Lecture:- Machine Learning Fundamentals

In this lecture, we will study about the fundamentals of machine learning, following is the topic we are going to study

- **What is Machine Learning?**
- **How it works?**
- **Types of Machine Learning Systems**

```
- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
```

- **Some Main Challenges in Machine Learning**
- **Dividing our data**
- **Overfitting**
- **Underfitting**
- **Cross Validation set**

What is Machine Learning?

Computer programs that uses algorithms to analyze data and make intelligent predictions based on the data without being explicitly programmed.

Slightly More Good definition:-

- **Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed. -> Arthur Samuel, 1959**
- **A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E. —> Tom Mitchell, 1997**

Machine Learning has taken place in almost every domain of industry, from bussiness to education or film industry or healthcare.

Some application of machine learning:-

- **Self Driving Cars**

- Robotics
- Real Estate
- In Healthcare
- In Bussiness

and lot lot more.

How it works?

Types of Machine Learning Systems

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Supervised Learning:- In this, we feed data to the algorithms, In which the data are labeled and we know what our output should like having the relationship between input values "X" and Output values "Y" .

Supervised Learning Problem can be divided in two problems:-

- Regression:- for continuous data
- Classification:- for categorical data

Examples of Supervised Learning:-

- Email Classification
- House Price Prediction
- Stock Price Prediction

Unsupervised Learning:- In this, we feed data to the algorithms which are not labeled or we can say that we don't know what our output should look like and there is not any kind of relationship between input var and output var. We have to recognize patterns based on the data, for doing so, we have different algorithms, we will study them later.

Examples of Unsupervised Learning:-

- T-Shirt Sizing

- T-Shirt Sizing
- Anomaly detection

Reinforcement Learning:-

Notation you should be familiar with

- **"Y" Often called the target variable which we have to predict.**
- **"X" is the input variable which we give input and Y we get**

Main Chanllenges in Machine Learning

1.) Insufficient Quantity of Training Data

2.) Poor-Quality Data

3.) Irrelevant Features

```
- Feature selection
- Feature Extraction
- Creating new features
```

Dividing our data

We divide our data into two sets, Training and Testing sets the reason is very simple, we want to test the model after we are done, so we make a test set for evaluating our model.

Overfitting the training data

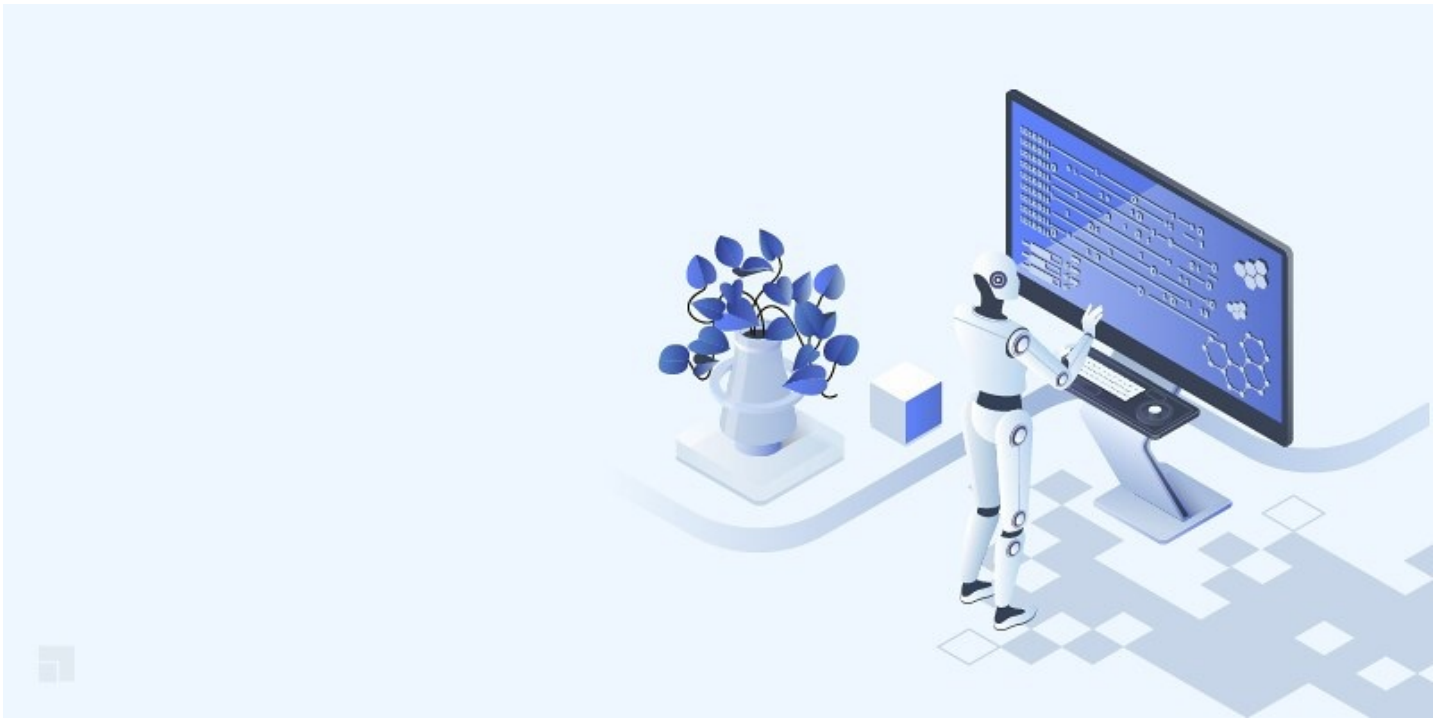
It means that the model performs well on the training data, but it does not generalize well.

Solution to this problem:-

- **To gather more data**
- **To reduce the noise in the training data**
- **To use regularization (we will study later on)**

Underfitting the Training Data

It means that the model performs bad on the trainig data,so it's obvious that it will also perform badly on testing data



Cross validation set

you simply hold out part of the training set to evaluate several candidate models and select the best one. The new heldout set is called the validation set (or sometimes the development set, or dev set). More specifically, you train multiple models with various hyperparameters on the reduced training set (i.e., the full training set minus the validation set), and you select the model that performs best on the validation set.