CHAPTER 1

Machine Learning

You easily find examples where the concepts of Machine Learning and Deep Learning are used interchangeably in the media. However, experts generally distinguish them. If you have decided to study this field, it's important you understand what these words actually mean, and more importantly, how they differ.

What occurred to you when you heard the term "Machine Learning" for the first time? Did you think of something that was similar to Figure 1-1? Then you must admit that you are seriously literal-minded.

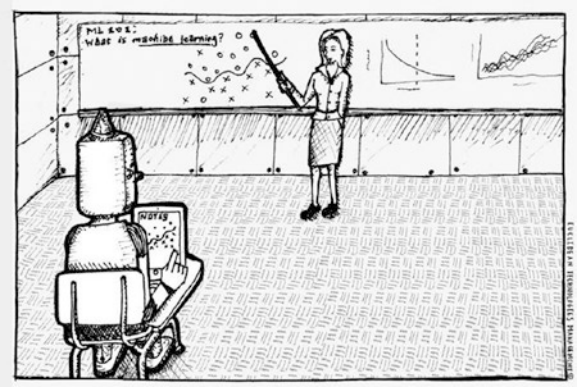

Figure 1-1. *Machine Learning or Artificial Intelligence? Courtesy of Euclidean Technologies Management (www.euclidean.com)*

Figure 1-1 portrays Artificial Intelligence much more than Machine Learning. Understanding Machine Learning in this way will bring about serious confusion. Although Machine Learning is indeed a branch of Artificial Intelligence, it conveys an idea that is much different from what this image may imply.

P. Kim, *MATLAB Deep Learning*, DOI 10.1007/978-1-4842-2845-6_1

In general, Artificial Intelligence, Machine Learning, and Deep Learning are related as follows:

> "Deep Learning is a kind of Machine Learning, and
> Machine Learning is a kind of Artificial Intelligence."

How is that? It's simple, isn't it? This classification may not be as absolute as the laws of nature, but it is widely accepted.

Let's dig into it a little further. Artificial Intelligence is a very common word that may imply many different things. It may indicate any form of technology that includes some intelligent aspects rather than pinpoint a specific technology field. In contrast, Machine Learning refers to a specific field. In other words, we use Machine Learning to indicate a specific technological group of Artificial Intelligence. Machine Learning itself includes many technologies as well. One of them is Deep Learning, which is the subject of this book.

The fact that Deep Learning is a type of Machine Learning is very important, and that is why we are going through this lengthy review on how Artificial Intelligence, Machine Learning, and Deep Learning are related. Deep Learning has been in the spotlight recently as it has proficiently solved some problems that have challenged Artificial Intelligence. Its performance surely is exceptional in many fields. However, it faces limitations as well. The limitations of Deep Learning stems from its fundamental concepts that has been inherited from its ancestor, Machine Learning. As a type of Machine Learning, Deep Learning cannot avoid the fundamental problems that Machine Learning faces. That is why we need to review Machine Learning before discussing the concept of Deep Learning.

What Is Machine Learning?

In short, Machine Learning is a modeling technique that involves data. This definition may be too short for first-timers to capture what it means. So, let me elaborate on this a little bit. Machine Learning is a technique that figures out the "model" out of "data." Here, the data literally means information such as documents, audio, images, etc. The "model" is the final product of Machine Learning.

Before we go further into the model, let me deviate a bit. Isn't it strange that the definition of Machine Learning only addresses the concepts of data and model and has nothing to do with "learning"? The name itself reflects that the technique analyzes the data and finds the model by itself rather than having a human do it. We call it "learning" because the process resembles being trained with the data to solve the problem of finding a model. Therefore, the data that Machine Learning uses in the modeling process is called "training" data. Figure 1-2 illustrates what happens in the Machine Learning process.

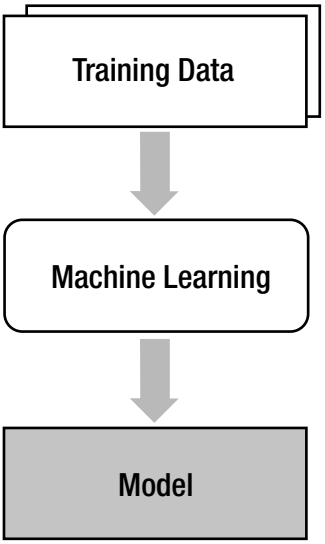

Figure 1-2. *What happens during the machine learning process*

Now, let's resume our discussion about the model. Actually, the model is nothing more than what we want to achieve as the final product. For instance, if we are developing an auto-filtering system to remove spam mail, the spam mail filter is the model that we are talking about. In this sense, we can say the model is what we actually use. Some call the model a *hypothesis*. This term seems more intuitive to those with statistical backgrounds.

Machine Learning is not the only modeling technique. In the field of dynamics, people have been using a certain modeling technique, which employs Newton's laws and describes the motion of objects as a series of equations called equations of motion, for a long time. In the field of Artificial Intelligence, we have the expert system, which is a problem-solving model that is based on the knowledge and know-how of the experts. The model works as well as the experts themselves.

However, there are some areas where laws and logical reasoning are not very useful for modeling. Typical problems can be found where intelligence is involved, such as image recognition, speech recognition, and natural language processing. Let me give you an example. Look at Figure 1-3 and identify the numbers.

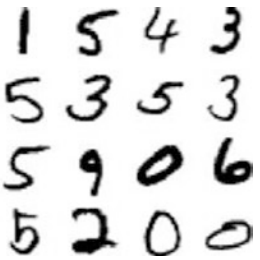

Figure 1-3. *How does a computer identify numbers when they have no recognizable pattern?*

I'm sure you have completed the task in no time. Most people do. Now, let's make a computer do the same thing. What do we do? If we use a traditional modeling technique, we will need to find some rule or algorithm to distinguish the written numbers. Hmm, why don't we apply the rules that you have just used to identify the numbers in your brain? Easy enough, isn't it? Well, not really. In fact, this is a very challenging problem. There was a time when researchers thought it must be a piece of cake for computers to do this, as it is very easy for even a human and computers are able to calculate much faster than humans. Well, it did not take very long until they realized their misjudgment.

How were you able to identify the numbers without a clear specification or a rule? It is hard to answer, isn't it? But, why? It is because we have never learned such a specification. From a young age, we have just learned that this is 0, and that this is 1. We just thought that's what it is and became better at distinguishing numbers as we faced a variety of numbers. Am I right?

What about computers, then? Why don't we let computers do the same thing? That's it! Congratulations! You have just grasped the concept of Machine Learning. Machine Learning has been created to solve the problems for which analytical models are hardly available. The primary idea of Machine Learning is to achieve a model using the training data when equations and laws are not promising.

Challenges with Machine Learning

We just discovered that Machine Learning is the technique used to find (or learn) a model from the data. It is suitable for problems that involve intelligence, such as image recognition and speech recognition, where physical laws or mathematical equations fail to produce a model. On the one hand, the approach that Machine Learning uses is what makes the process work. On the other hand, it brings inevitable problems. This section provides the fundamental issues Machine Learning faces.

Once the Machine Learning process finds the model from the training data, we apply the model to the actual field data. This process is illustrated in Figure 1-4. The vertical flow of the figure indicates the learning process, and the trained model is described as the horizontal flow, which is called inference.

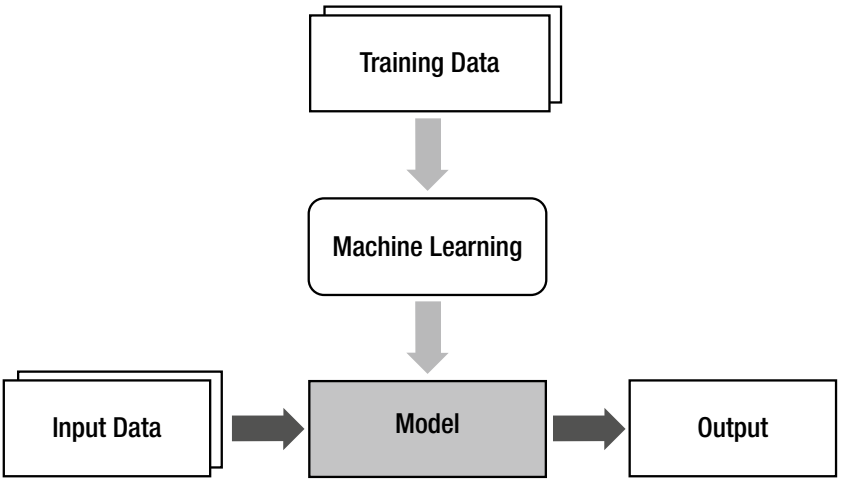

Figure 1-4. *Applying a model based on field data*

The data that is used for modeling in Machine Learning and the data supplied in the field application are distinct. Let's add another block to this image, as shown in Figure 1-5, to better illustrate this situation.

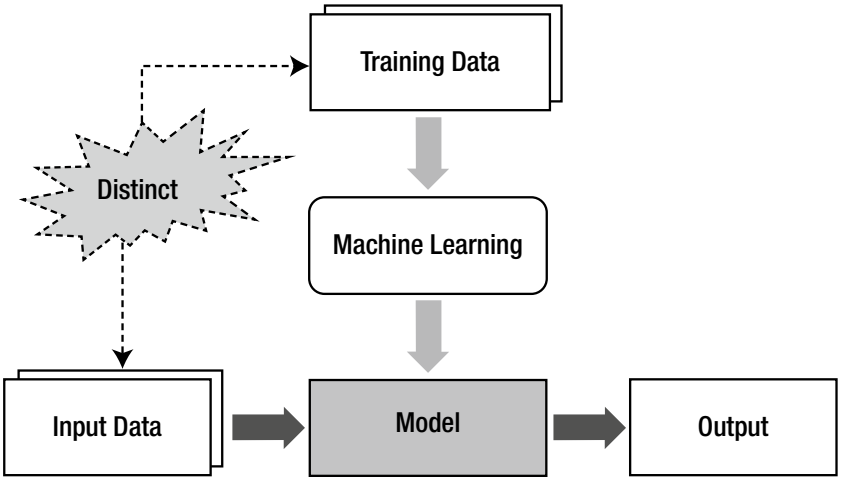

Figure 1-5. *Training and input data are sometimes very distinct*

The distinctness of the training data and input data is the structural challenge that Machine Learning faces. It is no exaggeration to say that every problem of Machine Learning originates from this. For example, what about using training data, which is composed of handwritten notes from a single person? Will the model successfully recognize the other person's handwriting? The possibility will be very low.

No Machine Learning approach can achieve the desired goal with the wrong training data. The same ideology applies to Deep Learning. Therefore, it is critical for Machine Learning approaches to obtain unbiased training data that adequately reflects the characteristics of the field data. The process used to make the model performance consistent regardless of the training data or the input data is called *generalization*. The success of Machine Learning relies heavily on how well the generalization is accomplished.

Overfitting

One of the primary causes of corruption of the generalization process is *overfitting*. Yes, another new term. However, there is no need to be frustrated. It is not a new concept at all. It will be much easier to understand with a case study than with just sentences.

Consider a classification problem shown in Figure 1-6. We need to divide the position (or coordinate) data into two groups. The points on the figure are the training data. The objective is to determine a curve that defines the border of the two groups using the training data.

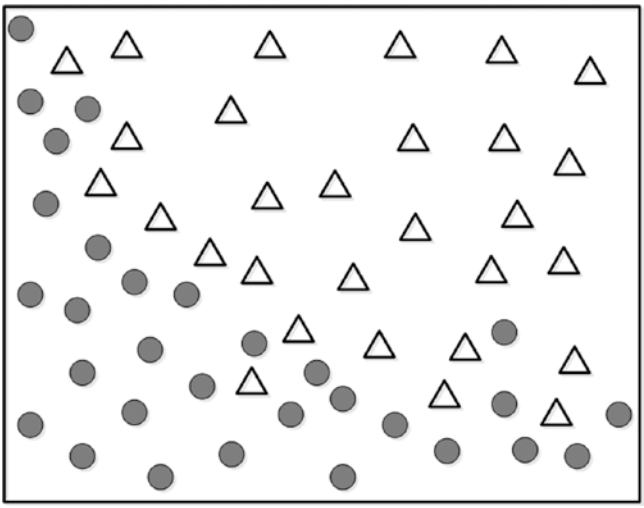

Figure 1-6. *Determine a curve to divide two groups of data*

Although we see some outliers that deviate from the adequate area, the curve shown in Figure 1-7 seems to act as a reasonable border between the groups.

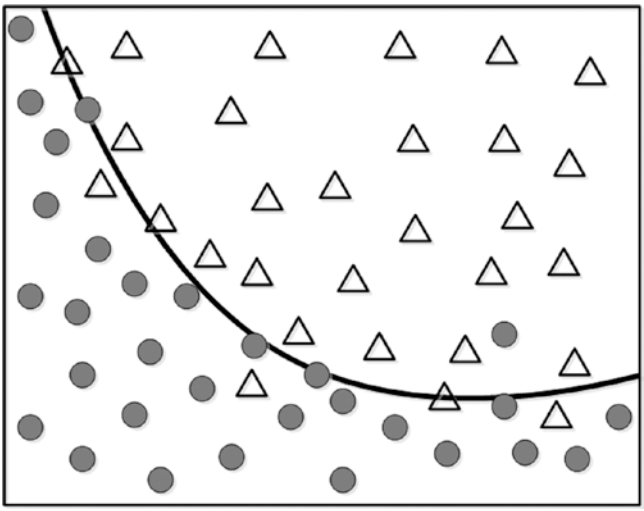

Figure 1-7. *Curve to differentiate between two types of data*

When we judge this curve, there are some points that are not correctly classified according to the border. What about perfectly grouping the points using a complex curve, as shown in Figure 1-8?

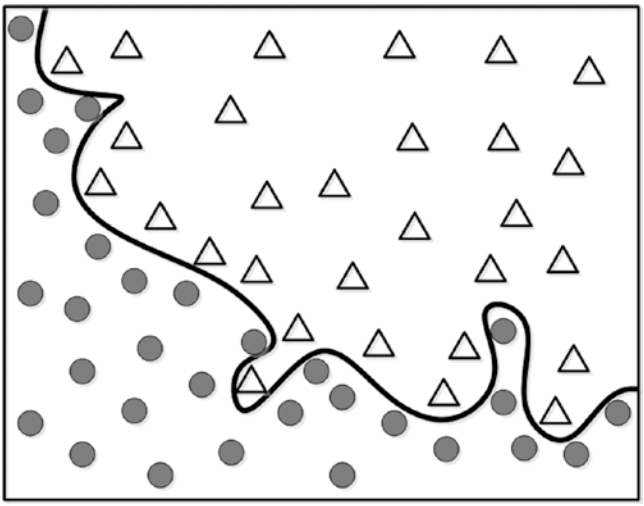

Figure 1-8. *Better grouping, but at what cost?*

This model yields the perfect grouping performance for the training data. How does it look? Do you like this model better? Does it seem to reflect correctly the general behavior?

Now, let's use this model in the real world. The new input to the model is indicated using the symbol ■, as shown in Figure 1-9.

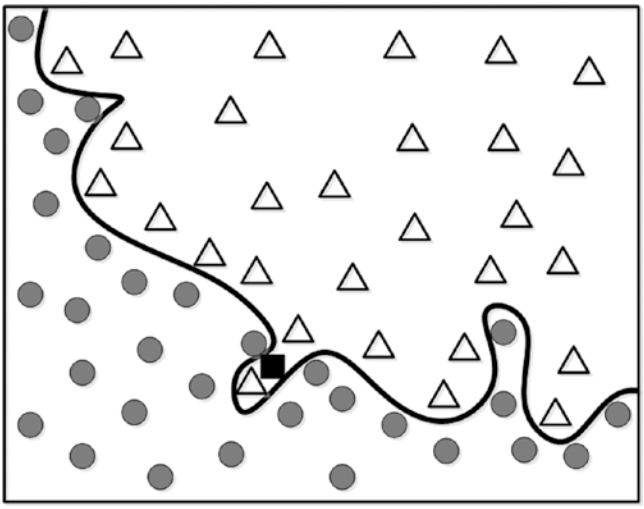

Figure 1-9. *The new input is placed into the data*

This proud error-free model identifies the new data as a class Δ. However, the general trend of the training data tells us that this is doubtable. Grouping it as a class ● seems more reasonable. What happened to the model that yielded 100% accuracy for the training data?

Let's take another look at the data points. Some outliers penetrate the area of the other group and disturb the boundary. In other words, this data contains much noise. The problem is that there is no way for Machine Learning to distinguish this. As Machine Learning considers all the data, even the noise, it ends up producing an improper model (a curve in this case). This would be penny-wise and pound-foolish. As you may notice here, the training data is not perfect and may contain varying amounts of noise. If you believe that every element of the training data is correct and fits the model precisely, you will get a model with lower generalizability. This is called *overfitting*.

Certainly, because of its nature, Machine Learning should make every effort to derive an excellent model from the training data. However, a working model of the training data may not reflect the field data properly. This does not mean that we should make the model less accurate than the training data on purpose. This will undermine the fundamental strategy of Machine Learning.

Now we face a dilemma—reducing the error of the training data leads to overfitting that degrades generalizability. What do we do? We confront it, of course! The next section introduces the techniques that prevent overfitting.

Confronting Overfitting

Overfitting significantly affects the level of performance of Machine Learning. We can tell who is a pro and who is an amateur by watching their respective approaches in dealing with overfitting. This section introduces two typical methods used to confront overfitting: regularization and validation.

Regularization is a numerical method that attempts to construct a model structure as simple as possible. The simplified model can avoid the effects of overfitting at the small cost of performance. The grouping problem of the previous section can be used as a good example. The complex model (or curve) tends to be overfitting. In contrast, although it fails to classify correctly some points, the simple curve reflects the overall characteristics of the group much better. If you understand how it works, that is enough for now. We will revisit regularization with further details in Chapter Three's "Cost Function and Learning Rule" section.

We are able to tell that the grouping model is overfitted because the training data is simple, and the model can be easily visualized. However, this is not the case for most situations, as the data has higher dimensions. We cannot draw the model and intuitively evaluate the effects of overfitting for such data. Therefore, we need another method to determine whether the trained model is overfitted or not. This is where *validation* comes into play.

The validation is a process that reserves a part of the training data and uses it to monitor the performance. The validation set is not used for the training process. Because the modeling error of the training data fails to indicate overfitting, we use some of the training data to check if the model is overfitted. We can say that the model is overfitted when the trained model yields a low level of performance to the reserved data input. In this case, we will modify the model to prevent the overfitting. Figure 1-10 illustrates the division of the training data for the validation process.

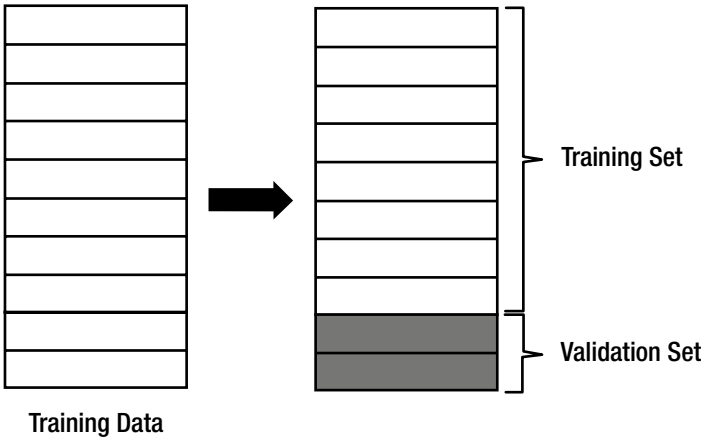

Figure 1-10. *Dividing the training data for the validation process*

When validation is involved, the training process of Machine Learning proceeds by the following steps:

1. Divide the training data into two groups: one for training and the other for validation. As a rule of thumb, the ratio of the training set to the validation set is 8:2.
2. Train the model with the training set.
3. Evaluate the performance of the model using the validation set.
    a. If the model yields satisfactory performance, finish the training.
    b. If the performance does not produce sufficient results, modify the model and repeat the process from Step 2.

Cross-validation is a slight variation of the validation process. It still divides the training data into groups for the training and validation, but keeps changing the datasets. Instead of retaining the initially divided sets, cross-validation repeats the division of the data. The reason for doing this is that the model can be overfitted even to the validation set when it is fixed. As the cross-validation maintains the randomness of the validation dataset, it can better detect the overfitting of the model. Figure 1-11 describes the concept of cross-validation. The dark shades indicate the validation data, which is randomly selected throughout the training process.

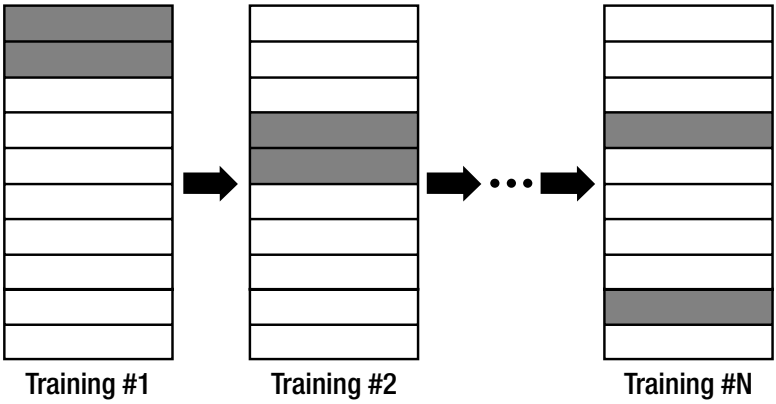

Figure 1-11. *Cross-validation*

Types of Machine Learning

Many different types of Machine Learning techniques have been developed to solve problems in various fields. These Machine Learning techniques can be classified into three types depending on the training method (see Figure 1-12).

- Supervised learning
- Unsupervised learning
- Reinforcement learning

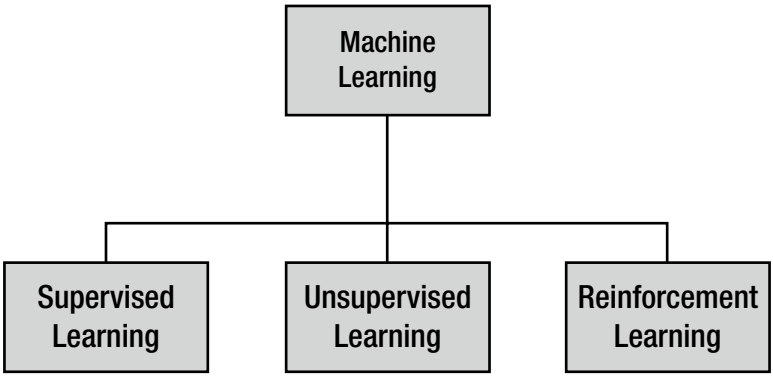

Figure 1-12. *Three types of Machine Learning techniques*

Supervised learning is very similar to the process in which a human learns things. Consider that humans obtain new knowledge as we solve exercise problems.

1. Select an exercise problem. Apply current knowledge to solve the problem. Compare the answer with the solution.
2. If the answer is wrong, modify current knowledge.
3. Repeat Steps 1 and 2 for all the exercise problems.

When we apply an analogy between this example and the Machine Learning process, the exercise problems and solutions correspond to the training data, and the knowledge corresponds to the model. The important thing is the fact that we need the solutions. This is the vital aspect of the supervised learning. Its name even implies the tutoring in which the teacher gives solutions to the students to memorize.

In supervised learning, each training dataset should consist of input and correct output pairs. The correct output is what the model is supposed to produce for the given input.

```
{ input, correct output }
```

Learning in supervised learning is the series of revisions of a model to reduce the difference between the correct output and the output from the model for the same input. If a model is perfectly trained, it will produce a correct output that corresponds to the input from the training data.

In contrast, the training data of the unsupervised learning contains only inputs without correct outputs.

```
{ input }
```

At a first glance, it may seem difficult to understand how to train without correct outputs. However, many methods of this type have been developed already. Unsupervised learning is generally used for investigating the characteristics of the data and preprocessing the data. This concept is similar to a student who just sorts out the problems by construction and attribute and doesn't learn how to solve them because there are no known correct outputs.

Reinforcement learning employs sets of input, some output, and grade as training data. It is generally used when optimal interaction is required, such as control and game plays.

```
{ input, some output, grade for this output }
```

This book only covers supervised learning. It is used for more applications compared to unsupervised learning and reinforcement learning, and more importantly, it is the first concept you will study when entering the world of Machine Learning and Deep Learning.

Classification and Regression

The two most common types of application of supervised learning are classification and regression. These words may sound unfamiliar, but are actually not so challenging.

Let's start with classification. This may be the most prevailing application of Machine Learning. The classification problem focuses on literally finding the classes to which the data belongs. Some examples may help.

Spam mail filtering service ➔ Classifies the mails by regular or spam

Digit recognition service ➔ Classifies the digit image into one of 0-9

Face recognition service ➔ Classifies the face image into one of the registered users

We addressed in the previous section that supervised learning requires input and correct output pairs for the training data. Similarly, the training data of the classification problem looks like this:

```
{ input, class }
```

In the classification problem, we want to know which class the input belongs to. So the data pair has the class in place of the correct output corresponding to the input.

Let's proceed with an example. Consider the same grouping problem that we have been discussing. The model we want Machine Learning to answer is which one of the two classes (Δ and ●) does the user's input coordinate (X, Y) belong (see Figure 1-13).

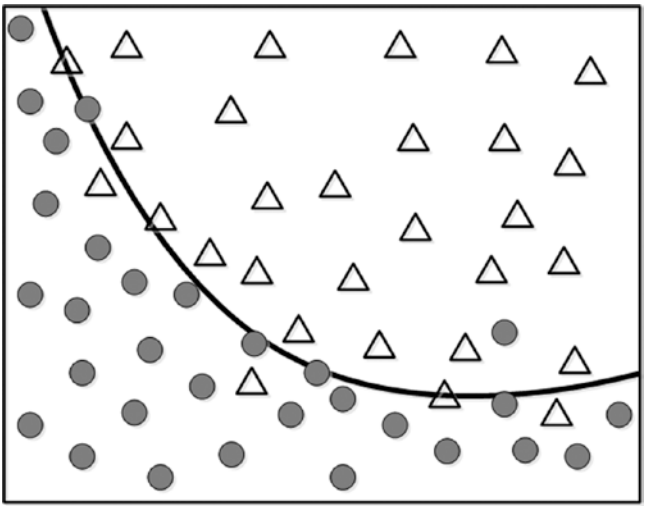

Figure 1-13. *Same data viewed from the perspective of classification*

In this case, the training data of N sets of the element will look like Figure 1-14.

$\{X_1, Y_1, \triangle\}$
$\{X_2, Y_2, \bullet\}$
...
$\{X_N, Y_N, \bullet\}$

Figure 1-14. *Classifying the data*

In contrast, the regression does not determine the class. Instead, it estimates a value. As an example, if you have datasets of age and income (indicated with a ●) and want to find the model that estimates income by age, it becomes a regression problem (see Figure 1-15).[1]

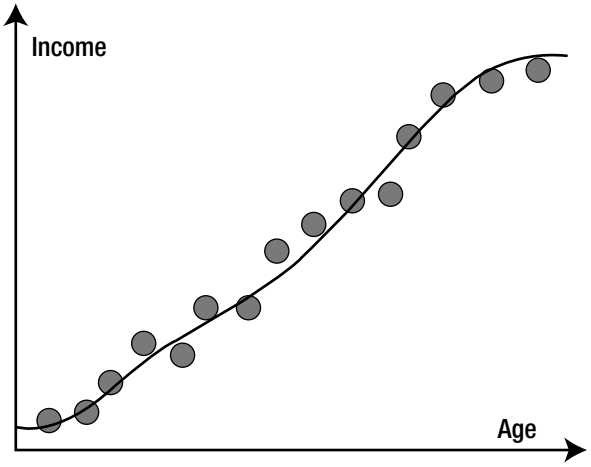

Figure 1-15. *Datasets of age and income*

The dataset of this example will look like the table in Figure 1-16, where X and Y are age and income, respectively.

[1]The original meaning of the word "regress" is to go back to an average. Francis Galton, a British geneticist, researched the correlation of the height of parents and children and found out that the individual height converged to the average of the total population. He named his methodology "regression analysis."

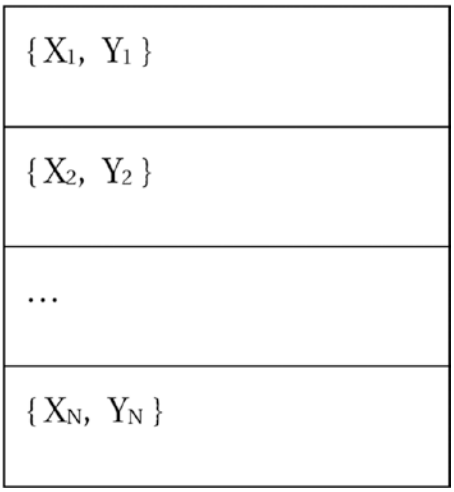

Figure 1-16. *Classifying the age and income data*

Both classification and regression are parts of supervised learning. Therefore, their training data is equally in the form of `{input, correct output}`. The only difference is the type of correct outputs—classification employs classes, while the regression requires values.

In summary, analysis can become classification when it needs a model to judge which group the input data belongs to and regression when the model estimates the trend of the data.

Just for reference, one of the representative applications of unsupervised learning is *clustering*. It investigates the characteristics of the individual data and categorizes the related data. It is very easy to confuse clustering and classification, as their results are similar. Although they yield similar outputs, they are two completely different approaches. We have to keep in mind that clustering and classification are distinct terms. When you encounter the term *clustering*, just remind yourself that it focuses on unsupervised learning.

Summary

Let's briefly recap what we covered in this chapter:

- Artificial Intelligence, Machine Learning, and Deep Learning are distinct. But they are related to each other in the following way: "Deep Learning is a kind of Machine Learning, and Machine Learning is a kind of Artificial Intelligence".

- Machine Learning is an inductive approach that derives a model from the training data. It is useful for image recognition, speech recognition, and natural language processing etc.
- The success of Machine Learning heavily relies on how well the generalization process is implemented. In order to prevent performance degradation due to the differences between the training data and actual input data, we need a sufficient amount of unbiased training data.
- Overfitting occurs when the model has been overly customized to the training data that it yields poor performance for the actual input data, while its performance for the training data is excellent. Overfitting is one of the primary factors that reduces the generalization performance.
- Regularization and validation are the typical approaches used to solve the overfitting problem. Regularization is a numerical method that yields the simplest model as possible. In contrast, validation tries to detect signs of overfitting during training and takes action to prevent it. A variation of validation is cross-validation.
- Depending on the training method, Machine Learning can be supervised learning, unsupervised learning, and reinforcement learning. The formats of the training data for theses learning methods are shown here.

Training Method	Training Data
Supervised Learning	`{ input, correct output }`
Unsupervised Learning	`{ input }`
Reinforced Learning	`{ input, some output, grade for this output }`

- Supervised learning can be divided into classification and regression, depending on the usage of the model. Classification determines which group the input data belongs to. The correct output of the classification is given as categories. In contrast, regression predicts values and takes the values for the correct output in the training data.