Individual Assignment MBD 2022

Mohammad Hadi Alipour Motlagh

3/13/2022

Introduction

Machine Learning Models

In machine learning (Machine Learning), the subject is the design of machines that learn from the examples given to them and their own experiences. In fact, in this science, an attempt is made to design a machine using algorithms in such a way that it can learn and operate without explicitly planning and dictating each action. In machine learning, instead of programming everything, the data is given to a general algorithm, and it is this algorithm that builds its logic based on the data given to it. Machine learning has a variety of methods, including supervised, unsupervised, and reinforcement learning. The algorithms used in machine learning fall into these three categories. Machine learning is the subject of studies that have emerged from artificial intelligence. Humans use artificial intelligence to build better and smarter machines. But researchers were initially unable to program machines to perform more complex tasks that are constantly challenging, except for a few simple tasks, such as finding the shortest path between points A and B. Accordingly, the perception was formed that the only possible way to achieve this was to design machines that could learn from themselves. In this approach, the machine is like a child learning from itself. Thus, machine learning was introduced as a new capability for computers. Today, this science is used in various fields of technology, and its use has become so widespread that people are often unaware of its existence in their daily tools and accessories.

There are three types of machine learning algorithms:

Supervised learning

Most machine learning methods use supervised learning. In supervised machine learning, the system tries to learn from the prior examples provided. In other words, in this type of learning, the system tries to learn the patterns based on the examples given to it.

Mathematically speaking, when input variable (X) and output variable (Y) are present and an algorithm can be used to derive an input-to-output mapping function based on them, learning is actually supervised. The mapping function is represented by $(Y = f(X))$.

Unsupervised learning

In unsupervised learning, the algorithm alone must look for interesting structures in the data. Mathematically speaking, unsupervised learning refers to when there are only input variables (X) in the data set and no output data variables. This type of learning is called unsupervised because unlike supervised learning, there is no correct answer given and the machine itself must look for the answer.

Reinforcement learning

A computer program that interacts with a dynamic environment must achieve a specific goal (such as playing with a competitor or driving a car). The program provides feedback on rewards and punishments and directs the issue accordingly. Using reinforcement learning, the machine learns to make specific decisions in an environment that is constantly subject to trial and error.

Supervised learning

As mentioned, supervised learning methods are more popular than other algorithms. In general, these algorithms can be divided into two main categories according to the type of objective variable, which include regression and classification. In regression, we seek to predict a range of data, such as forecasting sales, or revenue and expenditure. While in classification issues the target variable includes two or more limited options such as predicting customer exit or non-exit, profit or loss, fraud. If we want to study these problems mathematically, we can say that in the region problems, we seek to optimize the loss function by minimizing the cost in this function, which are briefly presented under the relevant functions.

In classification problems, however, we seek to maximize the probability function so that we can find a function that optimally separates the two groups. The following is probably the function.

Given the type of database used in this project, we are looking to find a function to predict credit card fraud. Therefore, this problem falls into the category of classification algorithms that use supervised learning. Therefore, in the following, I will review the 5 algorithms used in this field and then I will explain how to implement and measure the performance of each of these algorithms and on the data.

Logistic regression

Generalized Aditive Models

Decision Trees

Random Forest

Suppor Vector Machines

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.