go beyond managing lighting and temperature. It can be used to adjust calls, screens, shades, signal elevators and shuttles and so on. Machine learning implies that there will be some future state of multiple layers and levels of automation adjusting based on the current activity.

### Data Transformation is where the work lies

When you read through the different techniques of machine learning, you will probably assume that machine learning is mostly about selecting the right algorithm and tuning that algorithm to function accurately. The reality is prosaic – a large chunk of your time goes into cleansing the data and then transforming that data into raw features that would better signal the relationship between your data.

### Revolution of Machine Learning has begun

During the 1980s, there was a rapid development and advancement in computing power and computers. This gave rise to enormous amount of fear and excitement around artificial intelligence, computers and machine learning which could help the world solve a variety of ailments – right from household drudgery to diseases. As artificial intelligence and machine learning developed as formal fields of study, turning these ideas and hopes into reality was more difficult to achieve, and artificial intelligence retreated into the world of theory and fantasy. However, in the last decade, the advances in data storage and computing have changed the game again. Tasks that were once considered difficult for machines to learn have now become a reality.

### Machine Learning and Artificial Intelligence are interrelated

Machine learning is a subset of artificial intelligence that drives the process of data mining. What is the difference between these terms? These terms are often used interchangeably, and

experts spend hours debating on where they will need to draw the line between machine learning and artificial intelligence.

Artificial Intelligence can be defined as machines thinking like human beings. The brain can be considered as a computing machine. At any given minute, human beings can capture thousands of data points using the five different senses. The brain can recall memories from the past, draw conclusions based on causes and effects and also make decisions. Human beings learn to recognize patterns, but every being has a limit.

One can think of machine learning as a continuous and automated version of data mining. Data mining is a process that is used to detect certain patterns in data sets that human beings will not be able to find. Machine learning is a process that is capable of reducing the size of the data to detect and extrapolate patterns that will allow us to apply that information to identify new actions and solutions.

In smart building spaces, machine learning enables any building to run efficiently while also responding to occupants' changing needs. For instance, you can consider the variance in how a platform in any smart building may handle a repeated meeting and how any machine learning application could do much more. A scheduling system could be used to adjust the temperatures within the conference room to around 72 degrees on the day of the meeting right before the meeting starts. However, any machine-learning algorithm can make sense of more than a thousand variables at any given time of the year to create an ideal thermal environment during the business meeting.

### Deep Learning

Deep learning has earned a good name by catering to the advancement of machine learning algorithms and applications. Deep learning works towards automating some of the work through engineering. However, it is not a silver bullet and cannot be used if you have not invested some time in cleaning and transforming the data.

# Chapter Four: Types of Machine Learning

## Unsupervised Machine Learning

At this point, the reader should be familiar with the concept of supervised machine learning wherein the machine is trained using sets of inputs and outputs that are desired. However, there are other techniques of machine learning. One of these is known as reinforcement learning. In this method, the machine is designed to interact with its ambient environment through actions. Based on the environment's response to these actions, the machine receives rewards if the environment reacts positively or punishments if it reacts negatively. The machine learns from these reactions and is taught to perform in a manner such that it can maximize the rewards it will obtain in the future. The objective could also be to minimize future punishments. This technique of learning is related to the subjects of control theory in engineering and decision theory in statistics and management sciences.

The main problems studied in these two subjects are more or less equivalent, and the solutions are similar as well. However, the two subjects focus on different parts of the problem. There exists another technique of machine learning that is closely related to game theory and also uses reinforcement learning. The idea here is similar to that in reinforcement learning. The machine produces some actions that affect the surrounding environment, and it receives rewards or punishments depending on the reaction of the environment. However, the main difference is that the environment is not static. It is dynamic and can include other machines as well. These other machines are also capable of producing actions and receiving

rewards (or punishments). So, the objective of the machine is to maximize its future rewards (or minimize its future punishments) taking into account the effects of the other machines in the surroundings.

The application of game theory to such a situation with multiple, dynamic systems is a popular area of research. Finally, the fourth technique is called unsupervised machine learning. In this technique, the machine receives training inputs, but it does not receive any target outputs or rewards and punishments for its actions. This begs the question - how can the machine possibly learn anything without receiving any feedback from the environment or having information about target outputs? However, the idea is to develop a structure in the machine to build representations of the input vectors in such a manner that they can be used for other applications such as prediction and decision-making. Essentially, unsupervised learning can be looked at as the machine identifying patterns in input data that would normally go unnoticed. Two of the most popular examples of unsupervised learning are dimensionality reduction and clustering. The technique of unsupervised learning is closely related to the fields of information theory and statistics.

## Supervised Machine Learning

As mentioned earlier, an essential process of machine learning is called training where the machine is fed with data about past events so that the machine can anticipate future events. When this training data is supervised, it is called supervised machine learning. The data fed essentially consists of training examples. These examples consist of inputs and the desired outputs. These desired outputs are also known as supervisory signals.

The machine uses a supervised learning algorithm that generates an inferred function, which is used to forecast events. If the outputs are discrete, the function is called a classifier, and if the outputs are continuous, the function is known as a regression function. This function is responsible for predicting outputs of future inputs. The algorithm needs to conceive a generalized method of reaching the output from the input based on the previous data. An analogy that can be made in the spheres of human and animal learning is concept learning.

## Overview

Supervised learning is a method that uses a fixed algorithm. Given below are the steps involved in this algorithm:

- The first and foremost step in supervised learning is the determination of the type of examples to be used for training the machine. This is an extremely crucial step, and the engineer needs to be very careful in deciding the kind of data he wants to use as examples. For instance, for a speech recognition system, the engineer could either use single words, small sentences or entire paragraphs for training the machine.
- Once the engineer has decided on the type of data he wants to use, he needs to collect data to form a training set. This set needs to be representative of all the possibilities of that function. So, the second step requires the engineer to collect inputs and desired outputs for the training process.

- Now, the next step is to determine how to represent the input data to the machine. This is very important since the accuracy of the machine depends on the input representation of the function. Normally, the representation is done in the form of a vector. This vector contains information about various characteristic features of the input. However, the vector should not include information on too many features since this would increase the time taken for training. A larger number of features might also lead to mistakes made by the machine in prediction. The vector needs to contain exactly enough data to predict outputs.

- After deciding on the representation of input data, a decision must be made on the structure of the function. The learning algorithm to be used must also be agreed on. Most commonly used algorithms are decision trees or support vector machines.

- Now the engineer must complete the design. The learning algorithm chosen should be run on the data set that has been gathered for training. Sometimes, certain algorithms require the engineer to decide on some control parameters to ensure that the algorithm works well. Testing can estimate these parameters on a smaller subset or by using the method of cross-validation.

- After running the algorithm and generating the function, the accuracy of the function should be calculated. For this, engineers use a testing set. This set of data is different from the training data, and the

corresponding outputs to the inputs are already known. The test set inputs are sent to the machine, and the outputs obtained are checked with those in the test set.

There are some supervised learning algorithms in use, and each one has its strengths and weaknesses. Since no definitive algorithm can be used for all instances, the selection of the learning algorithm is a significant step in the procedure.

## Issues to consider in Supervised Learning

With the usage of supervised learning algorithms, there arise a few issues associated with it. Given below are four major issues:

### *Bias-variance tradeoff*

The first issue that needs to be kept in mind while working with machine learning is the bias-variance tradeoff. Consider a situation where we have various but equally good training sets. If when a machine is trained with different data sets, it gives systematically incorrect output predictions for a certain output, the learning algorithm is said to be biased towards that input. A learning algorithm can also be considered to have a high variance for input. This occurs when the algorithm causes the machine to predict different outputs for that input in each training set. The sum of the bias and variance of the learning algorithm is known as the prediction error for the classifier function. There exists a tradeoff between bias and variance. A requirement for learning algorithms with low bias is that they need to be flexible enough to accommodate all the data sets. However, if they are too flexible, the learning algorithms might

end up giving varying outputs for each training set and therefore increases the variance. Supervised learning methods need to be able to adjust this tradeoff. This is done automatically or by using an adjustable parameter.

### *Function complexity and amount of training data*

The second issue is concerned with deciding on the amount of training data based on the complexity of the classifier or regression function to be generated. Suppose the function to be generated is simple, a learning algorithm that is relatively inflexible with low variance and high bias will be able to learn from a small amount of training data. However, on many occasions, the function will be complex. This can be the case due to a large number of input features being involved or due to the machine being expected to behave differently for different parts of the input vector. In such cases, the function can only be learned from a large amount of training data. These cases also require the algorithms used to be flexible with low bias and high variance. Therefore, efficient learning algorithms automatically arrive at a tradeoff for the bias and variance depending on the complexity of the function and the amount of training data required.

### *Dimensionality of the input space*

Another issue that needs to be dealt with is the dimensionality of the input vector space. If the input vector includes a large number of features, the learning problem will become difficult even if the function only considers a few of these features as valuable inputs. This is simply because the extra and unnecessary dimensions could lead to confusion and could cause the learning algorithm to have high variance. So, when

the input dimensions are large, the classifier is adjusted to offset the effects by having low variation and high bias. In practice, the engineer could manually remove the irrelevant features to improve the accuracy and efficiency of the learning algorithm. However, this might not always be a practical solution. In recent times, many algorithms have been developed which are capable of removing unnecessary features and retaining only the relevant ones. This concept is known as dimensionality reduction that helps in mapping input data into lower dimensions to improve the performance of the learning algorithm.

### *Noise in the output values*

The final issue on this list is concerned with the interference of noise in the desired output values. It is possible that the values of the desired outputs (supervisory targets) can be wrong due to the noise that gets added at sensors. These values could also be wrong due to human error. In such cases, the learning algorithm should not look to match the training inputs with their specific outputs. For such cases, algorithms with high bias and low variance are desirable.

### *Other factors to consider*

- One important thing to be kept in mind is the heterogeneity of data. The level of heterogeneity of the data should also play a role in dictating the learning algorithm that is to be chosen. Some algorithms work better on data sets whose inputs are limited within small ranges. A few of these are support vector machines, logistic regression, neural networks, linear regression and nearest neighbor methods. Nearest neighbor methods and support vector machines with

Gaussian kernels work especially better with inputs limited to small ranges. On the other hand, there exist algorithms like decision trees that work very well with heterogeneous data sets.

- Another feature of the data sets that need to be considered is the amount of redundancy in the set. A few algorithms perform poorly in the presence of excessive redundancy. This happens due to numerical instabilities. Examples of these types of algorithms are logistic regression, linear regression, and distance-based methods. For such cases, regularization needs to be included so that the algorithms can perform better.
- While choosing algorithms, engineers need to consider the amount of non-linearities in the inputs and the interactions within different features of the input vector. If there is little to no interaction and each feature contributes independently to the output, algorithms based on distance functions and linear functions perform very efficiently. However, when there are some interactions with the input features, algorithms based on decision trees and neural networks are desirable. The reason for this is that these algorithms are designed to detect these interactions in the input vectors. If the engineer decides to use linear algorithms, he must specify the interactions that exist.

When an engineer is tasked with selecting an algorithm for a specific application, he may choose to compare various algorithms experimentally to decide which one is best suited for the application. However, a large amount of time needs to be invested by the engineer in collecting training data and tuning the algorithm. If provided with a large number of resources, it is advisable to spend more time collecting data than spending time on tuning the algorithm because the latter is extremely tedious. The most commonly used learning

algorithms are neural networks, nearest neighbor algorithms, linear and logistic regressions, support vector machines and decision trees.

# Chapter Five: Machine Learning Techniques

There are some concepts that are involved in machine learning, and these subjects have been covered earlier. This chapter covers some of the most important concepts used in machine learning.

## Dimension Reduction Methods

The databases that are used have thousands and millions of records and variables. It would be impossible to conclude that these variables are not dependent on one another with absolutely any correlation among them. It is essential for a data user to keep in mind that there could be multiple collinear ties between the variables – this is a condition where the predictor variables are all correlated in one way or another.

A lot of instability arises in the solution set when there is multicollinearity between variables. This will lead to results that are incoherent. For instance, if you look at multiple regressions, you will find that multiple correlations between predictor variables will result in a solution that will have a significant impact on the solution set, even though none of the variables may have any significant impact on the solution set when considered independently.

Even if one were to identify a way to avoid such instability, if the user were to include variables that have a high level of correlation between them, this would lead to overemphasis on

specific components of the model. This is because this particular component is being counted twice.

When too many predictor variables are used, there is an unnecessary complication that arises when we need to identify the way to model a relationship between a response variable and the predictor variables. This will also complicate the analysis and its interpretation, and it also violates the principle of parsimony. The principle states that an analyst should always stick to a certain number of predictor variables, which will make it easy to interpret the analysis. If one were to retain too many variables, there is a possibility that there could be over fitting, which would lead to a hindrance in the analysis. This is because the new data that is obtained would not behave in the same way as the training data or predictor data used.

There is also the question of how the analysis performed at the variable level would miss the relationships that lie between the predictors. For instance, there could be numerous single predictor variables that would fall into a single group or component that would address only one aspect of the data. If you were to look at a person's account, you would need to group the account balance and any deposits or savings made from that particular account into one category alone.

There are certain applications, such as image analysis, in which retaining the full dimensionality of the variable would make the problem at hand intractable. For instance, a face classification system based on 256×256 pixel images could potentially require vectors of dimension 65,536. Human beings can discern and understand certain patterns in an image at a glance; these patterns could elude the human eye if they were to be represented algebraically or graphically.

However, the most advanced visualization techniques also do not go beyond five dimensions. How do you think we will be able to identify the relationships that could exist between a massive data set that has thousands of variables?

The goal of dimension reduction methods is to use the structure of correlation among the different predictor variables to accomplish the following goals:

- Reduce the number of predictor components in the data set
- Ensure that these predictor components are independent of one another
- Provide a dynamic framework which would help in the interpretation of the analysis

The most common dimension reduction methods are Principal Component Analysis (PCA), User Defined Composites, and Factor Analysis.

## Fundamental Concepts of Probability

Probability is the most basic concept in statistics that one would need to know. Before you begin to understand data using statistics, you will need to learn to identify whether you are looking at inferential or descriptive statistics. You will also need to grasp the concepts of random variables, probability distributions and expectations. The sections that follow cover some of these aspects in detail.

## Probability and Inferential Statistics

When mathematical operations are performed on numerical data, you obtain a statistic. These statistics are often used to

make decisions for the firm. You always come across two types of statistics:

### *Descriptive Statistics*

This type of a statistic focuses on providing you with a description that provides information on some characteristics of your data.

### *Inferential Statistics*

Instead of focusing on just the descriptions of your dataset, inferential statistics helps to carve out smaller sections of the data to make a deduction about the larger sample. This type of statistics is often used to obtain information on some real-world measures in which the firm is interested.

Descriptive statistics helps one understand the characteristics of a numerical dataset. However, this does not help you understand why you would need to care about the data. Most data scientists are interested in descriptive statistics since they can understand the characteristics of certain real-world measures described by the dataset.

For instance, assume that a business owner would want to estimate the profits in the upcoming quarter. He can choose to take the average of the last few quarters and estimate how much of a profit he would make the following quarter. If the profits in the past quarters varied by a huge amount, a descriptive statistic called the variation could be used to understand how far off the predicted statistic is of the actual profit.

Inferential statistics reveal something about the data that you are interested in which is similar to what descriptive statistics

is. However, inferential statistics only provide information on smaller samples of data, which would help the data scientist make assumptions about the larger dataset, called the population.

If your dataset is too big, it would be easier to pull out a sample from that data and make inferences about the entire dataset from there. You can use inferential statistics where you simply cannot collect the data for the entire population. There are times when you may not have access to complete information. At such times, you will need to use inferential statistics to make assumptions about the population.

## Understanding Random Variables and Expectations

You are on holiday in Las Vegas and have decided to go to a casino. You have settled into your favorite chair at the roulette table and have just turned the wheel. You have understood that there is an equal chance that the ball can fall into any of the slots in the cylinder. The slot where the ball can fall is random, and the probability or likelihood of that happening is the same for each slot on the table. Since there is an equal probability of the ball landing in any slot, the random variable would follow the uniform distribution.

But, not every slot on the wheel is the same – twenty slots are either red or green, and there are 18 that are black. This would mean that there is an 18/38 probability that your ball will land on a black slot. You plan to make successive bets on your ball landing on the black slot.

Your net winnings can be considered to be a random variable here. A random variable is a measure of value or trait that is associated with a place, person or an object. This is something

that is unpredictable. That being said, it does not mean that one does not know anything about the random variable. You can use what you know about the random variable to help you make an informed decision.

You can take a weighted average – an average value over a large number of data points – of your winnings across the distribution, which yields the expectation of the random variable. This expectation is the expected value of all your winnings over many bets made. If you had to describe it formally, you would need to remember that expectation is a weighted average of any measure that is associated with the random variable that is being considered. If you are trying to derive a model for an unpredictable variable, you can always use probability and random variables.

Let us assume that a data scientist is walking down a street in California and is looking at the eye color of the people walking past him. She would notice people with green eyes, brown eyes, blue eyes and so forth. She will not have an idea about the eye color of the person she will see next. But, she has identified that brown and blue eyes are the most common. Since she has observed this, you will make an educated guess on what the eye color of the next person can be. The random variable here would be the eye color, and her guess on what the eye color of the next person would be based on the probability that distribution of the eye color of people walking along the street.

Are you ready to dig a little deeper? Let us get a little more quantitative. If the data scientist sat down and recorded how many people were observed with the same eye color, she would be able to create a frequency distribution that will help her understand what the percentage is. These percentages, called

percentiles in statistics will help her make an informed decision about the eye color of the population. These percentiles represent the probability distribution and the expectation calculated would be similar to how we calculated it in the roulette wheel example above.

There are many probability distributions that you would need to understand. However, you do not need to become a master at understanding these distributions since you can use programming languages like Python and R to identify the right distribution for your data.

### Regression Modeling

Regression modeling is an elegant and powerful method for estimating the value of target variables that are continuous. Multiple regression models could be used – one of the simplest models is the simple linear regression model. In this model, a straight line is used to estimate the relationship between a single continuous response variable and a single continuous predictor variable. There are also the multiple regression models where numerous predictor variables can be used to estimate one response.

Apart from the methods mentioned above, there is also the least shared regression method, which is a potent tool that can be used. However, there is a certain level of disparity when it comes to the assumptions of the model. It is essential that these assumptions be validated before the model is even built. If the user were to build and use a model that was based on assumptions that have not been verified, this would lead to failures that could cause too much damage to the company or the individual.

Once the user has obtained the results from the model, he or she will need to be certain that there exists absolutely no linear relationship between the variables used in the model. There could be a relationship that exists that would be very granular and difficult to identify. There is, however, a systematic approach to determining whether there is any linear relationship between the variables, and that is using inference. There are four inferential methods that could be used to determine the relationship:

- The t-test for the relationship between the response variable and the predictor variable
- The confidence interval for the slope, $\beta 1$
- The confidence interval for the mean of the response variable given a particular value of the predictor
- The prediction interval for a random value of the response variable given a particular value of the predictor

The inferential methods described above all depend on the adherence of data to the assumptions that are made at the beginning of the process. The level at which the data adheres to the assumptions can be identified using two graphical methods – a plot of the normal probability and a plot of the standardized residuals against the predicted or fitted values.

A normal probability plot is a quantile-quantile plot of the quantiles of a particular distribution against the quantiles of the standard normal distribution for the purposes of determining whether the specified distribution deviates from normality. In a normality plot, the values observed for the distribution of interest are compared against the same number of values that would be expected from the normal distribution. If the distribution is normal, the bulk of the

points in the plot should fall in a straight line; systematic deviations from linearity in this plot indicate non-normality. We evaluate the validity of the regression assumptions by observing whether certain patterns exist in the plot of the residuals versus fits, in which case one of the assumptions has been violated, or whether no such discernible patterns exists, in which case the assumptions remain intact.

If these graphs indicate violations of the assumptions, we may apply a transformation to the response variable y, such as the ln (natural log, log to the base e) transformation. Transformations may also be called if the relationship between the predictor and the response variables is not linear. We may use either "Mosteller and Tukey's ladder of re-expression" or a "Box-Cox transformation."

## Multiple Regression

In the previous section, we took a look at regression modeling using simple linear regression where we considered a single predictor variable and a single response variable. However, the only interest that data miners have is on the relationship that exists between a set of predictor variables and a target variable. Most applications built for data mining have a lot of data, with some sets including thousands or millions of variables, of which most have a linear relationship with the response or target variable. That is where a data miner would prefer to use a multiple linear regression models. These models provide improved accuracy and precision of prediction and estimation, similar to the improved accuracy of regression estimates over bivariate or univariate estimates.

Multiple linear regression models use linear surfaces like hyperplanes or planes to determine the relationship between a set of predictor variables and one continuous target or response variable. Predictor variables are often continuous, but there could be categorical predictor variables included in the model through the use of dummy or indicator variables. In a simple linear regression model, a straight line of dimension one is used to estimate the relationship between one predictor and the response variable. If we were to evaluate the relationship between two predictor variables and one response variable, we would have to use a plane to estimate it because a plane is a linear surface in two dimensions.

Data miners need to guard against multicollinearity, a condition where some of the predictor variables are correlated with each other. Multicollinearity leads to instability in the solution space, leading to possible incoherent results. For example, in a dataset with severe multicollinearity, it is possible for the F-test for the overall regression to be significant, whereas none of the t-tests for the individual predictors are significant. This situation is analogous to enjoying the whole pizza while not enjoying any of the slices.

The high variability associated with the estimates for different regression coefficients means that different samples may produce coefficient estimates with widely different values. For example, one sample may provide a positive coefficient estimate for x1, whereas a second sample may produce a negative coefficient estimate. This situation is unacceptable when the analytic task calls for an explanation of the relationship between the response and the predictors individually. If there was a chance to avoid such instability when variables that are highly correlated are included, those

variables tend to emphasize a particular component of the model being used because these elements are being counted twice. To avoid multicollinearity, the analyst should investigate the correlation structure among the predictor variables (ignoring the target variable for the moment).

However, suppose that we did not check for the presence of correlation among our predictors but went ahead and performed the regression anyway. Is there some way that the regression results can warn us of the presence of multicollinearity? The answer is yes; we may ask for the variance inflation factors (VIFs) to be reported. Note that we need to standardize the variables involved in the composite to avoid the possibility that the greater variability of one of the variables will overwhelm that of the other variable.

## Regression with Categorical Predictors

Thus far, our predictors have all been continuous. However, categorical predictor variables may also be used as inputs to regression models through the use of indicator variables (dummy variables).

For use in regression, a categorical variable with k categories must be transformed into a set of $k-1$ indicator variables. An indicator variable, also known as a dummy variable, is a binary 0/1 variable, which takes the value one if the observation belongs to the given category and takes the value 0 otherwise.

## Logistic Regression

Linear regression is used to approximate the relationship between a continuous response variable and a set of predictor

variables. However, the response variable is often categorical rather than continuous. For such cases, linear regression is not appropriate, but the analyst can turn to an analogous method, logistic regression, which is similar to linear regression in many ways. Logistic regression refers to methods for describing the relationship between a categorical response variable and a set of predictor variables.

One of the most attractive properties of linear regression is that closed-form solutions for the optimal values of the regression coefficients may be obtained by the least-squares method. Unfortunately, no such closed-form solution exists for estimating logistic regression coefficients. Thus, we must turn to maximum likelihood estimation, which finds estimates of the parameters for which the likelihood of observing the data is maximized.

The maximum likelihood estimators may be found by differentiating the likelihood function, $L(\beta|x)$, with respect to each parameter and then setting the resulting forms to be equal to zero. Unfortunately, unlike linear regression, closed-form solutions for these differentiations are not available. Therefore, other methods, such as iterative weighted least squares, must be applied.

In summary, linear regression is used to approximate the relationship between a continuous response variable and a set of predictor variables. Logistic regression, on the other hand, refers to methods for describing the relationship between a categorical response variable and a set of predictor variables.

Logistic regression assumes that the relationship between the predictor and the response is nonlinear. In linear regression, the response variable is considered to be a random variable Y

$= \beta 0 + \beta 1x + \varepsilon$ with conditional mean $\pi(x) = E(Y|x) = \beta 0 + \beta 1x$. The conditional mean for logistic regression takes on a different form from that of linear regression.

## Variable Selection Methods

To assist the data analyst in determining which variables should be included in a multiple regression model, several different variable selection methods have been developed, including forward selection, backward elimination, stepwise selection, and best subsets. These variable selection methods are essentially algorithms to help construct the model with the optimal set of predictors.

### *Forward Selection Procedure*

No variables are used in the model when employing this method.

Step 1: For the first variable to enter the model, select the predictor most highly correlated with the target. (Without loss of generality, denote this variable x1.) If the resulting model is not significant, stop and report that no variables are significant predictors; otherwise, proceed to step 2.

Step 2: For each remaining variable, compute the sequential F-statistic for that variable given the variables already in the model. For example, in this first pass through the algorithm, these sequential F-statistics would be F (x2|x1), F (x3|x1), and F (x4|x1). On the second pass through the algorithm, these might be F (x3|x1, x2), and F (x4|x1, x2). Select the variable with the largest sequential F-statistic.

Step 3: For the variable selected in step 2, test for the significance of the sequential F-statistic. If the resulting model

is not significant, stop, and report the current model without adding the variable from step 2. Otherwise, add the variable from step 2 into the model, and return to step 2.

### *Backward Elimination Procedure*

The backward elimination procedure begins with all the variables or all of a user-specified set of variables in the model.

Step 1: Perform the regression on the full model, that is, using all available variables. For example, perhaps the full model has four variables, namely, x1, x2, x3, and x4.

Step 2: For each variable in the current model, compute the partial F-statistic. In the first pass through the algorithm, these would be F (x1|x2, x3, x4), F (x2|x1, x3, x4), F (x3|x1, x2, x4), and F (x4|x1, x2, x3). Select the variable with the smallest partial F-statistic. Denote this value as F-min.

Step 3: Test for the significance of F-min. If F-min is not significant, remove the variable associated with F-min from the model, and return to step 2. If F-min is significant, stop the algorithm, and report the current model. If this is the first pass through the algorithm, the current model is the full model. If this is not the first pass, the current model has been reduced by one or more variables from the full model.

### *Stepwise Procedure*

The stepwise procedure represents a modification of the forward selection procedure. A variable that has been entered into the model early in the forward selection process may turn out to be nonsignificant once other variables have been entered into the model. The stepwise procedure checks on this possibility by performing at each step a partial F-test using the

partial sum of squares for each variable currently in the model. If there is a variable in the model that is no longer significant, the variable with the smallest partial F-statistic is removed from the model. The procedure terminates when no further variables can be entered or removed.

### *Best Subsets Procedure*

For data sets where the number of predictors is not too large, the best subsets procedure represents an attractive variable selection method. However, if there are more than 30 or so predictors, the best subsets method encounters a combinatorial explosion and becomes intractably slow. The best subsets procedure works as follows:

Step 1: The analyst specifies how many (k) models of each size he or she would like to have reported, as well as the maximum number of predictors (p) the analyst wants in the model.

Step 2: All models of one predictor are built: for example, $y = \beta 0 + \beta 1$ (sugars) $+\varepsilon$, $y = \beta 0 + \beta 2$ (fiber)$+\varepsilon$, and so on. Their R2, R2adj, Mallows' Cp, and s values are calculated. The best k models are reported based on these measures.

Step 3: All models of the two predictors are built: for example, $y = \beta 0 + \beta 1$ (sugars) $+\beta 2$ (fiber)$+\varepsilon$, $y = \beta 0 + \beta 1$ (sugars)$+\beta 4$ (shelf2)$+\varepsilon$, and so on. Their R2, R2adj, Mallows' Cp, and s values are calculated, and the best k models are reported.

The procedure continues in this way until the maximum number of predictors (p) is reached. The analyst then has a listing of the best models of each size 1, 2... p to assist in the selection of the best overall model.

## Naïve Bayes Estimation and Bayesian Networks

In the field of statistics, probability is approached in two ways – the classical approach or the Bayesian approach. Probability is often taught using the classical approach or the frequentist approach. This is a method that is followed in all beginners' classes in statistics. In a frequentist method of probability, the population constraints are fixed constants with values that are indefinite. These prospects are defined as "*the relative frequencies of the various categories, where the experiment is repeated an indefinitely large number of times*." For example, if we toss a coin ten times, it may not be very unusual to observe 80% heads, but if we toss the same coin 10 trillion times, we can be fairly certain that the proportion of heads will be near 50%. It is this behavior that defines prospect of the frequentist approach.

However, certain situations do arise in which the classical definition of probability makes it difficult to understand the situation. For example, what is the likelihood that terrorists will strike New York City with a dirty bomb? Given that such an occurrence has never occurred, it is difficult to conceive what the long-term behavior of this gruesome experiment might be. Another approach to probability, the frequentist approach, uses parameters that are fixed so that the randomness lies only in the data. This randomness is viewed as a random sample from a given distribution with unknown but fixed parameters.

These assumptions are turned around in the Bayesian approach to probability. In this approach to probability, the parameters are all considered to be random variables with data that is known. The parameters are regarded as coming from a distribution of possible values, and Bayesian look to the

observed data to provide information on likely parameter values.

Criticism of the Bayesian framework has focused primarily on two potential drawbacks. First, the elicitation of a prior distribution may be subjective. That is, two different subject matter experts may provide two different prior distributions, which will presumably percolate through to result in two different posterior distributions. The solution to this problem is (1) to select non-informative priors if the choice of priors is controversial and (2) to apply a large amount of data so that the relative importance of the prior is diminished. Failing this, model selection can be performed on the two different posterior distributions using model adequacy and efficacy criteria, resulting in the choice of the better model. Is reporting more than one model a bad thing?

The second criticism has been that Bayesian computation has been intractable in data mining terms for most interesting problems where the approach suffered from scalability issues. The curse of dimensionality hits Bayesian analysis rather hard, given that the normalizing factor requires integrating (or summing) over all possible values of the parameter vector, which may be computationally infeasible when applied directly. However, the introduction of Markov chain Monte Carlo (MCMC) methods, such as Gibbs sampling and the Metropolis algorithm, has greatly expanded the range of problems and dimensions that Bayesian analysis can handle.

### Genetic Algorithms

Genetic algorithms (GAs) attempt to mimic computationally the processes by which natural selection operates and apply

them to solve business and research problems. Developed by John Holland in the 1960s and 1970s, genetic algorithms provide a framework for studying the effects of such biologically inspired factors as mate selection, reproduction, mutation, and crossover of genetic information. In the natural world, the constraints and stresses of a particular environment force different species (and different individuals within species) to compete to produce the fittest offspring. In the world of genetic algorithms, the fitness of various potential solutions are compared, and the fittest potential solutions evolve to produce ever more optimal solutions.

Unsurprisingly, the field of genetic algorithms has borrowed heavily from genomic terminology. Each cell in the body contains the same set of chromosomes – strings of DNA that function as a blueprint for making one of us. Then, each chromosome can be partitioned into genes, which are blocks of DNA designed to encode a particular trait, such as eye color. A particular instance of the gene (e.g., brown eyes) is an allele. Each gene is to be found at a particular locus on the chromosome. Recombination, or crossover, occurs during reproduction, where a new chromosome is formed by combining the characteristics of both parents' chromosomes. Mutation, the alteration of a single gene in a chromosome of the offspring, may occur randomly and relatively rarely. The offspring's fitness is then evaluated, either regarding viability (living long enough to reproduce) or in the offspring's fertility.

In the field of genetic algorithms, a chromosome refers to one of the candidate solutions to the problem, a gene is a single bit or digit of the candidate solution, and an allele is a particular instance of the bit or digit (e.g., 0 for binary-encoded solutions or the number 7 for real-valued solutions). Recall that binary

numbers have base 2, such that the first "decimal" place represents "ones," the second represents "twos," the third represents "fours," the fourth represents "eights," and so on.

Genetic algorithms use the following three operators:

- Selection: The selection operator refers to the method used for selecting which chromosomes will be reproducing. The fitness function evaluates each of the chromosomes (candidate solutions), and the fitter the chromosome, the more likely it will be selected to reproduce.
- Crossover: This operator implements recombination, where it creates two original offspring by casually choosing a locus and exchanging subsequences towards the left and right of that locus between two chromosomes chosen during selection. For example, in binary representation, two strings, 11111111 and 00000000, could be crossed over at the sixth locus in each to generate the two new offspring, 11111000 and 00000111.
- Mutation: The mutation operator randomly changes the bits or digits at a particular locus in a chromosome. Usually, however, it has a minimal probability. For example, after crossover, the 11111000-child string could be mutated at locus two to become 10111000. It presents new information to the genetic pool and safeguards against converging too quickly to a local optimum.

Most genetic algorithms function by iteratively updating a collection of potential solutions called a population. Each member of the population is evaluated for fitness on each cycle. A new population then replaces the old population using the operators above, with the fittest members being chosen for reproduction or cloning. The fitness function f (x) is a real-

valued function operating on the chromosome (potential solution), not the gene, such that the x in f (x) refers to the numeric value taken by the chromosome at the time of fitness evaluation.

# Chapter Six: Top Six Real Life Applications of Machine Learning

We use some applications regularly that involve machine learning. This chapter covers some of the top six applications.

## Image Recognition

The most common application of machine learning is image recognition, and most laptops and phones have this tool. There are many situations when you can classify a certain object as an image. The measurements of every digital image always give the user an idea about the output of each pixel in the image.

If you were to look at a black and white image, the intensity of every pixel in the image serves as a measurement. If the image has M*M pixels, the measurement would be M^2.

In a colored image, every pixel is considered as three measurements where each measurement provides intensity to the main color (RGB). So if there were an M*M image, there would be three M^2 measurements.

## Face Detection

The most common category would be the presence of a face versus the presence of no face. There can also be a separate category for every person in a database with multiple individuals.

## Character Recognition

You can segment pieces of writing into images of small sizes where each image contains one character. These categories may comprise the 26 letters of the English alphabet, the first ten numbers and some special characters.

### Speech Recognition

This application is the translation of spoken words into actual text. It is often called Automatic Speech Recognition (ASR), Speech to text (STT) or Computer Speech Recognition (CSR).

When it comes to speech recognition, the software application is trained to recognize spoken words. The measurements that are calculated are often a set of numbers that represent the signal of speech. These signals can be separated into portions that contain phonemes or distinct words. Speech signals in every segment can be represented by the energies or intensities in distinct frequency – time bands.

The details of the representation of signals are outside the scope of this book, but it is good to know that these signals can be represented as a set of real values. Applications on speech recognition often include voice user interfaces. These interfaces are call routing, voice dialing, and other similar applications. These applications can also use data entry and other simple methods of processing information.

### Medical Diagnosis

Machine learning (ML) provides techniques, tools, and methods that help a doctor solve prognostic and diagnostic problems in many medical domains. This is being used for the analysis of clinical parameters and also an analysis of their combinations for prognosis. The result of this analysis would

help enhance the medical knowledge most doctors have. ML is being used for data analysis to help doctors identify the irregularities in dealing with incorrect and unstructured data, the interpretation of continuous data and to monitor results efficiently.

The successful use of different ML methods can always help with integrating computer-based systems in the healthcare environment thereby providing the medical world with opportunities to enhance the treatments being provided.

In medical diagnosis, the interest is to establish the existence of a disease and then identify the disease accurately. There are different categories for each disease that are under consideration and one category where the disease may not be present. Machine learning helps to improve the accuracy of a diagnosis and also analyzes the data of the patients. The measurements used are the results of the many medical tests conducted on the patient. The doctors identify the disease using these measurements.

### Statistical Arbitrage

Statistical Arbitrage, a term often used in finance, refers to trading strategies that are used to identify the short-term securities that can be invested in. In these strategies, the user always tries to implement an algorithm on an array of securities that are based on the general economic variables and historical correlation of the data. The measurements are cast as estimation or classification problems. The basic assumption made is that the price will always move towards a historic average.

Machine learning methods are applied to obtain a strategy called the index arbitrage. Linear regression and support vector regression is used to different prices of a fund and a stream of stocks. Then, Principal Component Analysis is used to reduce the dimensions in the dataset. The residuals are modeled to identify the trading signals as a mean reverting process.

In this study, the case of classification could be sold, buy, hold, do nothing for each security. The expected return for each security could be predicted over a future time horizon. The estimates are often used to decide on whether the investor should buy or sell securities. If you are unclear of the concepts mentioned in this section, please refer to Chapter Five.

### Prediction

Let us assume that a bank is trying to calculate the probability of a loan applicant defaulting on a repayment. To calculate this probability, the system will first have to identify, clean, and classify the data that is available into groups. This classification is performed based on certain criteria set out by the analysts. Once the classification of data has been completed, the probability can be calculated. These calculations can be made across different sectors for a variety of purposes.

Prediction is one of the sought after machine learning algorithms. If you were to look at a retailer, you would only be able to get reports on the sales that happened in the past. This type of reporting is called historical reporting. Now, you can predict what the sales may be shortly. This would help the business make the right decision in the future.

### Learning Associations

The process of developing an insight into the association between groups of products is called learning association. There are many products that reveal an association with one another although they seem unrelated. The association can be identified based on the buying habits of customers.

Basket learning analysis, which deals with studying about the association between products purchased by different customers, is an application of machine learning. If we assume that Amy has bought a product X, we can try to identify if she would buy product Y based on an association between the two products. This can be identified based on an example of fish and chips. If there were a new product launched into the market, then the association between the existing products would change. If one were to know these relationships, they would be able to identify the right products to give to their customers. Products could also be bundled together to increase their purchasing power.

Big Data analysts use machine-learning algorithms to identify if there is a relationship between different products. Conditional probabilities can be used to attest to the relationship between these products.

### Information Extraction

Data comes in both the unstructured and structured form. Information Extraction (IE) is an application of machine learning where structured data is extracted from unstructured data, like emails, articles, web pages, blogs, and articles. It is known that large volumes of data are being generated. Most of this data is unstructured which is makes it difficult to handle the large volumes of data. The process of extraction takes the

input as a set of unstructured data in the form of documents and extracts structured data. The extracted structured data is stored in a relational database. This process of extraction has become extremely important in the big data industry.

### Personal Security

If you attend large public events or fly on an airplane frequently, you will have waited in long lines where you and your belongings have been screened. This is machine learning proving to the world that it can help people spot alarms and overcome human error. The application of machine learning technologies in the area of security helps to speed up the process and also ensure that people everywhere are safe.

# Chapter Seven: Glossary on Important Machine Learning terms

A few terms are often used when it comes to learning machine learning. This chapter covers some of these terms.

## Data Science

Data science, in its truest form, represents the resource and process optimization of data analysis. Through data science, one can produce data insights, which can be used to improve your investments, business, health, lifestyle and social life. For any pursuit or goal, you can use data science methods to help you understand and predict the direct route from where you are right now to where you would like to be shortly. You will also be able to anticipate any obstacles or hurdles that come your way.

## Data Mining

Data mining is a process that is used by numerous businesses to convert the collected raw data into information that can be used by the business. There are specialized tools that can be used to detect patterns in large-scale information, which would help you learn more about your consumers and also respond to their concerns while developing strategies that would help to increase your revenue. The ultimate goal of data mining is to increase your profits. Data mining is said to be useful only when data is gathered and stored efficiently and processed in the right manner for future use.

## Artificial Intelligence

Artificial Intelligence or AI is a field of computer science that is aimed at developing computers that are capable of performing tasks that can be done by people, especially those tasks that are considered to be performed by intelligent people.

## Additive Property

The next axiom that is important to note can only be true when both events A and B are mutually exclusive.

P (A+B) =P (A) +P (B)

This axiom states that the probability of both events A and B occurring is the same as the sum of the individual probabilities of the events occurring if and only if both events exclude each other. For example, if A is the event that we get a six on rolling a die and B is the event that we get a five rolling a die, then this axiom hold. But, if B were assumed to be the event where we get an even number, this set would include the number six making the axiom false.

## Regression

Regression is closely related to classification. Classification is directly concerned with the prediction of discrete classes whereas regression is functional while the class is required to be projected when it is made up of a continuous set of numerical values. Linear regression is an example of regression techniques.

## Joint Probability

This property can be expressed as follows:

P (a, b) = P (A=a, B=b)

This property can be read as the probability of a and b is the same as the probability that event A turns out in state 'a' and event B turns out in state 'b.'

Bayes' Rule

This rule is often used to identify the conditional probability when P (A, B) is not known. The equation used is as follows:

P (A|B) = [P (B|A)*P (A)]/P (B)

These various axioms are a great way to understand the logic behind Markov Models. Let us now take a look at the mathematical aspects of the Markov Model. As mentioned above, Markov Models was discovered in the year 1916 by Andreevich Markov, a scientist who was studying and analyzing the frequency of different types of words in Pushkin's poems. These models have now become an integral model to use while working with data science, artificial intelligence, and machine learning.

## Classification

Classification is concerned with separating data into unique classes using models. These models are always built using training data sets for which the classes have already been named to help the algorithm learn. These models are then used by inputting real-time data where the model holds all the classes. This will help the model predict the relationship that exists within the data based on what the model has learned from the training dataset. Well – known classification schemes are support vector machines and decision trees. Since these

algorithms will need an explicit definition of classes, classification is a form of supervised machine learning.

### Support Vector Machines

Support Vector Machines (SVMs) allow the user to classify linear and nonlinear data. They work by transforming the training dataset into higher dimensions that are then inspected for the optimal boundary separation between classes. In SVMs, these boundaries are often referred to as hyperplanes which are identified using support vectors or the instances that define the classes and their margins which are the lines that are parallel the hyperplanes. The shortest distance between the hyperplane and the support vector associated with it defines these.

The goal behind using an SVM is to identify the hyperplane that separates two classes if there are a large number of dimensions. This process helps to delineate the member classes in the dataset. When this process is repeated a number of times, there are enough hyperplanes that are generated that can help to separate dimensions in an n – dimension space.

### Clustering

Clustering is a technique that is used to analyze data that doesn't comprise of pre-labeled classes or any class characteristic at all. The instances in the data are grouped with the theory of maximizing the similarity within classes and minimize the similarity between classes. This loosely translates into the bundling algorithm, which identifies and groups the instances which are similar to each other when compared to ungrouped instances that do not have too much of a similarity. The most well-known clustering algorithm is k – means

clustering. Clustering does not require the pre-labeling of instance classes; therefore it is a form of unsupervised machine learning meaning that the algorithm learns more from observation as opposed to learning by example.

### Association

Association is easily explained by introducing a market basket analysis that is a task that it is well – known for. This type of analysis always tries to identify the association that exists between different data instances that have been chosen by any particular shopper and placed in their basket. This could either be real or virtual, and the algorithm always assigns confidence and support measures for comparison. The value of this always lies in customer behavior analysis and cross marketing. Association algorithms are generalizations of market-based analyses and are similar to classification algorithms in the sense that any attribute can be predicted when using association. Apriori is one the best–known association algorithms. If you have deduced that association is an example of unsupervised machine learning, then you are right.

### Machine Learning

Although this has been covered in detail above, we will just look at it one more time. Machine learning is concerned with how a computer can be constructed to improve the experience of the user. Machine learning is an interdisciplinary science that employs techniques from different fields like computer science, artificial intelligence, mathematics, and statistics and so on. The main aspects of machine learning research include algorithms that help to facilitate this improvement from

experience. These algorithms can be applied in some fields like artificial intelligence, data mining, and computer vision.

## Decision Trees

Decision trees are recursive, divide–and–conquer and top-down classifiers. These trees are composed of two main tasks: tree pruning and tree induction. The latter is the task where a set of pre-classified instances are taken as inputs after which decisions are made based on which attributes are split on thereby splitting the dataset and recursing on the resulting split datasets until every training instance has been categorized. While building the tree, the main goal is to split all the attributes to create the child nodes that are pure. This would ensure that the number of splits needed to classify the instances in the dataset, would be few. The purity of the child nodes is always measured based on information that relates to the how much information is to be known about a formerly unseen case to classify it accurately.

A complete decision tree prototype can always be complicated and may contain some unnecessary structure that can be tough to interpret. The procedure of eliminating any unnecessary structure from the decision tree in order to make it easily readable, more efficient and accurate for human beings to comprehend is called as tree pruning. This increased accuracy due to tree pruning helps to reduce over fitting.

## Fundamental Axioms

Let us take a look at the various axioms that are used in machine learning to understand the math that supports the

model. One of the most fundamental axioms can be expressed as follows:

$0<P(A)<1$

This states that the probability of any event occurring is always going to be greater than zero but less than one, both inclusive. This implies that the probability of the occurrence of any event can never be negative. This makes sense since the probabilities can never be more certain than hundred percent and least certain than zero percent.

## Deep Learning

Deep learning is the process of applying deep neural network technologies – that is using neural network architectures that have multiple hidden layers of neurons – to solve different problems. These deep neural network architectures are certain machine learning algorithms.

## Generative Model

In statistics and probability, a generative model is used to generate data sets when some parameters are hidden. These models are used in machine learning to either model the data directly or used as an intermediate step to form a conditional probability density function.

# Conclusion

Machine learning has earned a great deal of importance over the last few years. People from different fields have begun to research how they can incorporate machine learning in their field of study. Therefore, it is of utmost importance to understand what machine learning is and how it is linked to different fields of study.

This book provides you with all the information you would need on machine learning. You will gather an idea on the different subjects that are linked to machine learning and some facts about machine learning that make it an interesting subject to learn. Machine learning has been linked to artificial intelligence and data mining since the beginning of time. Therefore, it is important to gather some information about these fields of study too.

Thank you for purchasing the AI Sciences book. We hope you have gathered all the information necessary for machine learning.

# Thank you !

If you enjoyed this book and felt that it added value to your life, we ask that you please take the time to review our books in amazon.

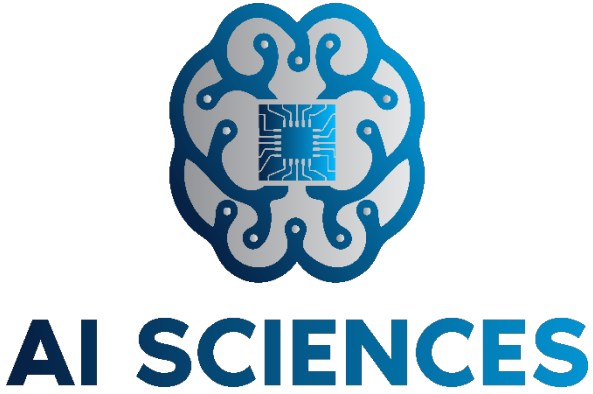
**Your honest feedback would be greatly appreciated. It really does make a difference.**

**If you noticed any problem, please let us know by sending us an email at review@aisciences.net before writing any review online. It will be very helpful for us to improve the quality of our books.**


**We are a very small publishing company and our survival depends on your reviews.**
**Please, take a minute to write us an honest review.**

**If you want to help us produce more material like this, then please leave an honest review on amazon. It really does make a difference.**

https://www.amazon.com/dp/B07FTPKJMM


AI SCIENCES