# 1 Beginning with Machine Learning

**Chapter Objectives**

- ✓ To understand the concept of machine learning and its applications.
- ✓ To understand what are supervised and unsupervised machine learning strategies.
- ✓ To understand the concept of regression and classification.
- ✓ To identify the strategy to be applied to a given problem.

## 1.1 Introduction to Machine Learning

Machine Learning (ML) has emerged as the most extensively used tool for web-sites to classify surfers and address them appropriately. When we surf the Net, we are exposed to machine learning algorithms multiple times a day, often without realizing it. Machine learning is used by search engines such as Google and Bing to rank web pages or to decide which advertisement to show to which user. It is used by social networks such as Facebook and Instagram to generate a custom feed for every user or to tag the user by the picture that was uploaded. It is also used by banks to detect whether an online transaction is genuine or fraudulent and by e-commerce websites such as Amazon and Flipkart to recommend products that we are most likely to buy. Even email providers such as Gmail, Yahoo, and Hotmail use machine learning to decide which emails are spam and which are not. These are only a few examples of applications of machine learning.

The ultimate aim of machine learning is to build an Artificial Intelligence (AI) platform that is as intelligent as the human mind. We are not very far from this dream and many AI researchers believe that this goal can be achieved through machine learning algorithms that try to mimic the learning processes of a human brain.

Actually, ML is a branch of AI. Many years ago researchers tried to build intelligent programs with pre-defined rules like in the case of a normal program. But this approach did not work as there were too many special cases to be considered. For instance, we can define rules to find the shortest

path between two points. But it is very difficult to make rules for programs such as photo tagging, classifying emails as spam or not spam, and web page ranking. The only solution to accomplish these tasks was to write a program that could generate its own rules by examining some examples (also called training data). This approach was named Machine Learning. This book will cover state of art machine learning algorithms and their deployment.

## 1.2 Applications of Machine Learning

Machine Learning or ML is everywhere. It is a definite possibility that one is using it in one way or the other and doesn't even know about it. Some common applications of machine learning that we come across every day are:

### *Virtual Personal Assistants*

There are many Virtual Personal Assistants such as Siri, Alexa, or Google Assistant that we interact with in our daily life. As the term suggests, they help in discovering information, when asked by voice. You have to train them before asking 'What is my calendar for now?', 'How is climate today', or similar inquiries. For answering, the personal assistant searches for the information over the Internet and recalls your related queries, to solve your request. These assistants can be trained for certain tasks like 'Set an alarm for 5 AM next morning', 'Remind me to visit doctor tomorrow at 6 PM', and so on. Smart Speakers like Amazon Echo and Google Home are the outcomes of this innovation. Above-mentioned assistants use machine learning to achieve these objectives.

### *Traffic predictions*

All of us are familiar with Google maps; it uses machine learning to predict the expected time of arrival at the destination and also to model traffic congestion on real time basis.

### *Online transportation networks*

We all book cabs by using mobile apps like Ola and Uber. These apps estimate the price of the ride by using machine learning. They also use ML to determine price surge hours by predicting the rider's demand.

### *Video surveillance*

CCTV cameras have become common for video surveillance. The manual monitoring of these cameras is a very difficult job and boring as well. This is why the idea of training computers to do this job makes sense and machine learning helps to achieve this objective. ML based video surveillance systems can even detect crime before it happens. It can track unusual behavior in people such as standing motionless for an unnaturally long time, hesitancy, or napping on benches and such-like. The system can thus alert human attendants, for suitable response.

### Social media services

Machine learning is also playing a vital role in personalizing news feed in order to better advertisement targeting over social media. Facebook uses machine learning to show news feed to the user based on his or her interests by considering items clicked earlier by that user. Facebook also continuously takes note of the friends that you connect with, the profiles that you often visit, your interests, workplace, and such; and, on the basis of this continuous learning, a list of Facebook users are suggested for you to become friends with.

The Face Recognition feature of Facebook also uses ML to tag the friends in a picture. Facebook checks the poses and projections in the picture, notices the unique features, and then matches them with the people in your friends list. The entire process is performed with the help of ML and is performed so quickly at the backend that it tags the person as soon as we upload his or her picture.

Pinterest also uses ML for computer vision to identify the objects (or pins) in the images and recommend similar pins accordingly.

### Email spam and malware filtering

Email spam and malware filters have inbuilt machine learning to identify spam emails. On the basis of emails we earlier marked as spam or not, the system learns and identifies new mail as spam or not, automatically.

Usually, a malware's code is 90–98% similar to its previous versions. The system security programs that incorporate machine learning understand the coding pattern and detect new malware very efficiently and offer protection against them.

### Online customer support

Many sites these days offer the surfer the option to talk with them. While doing so, these sites have 'bots' hunting the website internals to come up with an appropriate response. In any case, very few sites have a live official to answer your questions. In most cases, you converse with a chatbot. These bots extract data from the site and present it to clients through machine learning.

### Search engine result refining

Google and other search engines use machine learning to improve search results for you. Every time you execute a search, the algorithms at the backend keep a watch on how you respond to the results. If you open the top results and stay on the web page for long, the search engine assumes that the results it displayed were in accordance with the query. Similarly, if you reach the second or third page of the search results but do not open any of the results, the search engine estimates that the results served did not match requirement. This is the way, machine learning trains itself at the backend to improve search results.

### Product recommendations

Whenever we make an online purchase on websites such as Amazon or Flipkart or similar, we usually keep receiving emails from them with shopping suggestions. They also recommend us items that

somehow match our tastes. This improves the shopping experience and again it is machine learning that makes it happen. On the basis of our behavior on the website or app, past purchases, items liked or added to the cart, brand preferences and other such factors, the product recommendations are made.

### Online fraud detection

Machine learning is also helping in making cyberspace more secure and tracking monetary frauds online. For example, Paypal is using ML for protection against money laundering. The company uses a set of ML tools that helps them to compare millions of transactions taking place and distinguish between legitimate or illegitimate transactions taking place between the buyers and sellers.

### Medicine

With the advent of automation, medical records are now available in electronic form. The ML algorithms are helping doctors to understand diseases in a better manner by turning the medical records into medical knowledge.

### Computational biology

Biologists are also collecting an enormous amount of data about human DNA. The ML algorithms are helping them understand and identify the relationship between various genes and related human features.

### Handwriting recognition

ML can not only recognize handwriting but also read different ones. So, it is a versatile tool for many applications. For instance it can be used to route postal mail all over the country once it has been trained to read addresses written in anyone's handwriting.

### Machine translation

We use Google Translate that translates text/website instantly between 100 different human languages as if by magic. Available on our phones and smart watches, the technology behind Google Translate is Machine Translation. It has changed the world by allowing people with different language to communicate with each other.

### Driverless cars and autonomous helicopters

As pointed out, ML is used for programs in situations for which it is very difficult to define rules. One such application is self-driving cars or autonomous helicopters. It takes years of experience for a person to become a good driver and much of this is intuitive. It is very difficult to define this experience in terms of formulating all the myriad rules which are necessary for a program to drive a car on its own. The only possible solution is machine learning, i.e., having a computer program that can learn by itself how to drive a car or fly a helicopter.

With such enormous scope for employing machine learning there is a corresponding huge demand for machine learning experts all over world. It is one of the top ten most required IT skills.

## 1.3 Defining Machine Learning

There are two commonly used definitions of machine learning.

**Arthur Samuel (1959)** coined the term machine learning and defined it as: '*the field of study that gives computers the ability to learn without being explicitly programmed*.' This is an informal and old definition of machine learning.

From this definition, formulated more than half a century ago, it is evident that machine learning is not a new field. But during those times computers were not fast enough to implement the proposed techniques. The concept of machine learning gained popularity in the early 1990s due to the availability of the computers with huge storage space and processing powers.

In 1998, Tom Mitchell redefined the concept of machine learning as '[A] computer program is said to learn from experience E with respect to some class of tasks T and performance measures P, if its performance at tasks in T, as measured by P, improves with experience E.'

For instance, let's say there is an email program which tracks the emails that a person marks as spam or not spam and based on that learns how spam can be filtered in a better way.

Thus, Classification of emails as spam or not spam is the **Task T**.

Tracking the user and marking emails as spam or not spam becomes **Experience E**.

The number of emails correctly classified as spam or not spam is **Performance P**.

## 1.4 Classification of Machine Learning Algorithms

The classification of machine learning algorithms is shown in Figure 1.1 below.

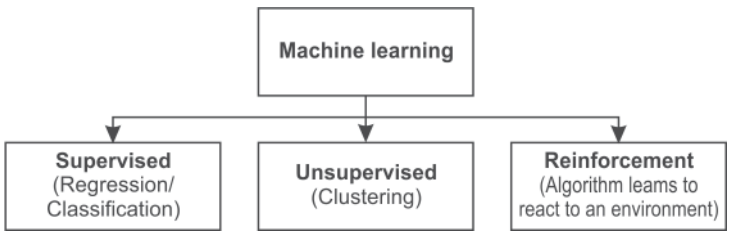


**Figure 1.1** Classification of machine learning algorithms

Some examples are given below for better understanding of these classifications. Further definition will follow in a later section.

### 1.4.1 Supervised learning

Suppose Sonam wants to *accurately* estimate the price of some houses in New Delhi. She first compiles a dataset having covered area of each house with its corresponding price from the city of

New Delhi. She does this for a number of houses of different sizes and plots this data as shown in Figure 1.2. Here, the *Y-axis* represents the price of different houses in lakhs of rupees and the *X-axis* indicates the size of different houses in square feet.

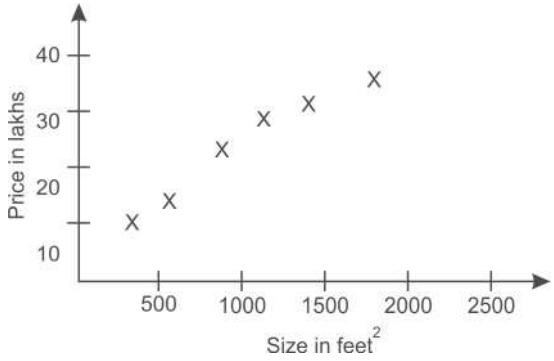


**Figure 1.2** Data plot for size of plot and cost

Let's suppose Sonam's friend owns a house that is 850 square feet in size and she wants to know the selling price for the house. In this case, how can a machine learning algorithm help her? An ML algorithm might be able to do so by fitting a straight line through the plotted data and based on the curve obtained it can be observed that house can be sold about approximately 18 lakh rupees as shown in Figure 1.3. But this is just an opening step.

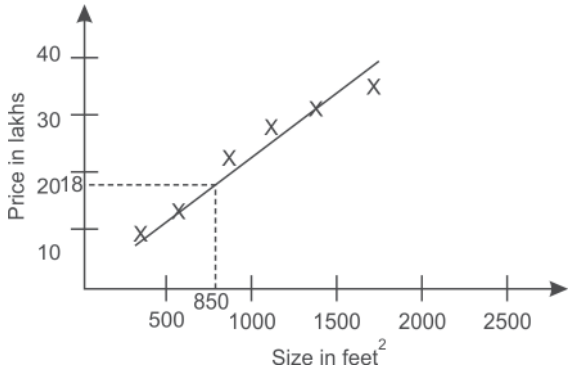


**Figure 1.3** Estimation (prediction) of cost of house with a small dataset

As more and more data is added to this dataset, i.e., as the number of experiences plotted are increased, the graph looks as shown in Figure 1.4. Then, instead of fitting a straight line, a second order polynomial will make a better fit for this data. By using a second order polynomial, we get better prediction and it seems that Sonam's friend should expect a price of close to 20 lakhs as shown in Figure 1.4.

It is important to note that, here, by increasing the dataset, i.e., the experience of the machine, the performance of the system for the task 'predicting house price' has improved. And thus, the machine has learnt how to better estimate or predict the prices of houses. The example discussed

above is an example of supervised machine learning and the term 'supervised' signifies the fact that the dataset with the 'right answers' is given to the algorithm. The example given above is also a case of regression problem. In a regression problem, the system predicts a continuous–valued output (here, it is the price of the house).

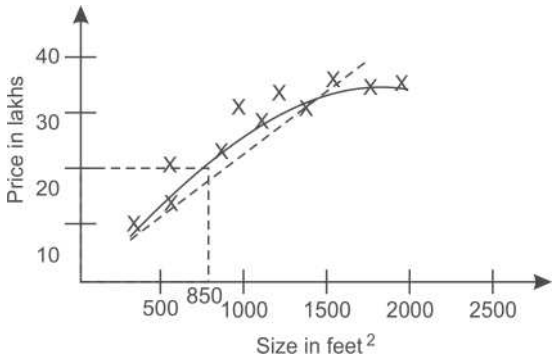


**Figure 1.4** Prediction of cost of house with large dataset

Take another example of supervised learning: suppose a doctor looks at a dataset of medical records to try and find out whether a tumour is benign or malignant. A benign tumor is a harmless tumor and a malignant tumor is a tumor that is dangerous and harmful.

Now the doctor wishes to predict whether a tumor is cancerous or not based on the tumor size. For this, the doctor can collect the data of breast tumor patients and plot a graph in which size of the tumor is on X-axis and type of tumor, i.e, cancerous or not, is on Y-axis as shown in Figure 1.5. In this example, we don't try to predict a continuous value but rather try to classify a tumor as being either benign or malignant.

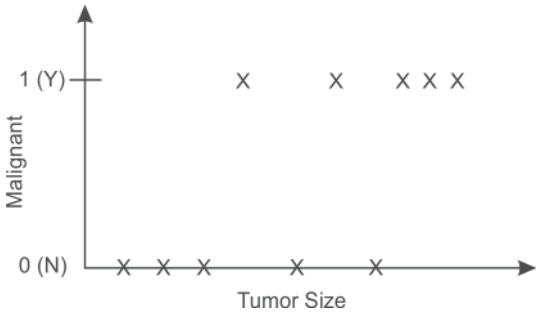


**Figure 1.5** Data plot for tumor size and malignancy

In the above plot, we have five examples of malignant or cancerous tumors having value one and five samples of benign or non-cancerous tumors with a value of zero on the Y-axis. Suppose a person has a breast tumor and the size of the tumor is somewhere around the value marked as A as shown in Figure 1.6. The machine learning question here is to predict the chances of that tumor being benign or malignant?

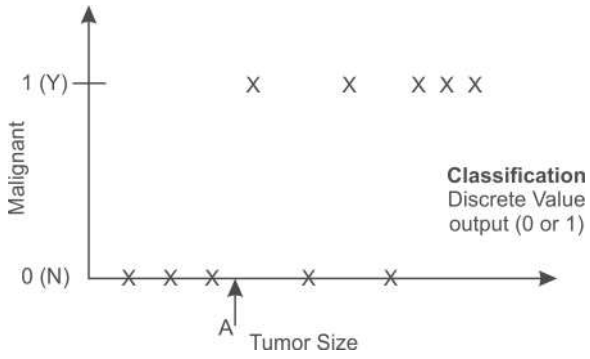


**Figure 1.6** Prediction about a tumor of size A

This then is a '***Classification***' issue. The term 'classification' signifies that the system has to predict the output as a discrete value, i.e., one or zero (either benign or malignant in the above example). It should be noted that in a classification problem, the output can have more than two possible values. For example, there may be three types of breast cancers and one can try to predict the discrete value, i.e., zero, one or two. Here, zero may represent a benign tumor or not harmful cancer, one may represent type one cancer and the discrete value two may indicate a type two cancer. Hence, in a classification problem, we may have N classes in the output where N is always a finite number.

In this example, only one attribute or feature namely the tumor size has been used with the aim to predict whether the type of tumor is benign or malignant. In other machine learning situations, there can be more than one attribute or feature. For example, the age of the patient can also be considered instead of just knowing the size of tumor only. In that case, our dataset would look like as shown in Figure 1.7.

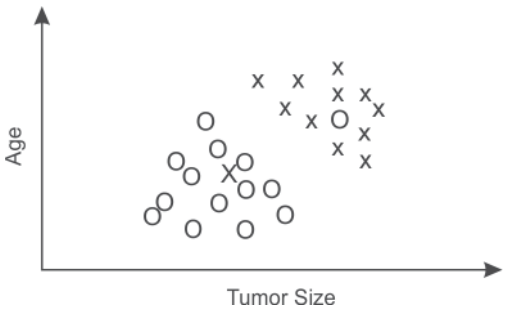


**Figure 1.7** Considering tumor size and age as features for classification

Let's suppose a person has a tumour, of size and age as depicted by B in Figure 1.8.

In this dataset, the ML algorithm is able to fit a straight line to separate out the benign tumors from the malignant tumors as shown in Figure 1.9. Thus, according to ML algorithm, a straight line can, hopefully, help us determine a person's tumor by separating out the tumors. And if a person's tumor falls in this benign area then the type of cancer is more likely to be benign than malignant.

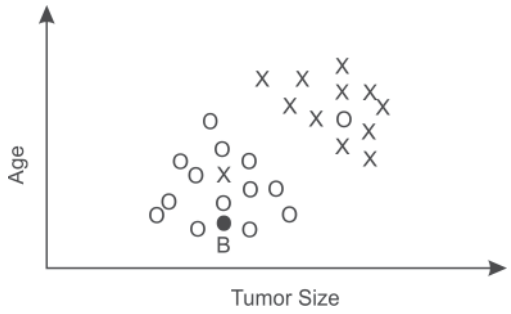


**Figure 1.8** Prediction for a tumor of size B

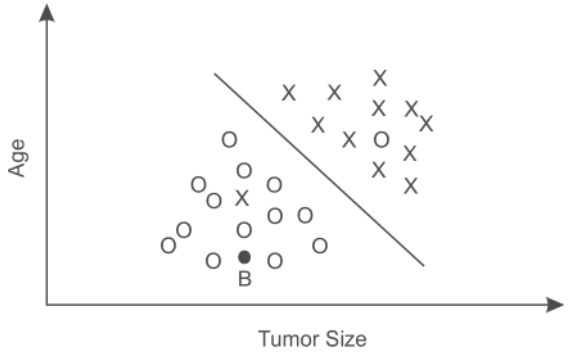


**Figure 1.9** Prediction for tumor size B being benign

In this example, two features such as patient's age and tumor size have been considered. However, we can increase the number of features to further increase the accuracy of prediction. It is important to note that as more and more relevant features are added into a model, the model will become more complex but the accuracy may increase. Commonly, machine learning algorithms have more than one feature.

## Test your progress

Imagine that you own a company and to address each of its problems, you wish to develop a machine learning algorithm.

**Problem 1:** For each customer account, you'd like to have a software to check the account and decide if it has been compromised (hacked).

**Problem 2:** You have a large stock of similar goods. You wish to know that in the next three months how many of these goods will sell.

You have the following options:

(a) Option A: First problem is regression problem and second problem is a classification problem.
(b) Option B: Both are classification problems.

(c) Option C: Both are regression problems.
(d) Option D: First problem is a classification problem and second problem as a regression problem.

The correct answer is option D.

In Problem 1, we have to predict whether the account has been hacked or not. So, if we consider 0 as not hacked and 1 as hacked then we have to predict either 0 or 1. Thus, this is a case of binary classifier (having only two possible outputs). In Problem 2, we have to predict the number of goods that will be sold in next 3 months, so it will be a continuous value, thus it is a case of regression.

Thus, supervised learning problems are categorized into either 'classification' or 'regression' problems. In a regression problem, we try to predict the results within a continuous output. This means that we try to map input variables to some continuous function. In a classification problem instead, we try to predict results in a discrete output. In other words, we try to map input variables into discrete classes or categories.

In the example about predicting the price of houses based on given data about their size: the price as a function of house size is a *continuous* output, so this is a regression problem.

We could turn this example into a classification problem by instead asking whether a house will sell for more than or less than a certain amount. Here, we are classifying the houses based on price into two *discrete* categories.

Further examples should help in better understanding regression and classification. The prediction of marks of a student is a case of regression while prediction about his grades or division is classification. The prediction of a score in a cricket match is an example of regression while to predict if the team will win or lose the match is an example of classification. The prediction about tomorrow's temperature is a case of regression, while to predict whether tomorrow will be cooler or hotter than today is an example of a classification problem.

### 1.4.2 Unsupervised learning

Unsupervised learning, on the other hand, allows us to approach problems with little or no idea about what the results will look like. We can derive structure from data where we don't necessarily know the effect of the variables.

In unsupervised learning data is not labeled, it means that there is no output attribute. We only have input attributes and on the basis of values of input attributes grouping or clustering is performed on the input data to group them into similar classes.

We can only construct the clusters of data based on relationships among the variables in the data. With unsupervised learning, there is no feedback about the results predicted, i.e., there is no teacher to correct you.

#### *Examples of unsupervised learning*

Google news as depicted in Figure 1.10 is an excellent example of clustering that uses unsupervised learning to group news items based on their contents. Google has a collection of millions of news items written on different topics and their clustering algorithm automatically groups these news items into a small number that are somehow similar or related to each other by using different attributes, such as word frequency, sentence length, page count, and so on.