Individual Assignment

Statistical & Machine Learning for Marketing

Student

Mohammad Hadi Alipour Motlagh

MBD IESEG 2021-2022

Contents

[Introduction 3](#_Toc98845708)

[Machine Learning Models 3](#_Toc98845709)

[Supervised learning 3](#_Toc98845710)

[Unsupervised learning 3](#_Toc98845711)

[Reinforcement learning 3](#_Toc98845712)

[Supervised learning 4](#_Toc98845713)

[Logistic regression 4](#_Toc98845714)

[KNN 6](#_Toc98845715)

[Decision Trees 6](#_Toc98845716)

[Random Forest 9](#_Toc98845717)

[Support Vector Machines 10](#_Toc98845718)

[Application of Models 11](#_Toc98845719)

[Data description 11](#_Toc98845720)

[Data Manipulation 11](#_Toc98845721)

[Feature Selection 11](#_Toc98845722)

[Performance of Models 11](#_Toc98845723)

[Concolusion 12](#_Toc98845724)

[Refrences 12](#_Toc98845725)

# 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

One of the methods of "classification" in the topic of "supervised machine learning" (Supervised Machine Learning) is logistic regression. In this regression method, the concept and method of calculating the "odds ratio" (Odds Ratio) is used. Therefore, it is better to get acquainted with this concept first.

Since in the previous section the prediction value for the dependent variable, with probability

p

(

x

)

Done, to determine the model of the relationship between the dependent and independent variables instead of the linear relationship, we need a function that changes from about 0 to 1. In logistic regression method, a function called "Logistic Function" is used. For this reason, this regression method is called logistic regression. In the continuation of this function, the introduction and the related diagram based on the parameters

b

1

=

1

,

b

0

=

0

Can be seen in the image.

As can be seen by increasing the value of x (

x

→

∞

) The logistics function will be close to 1. Also by reducing the value of x (

x

→

-

∞

The value of the function tends to zero. Now suppose this function is used for logistic regression to express the probability of a dependent variable. So we will have:

In order to estimate the parameters of this model, "Logit Transformation" can be used. This conversion is on luck

p

(

x

)

1

-

P

(

x

)

As stated earlier, we execute. In this case, the relationship can be written as follows:

## KNN

## Decision Trees

In general, decision tree analysis is a predictive modeling tool that can be used in many fields. Decision trees can be created using an algorithmic solution that can differentiate data sets based on different conditions in different ways. Decision trees are one of the most powerful algorithms that are considered as a subset of supervised algorithms.

They can be used for both classification and regression tasks. The two main elements of a tree are the decision nodes where the data is distributed, and the leaves from which we get the output. The following is an example of a binary tree that provides a variety of information.

Gini Index

It is the name of a cost function used to evaluate the binary separation of a data set, and works with the definite objective variables "success" and "failure".

The higher the Gini index, the higher the homogeneity. The ideal value for the Gini index is 0, and the worst value for a problem with two classes is 0.5. Using the following steps, the Gini index can be calculated for a breakdown.

First, calculate the Gini index for the sub-nodes using the formula p ^ 2 + q ^ 2, which is the sum of the squares of the probabilities of success and failure. Then, use the weighted Gini score for each node to calculate the Gini index.

The classification and regression tree (CART) algorithm generates binary segregation using the Gini method.

A subdivision basically contains an attribute in the data set and a value. With the help of the following three steps, a separation can be created in the data set.

Step 1: ‌ ‌

Calculate Gini Score

We talked about this in the previous section.

Step 2: ‌ ‌

Separation of a data set

It can be defined as separating a data set into two lists of rows that have an index of a property, and a split value of that property. After taking the two groups of right and left from the data set, we can use the Gini score, which was calculated in the first step, to calculate the separation value. The value of the separation will decide which group the attribute will remain in.

Step 3: ‌ ‌

Evaluation of segregations

The next step after finding the Gini score and sorting the data set is to evaluate all the sorts. To this end, we must first consider all the values ​​associated with each attribute as a separation candidate. Then, we need to estimate the cost of segregation to find the best possible segregation. The best separation is used as a node in the decision tree.

Build a tree: ‌

As we know, a tree has a root node and end nodes. After making the root knot, we can make the tree by following the two parts below.

Section 1: ‌ ‌

Final node construction:

When making the final nodes of a decision tree, deciding when to complete the growth of the tree or making more end nodes is one of the most important points. This can be done using two criteria called maximum tree depth and minimum node records, which are introduced below.

Maximum Tree Depth:

As the name implies, this is the maximum number of nodes in a tree after the root node. When a tree reaches its maximum depth, we must stop adding end nodes. For example, when a tree has the maximum number of end nodes.

Minimum Node Records:

Defined as the minimum number of learning patterns for which a given node is responsible. When the tree reaches these minimum node records or less, we must stop adding end nodes.

The final node is used to make the final prediction.

Section 2: ‌ ‌

Recursive splitting:

Now that we know when to make the final knots, we can start building our own tree. Return separation is a method of building a tree. In this method, when a node is created, we can call the child nodes (nodes added to an existing node) in each data group, which is generated by separating the data set, by calling the same thing over and over again. Create the function recursively.

Prediction:

After building a decision tree, we need to make a prediction about it. Basically, forecasting involves guiding the decision tree by providing specific rows of data.

According to the above, a prediction can be made with the help of the recursive function. This similar prediction procedure is called again for the left and right child nodes.

assumptions

The following are some of the assumptions made at the time of making the decision tree.

When preparing decision trees, the learning set will be the root node.

Decision tree classifiers prefer to have definite attribute values. If we intend to use continuous values, we must discretize them before constructing the model.

Records are distributed recursively based on attribute values.

Statistical methods are used to place attributes at the location of each node - such as the root node or intermediate nodes.

## Random Forest

Random forest is basically like a bag containing n decision trees that have different sets and are taught under different data sets. Suppose we decide to have 100 trees in a random forest !! As I said, these decision trees have a different set of extraordinary parameters and a subset of different training data, so the decision or prediction provided by these trees can be very different.

Let's say we somehow train all 100 of these trees with their data sets. All 100 existing trees are now being examined for their predictions about test data. Now we just have to make a decision about an example or a test data, we do it by simple voting. We accept as a result what most trees have predicted for that example.

The Random Forest random forest algorithm is essentially a supervised learning algorithm. This can be used for both regression and classification tasks. But in this article we will discuss its use for random Forrest classification because it is more intuitive and easier to understand. Random forest is one of the most widely used algorithms due to its simplicity and stability. The word "random" is displayed when constructing a subset of data for trees. A subset of data is created by randomly selecting x the number of attributes (columns) and the number of y samples (rows) from the original set of n attributes and m examples.

Random forests are more stable and reliable than a decision tree. This is just an interpretation of this sentence: "It is better to get a vote from all cabinet ministers than to accept only the decision made by the prime minister." As we have seen, random forests are nothing more than a collection of decision trees, so knowing the decision tree is essential.

In the following, we will examine the Random Forest random forest algorithm. The first two phases are the creation of a random forest with a combination of N decision trees and the second phase is the prediction for each tree created in the first phase. The work process can be explained in the following steps and diagrams:

Step 1: Select random K data points from the tutorial set.

Step 2: Create decision trees related to selected data points (subsets).

Step 3: Select the N number for the decision trees you want to create.

Step 4: Repeat steps 1 and 2.

Step 5: For the new data points, find the predictions of each decision tree and assign the new data points to the group that gets the majority of votes.

## Support Vector Machines

SVM is basically a binary separator. In the previous section, the theoretical foundations of support vector machines for classifying two classes were described. A multi-class pattern recognition can be achieved by combining two-class support vector machines. There are usually two perspectives on this. One is the one-on-one strategy for categorizing each pair of classes and the remaining classes. Another one-on-one strategy is to categorize each pair. In situations where the first classification leads to ambiguous classification. For multi-class problems, the general approach is to reduce the multi-class problem to multiple binary problems. Each problem is solved with a binary separator. The output of the SVM binary separators is then combined to solve a multi-class problem.

Support vector machines (SVMs) are one of the supervised learning methods [1] used for classification [2] and regression [3].

This method is one of the relatively new methods that has shown good performance in recent years compared to older methods for classification. The basis of the SVM classifier is linear classification of data, and in linear segmentation of data we try to choose the line that has the most reliable margin. Solving the optimal line finding equation for data is done by QP methods, which are known methods for solving constrained problems. Before linear division, in order for the machine to be able to categorize high-complexity data, we move the data to a much larger space by means of the phi function [4]. In order to be able to solve the problem of very large dimensions using these methods, we use the Lagrangian duality theorem [5] to convert the problem of minimization into a dual form in which instead of the complex phi function that elevates us to a space with dimensions, We use a simpler function called the kernel function, which is a multiplication of the phi function. Various kernel functions can be used, including exponential, polynomial, and sigmoid kernels.

Benefits and nega

Training is relatively simple.

Unlike neural networks, it is not trapped in local maxima. It works almost perfectly for high-dimensional data. The trade-off between classifier complexity and error rate is clearly controlled. Requires a good kernel function and C parameter selection.

# Application of Models

## Data description

## Data Manipulation

## Feature Selection

## Performance of Models

# Concolusion

# Refrences

1. Wikipedia pages related to selected Models, and Macine Learning.
2. Course Book.
3. The Codes of Book and codes of theat provided in the course.