**MACHINE LEARNING**

**Supervised learning:** Supervised learning is the learning of the model where with input variable ( say, x) and an output variable (say, Y) and an algorithm to map the input to the output.  
That is, **Y = f(X)**

**Why supervised learning?**  
The basic aim is to approximate the mapping function(mentioned above) so well that when there is a new input data (x) then the corresponding output variable can be predicted.

It is called supervised learning because the process of an learning(from the training dataset) can be thought of as a teacher who is supervising the entire learning process. Thus, the “learning algorithm” iteratively makes predictions on the training data and is corrected by the “teacher”, and the learning stops when the algorithm achieves an acceptable level of performance(or the desired accuracy).

**Example of Supervised Learning**  
Suppose there is a basket which is filled with some fresh fruits, the task is to arrange the same type of fruits at one place.  
Also, suppose that the fruits are apple, banana, cherry, grape.

Suppose one already knows from their *previous work* (or experience) that, the shape of each and every fruit present in the basket so, it is easy for them to arrange the same type of fruits in one place.

Here, the previous work is called as **training data** in Data Mining terminology. So, it learns the things from the training data. This is because it has a response variable which says y that if some fruit has so and so features then it is grape, and similarly for each and every fruit.

This type of information is deciphered from the data that is used to train the model.  
This type of learning is called **Supervised Learning**.  
Such problems are listed under classical *Classification Tasks*.  
   
**Unsupervised Learning:** Unsupervised learning is where only the input data (say, X) is present and no corresponding output variable is there.

**Why Unsupervised Learning?**  
The main aim of Unsupervised learning is to model the distribution in the data in order to learn more about the data.

It is called so, because there is no correct answer and there is no such teacher(unlike supervised learning). Algorithms are left to their own devises to discover and present the interesting structure in the data.

**Example of Unsupervised Learning**  
Again, Suppose there is a basket and it is filled with some fresh fruits. The task is to arrange the same type of fruits at one place.

This time there is no information about those fruits beforehand, its the first time that the fruits are being seen or discovered

So how to group similar fruits without any prior knowledge about those.  
First, any physical characteristic of a particular fruit is selected. Suppose *color*.

Then the fruits are arranged on the basis of the color. The groups will be something as shown below:  
**RED COLOR GROUP**: apples & cherry fruits.  
**GREEN COLOR GROUP**: bananas & grapes.

So now, take another physical character say, *size*, so now the groups will be something like this.  
**RED COLOR** AND **BIG SIZE**: apple.  
**RED COLOR** AND **SMALL SIZE**: cherry fruits.  
**GREEN COLOR** AND **BIG SIZE**: bananas.  
**GREEN COLOR** AND **SMALL SIZE**: grapes.  
The job is done!  
Here, there is no need to know or learn anything beforehand. That means, no train data and no response variable. This type of learning is known as **Unsupervised Learning**.

**Difference b/w Supervised and Unsupervised Learning :**

|  |  |  |
| --- | --- | --- |
|  | **SUPERVISED LEARNING** | **UNSUPERVISED LEARNING** |
| Input Data | Uses Known and Labeled Data as input | Uses Unknown Data as input |
| Computational Complexity | Very Complex | Less Computational Complexity |
| Real Time | Uses off-line analysis | Uses Real Time Analysis of Data |
| Number of Classes | Number of Classes are known | Number of Classes are not known |
| Accuracy of Results | Accurate and Reliable Results | Moderate Accurate and Reliable Results |