

Supervised Machine Learning

Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output.

In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly. It applies the same concept as a student learns in the supervision of the teacher.

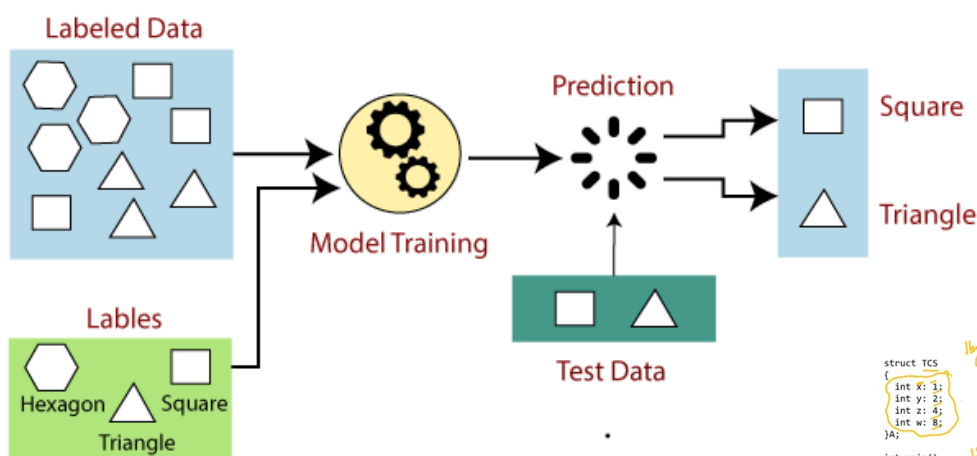
Supervised learning is a process of providing input data as well as correct output data to the machine learning model. The aim of a supervised learning algorithm is to **find a mapping function to map the input variable(x) with the output variable(y).**

In the real-world, supervised learning can be used for **Risk Assessment, Image classification, Fraud Detection, spam filtering**, etc.

How Supervised Learning Works?

In supervised learning, models are trained using labelled dataset, where the model learns about each type of data. Once the training process is completed, the model is tested on the basis of test data (a subset of the training set), and then it predicts the output.

The working of Supervised learning can be easily understood by the below example and diagram:



Programming with C

```
struct TCS
{
    int x: 1;
    int y: 2;
    int z: 4;
    int w: 8;
}A;

int main()
{
    printf("%d", sizeof(A));
    return 0;
}
```

Handwritten notes: *16 bits*, *15 bits*, *4 bytes*, *16*, *15*

What will be the output of above code in bytes? , if size of integer variable is consider to be as 4 bytes

Variable	Size (bytes)
A	4
B	16
C	8
D	15

Suppose we have a dataset of different types of shapes which includes square, rectangle, triangle, and Polygon. Now the first step is that we need to train the model for each shape.

- If the given shape has four sides, and all the sides are equal, then it will be labelled as a **Square**.
- If the given shape has three sides, then it will be labelled as a **triangle**.
- If the given shape has six equal sides then it will be labelled as **hexagon**.

Now, after training, we test our model using the test set, and the task of the model is to identify the shape.

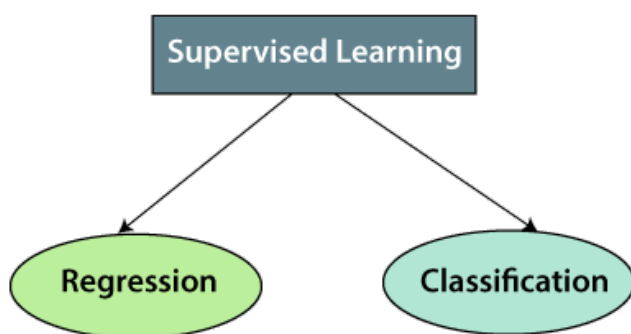
The machine is already trained on all types of shapes, and when it finds a new shape, it classifies the shape on the bases of a number of sides, and predicts the output.

Steps Involved in Supervised Learning:

- First Determine the type of training dataset
- Collect/Gather the labelled training data.
- Split the training dataset into training **dataset, test dataset, and validation dataset**.
- Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
- Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.
- Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.
- Evaluate the accuracy of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.

Types of supervised Machine learning Algorithms:

Supervised learning can be further divided into two types of problems:



1. Regression

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecast

popular Regression algorithms which come under supervised learning:

Programming with C

```
struct TCS
{
    int X; 1;
    int y; 2;
    int z; 4;
    int w; 8;
}A;

int main()
{
    printf("%d", sizeof(A));
    return 0;
}
```

Handwritten notes: 16 bits, 15 bits, 4 bytes, 32 bits

A	4	B	16
C	8	D	15

What will be the output of above code in bytes? , if size of integer variable is consider to be as 4 bytes

- Linear Regression
- Regression Trees
- Non-Linear Regression
- Bayesian Linear Regression
- Polynomial Regression

2. Classification

Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.

Spam Filtering,

- Random Forest
- Decision Trees
- Logistic Regression
- Support vector Machines

Note: We will discuss these algorithms in detail in later chapters.

Advantages of Supervised learning:

- With the help of supervised learning, the model can predict the output on the basis of prior experiences.
- In supervised learning, we can have an exact idea about the classes of objects.
- Supervised learning model helps us to solve various real-world problems such as **fraud detection, spam filtering**, etc.

Disadvantages of supervised learning:

- Supervised learning models are not suitable for handling the complex
- Supervised learning cannot predict the correct output if the test data
- Training required lots of computation times.
- In supervised learning, we need enough knowledge about the classes

Programming with C

```

struct TCS
{
    int x; 1;
    int y; 2;
    int z; 4;
    int w; 8;
}A;

int main()
{
    printf("%d", sizeof(A));
    return 0;
}

```

Handwritten notes: 16 bits, 15 bits, 4 bytes, 32 bits

A	4	B	16
C	8	D	15

What will be the output of above code in bytes? , if size of integer variable is consider to be as 4 bytes