DAY 02 SUPERVISED LEARNING

DEFINITION :

- Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output.

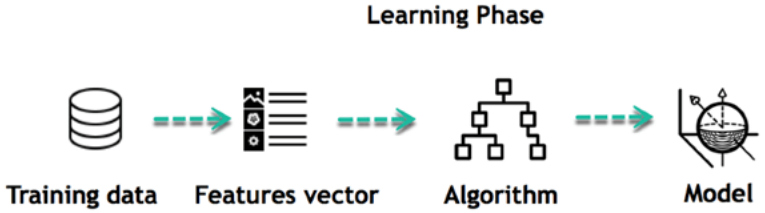
Y = f(X)

- The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.
- In other word Supervised learning, you train the machine using data which is well **"labeled"**. It means some data is already tagged with the correct answer. It can be compared to learning which takes place in the presence of a supervisor or a teacher.
- A supervised learning algorithm learns from labeled training data, helps you to predict outcomes for unforeseen data.

WHY IT IS CALLED SUPERVISED ?

- It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process.
- We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher. Learning stops when the algorithm achieves an acceptable level of performance.

WORK FLOW :



TYPES IN SUPERVISED LEARNING :

- Supervised learning problems can be grouped into regression and classification problems.
    1. **Classification:**
        - A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”.
    2. **Regression:**
        - A regression problem is when the output variable is a real value, such as “dollars” or “weight”.
- Some popular examples of supervised machine learning algorithms are:
    1. Linear regression (Simple/Multiple) for regression problems.
    2. Decision Tree and Random forest for classification and regression problems.
    3. Support vector machines, KNN, Naïve Bayes and Logistic Regression for classification problems.

ADVANTAGES :

- Supervised learning allows collecting data and produce data output from the previous experiences.
- Helps to optimize performance criteria with the help of experience.
- Supervised machine learning helps to solve various types of real-world computation problems.

CHALLENGES :

Here, are challenges faced in supervised machine learning :

- Irrelevant input feature present training data could give inaccurate results
- Data preparation and pre-processing is always a challenge.
- Accuracy suffers when impossible, unlikely, and incomplete values have been inputted as training data

- If the concerned expert is not available, then the other approach is "brute-force." It means you need to think that the right features (input variables) to train the machine on. It could be inaccurate.

DISADVANTAGES :

- Decision boundary might be overtrained if your training set which doesn't have examples that you want to have in a class
- You need to select lots of good examples from each class while you are training the classifier.
- Classifying big data can be a real challenge.
- Training for supervised learning needs a lot of computation time.

APPLICATION :

- **Regression Example**
  - **Example :** You can use regression to predict the house price from training data. The input variables will be locality, size of a house, etc.
  - **Strengths :** Outputs always have a probabilistic interpretation, and the algorithm can be regularized to avoid overfitting.
  - **Weaknesses :** Logistic regression may underperform when there are multiple or non-linear decision boundaries. This method is not flexible, so it does not capture more complex relationships.
- **Classification Example**
  - **Example :** Determining whether or not someone will be a defaulter of the loan.
  - **Strengths :** Classification tree perform very well in practice
  - **Weaknesses :** Unconstrained, individual trees are prone to overfitting.

SUMMARY :

- In Supervised learning, you train the machine using data which is well "labelled."
- You want to train a machine which helps you predict how long it will take you to drive home from your workplace is an example of supervised learning
- Regression and Classification are two types of supervised machine learning techniques.
- Supervised learning is a simpler method while Unsupervised learning is a complex method.
- The biggest challenge in supervised learning is that Irrelevant input feature present training data could give inaccurate results.
- The main advantage of supervised learning is that it allows you to collect data or produce a data output from the previous experience.
- The drawback of this model is that decision boundary might be overstrained if your training set doesn't have examples that you want to have in a class.
- As a best practice of supervise learning, you first need to decide what kind of data should be used as a training set.