Nitte Meenakshi Institute of Technology

(AN AUTONOMOUS INSTITUTION AFFILIATED TO VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELGAUM)
PB No. 6429, Yelahanka, Bangalore 560-064, Karnataka
Telephone: 080- 22167800, 22167860, Fax: 080 - 22167805

Department of Computer Science & Engineering

Artificial Intelligence and Machine Learning (22CSG53)

Unit-3:

Introduction to Machine Learning, Supervised Learning Models

Semester – V Section: A

Faculty:
Mamatha Bai B G
Asst. Prof.,
Dept. of CSE,
NMIT

Content

Introduction to Machine Learning

- Overview of machine learning
- Types of machine learning
- Supervised
- Unsupervised
- Reinforcement learning
- Bias-variance tradeoff
- Model evaluation and performance metrics
- Overfitting and underfitting

Supervised Learning Models

- Linear regression
- logistic regression
- Bayesian Learning
- Bayes theorem and example
- Naïve Bayes Classifier and example.

Introduction to Machine Learning

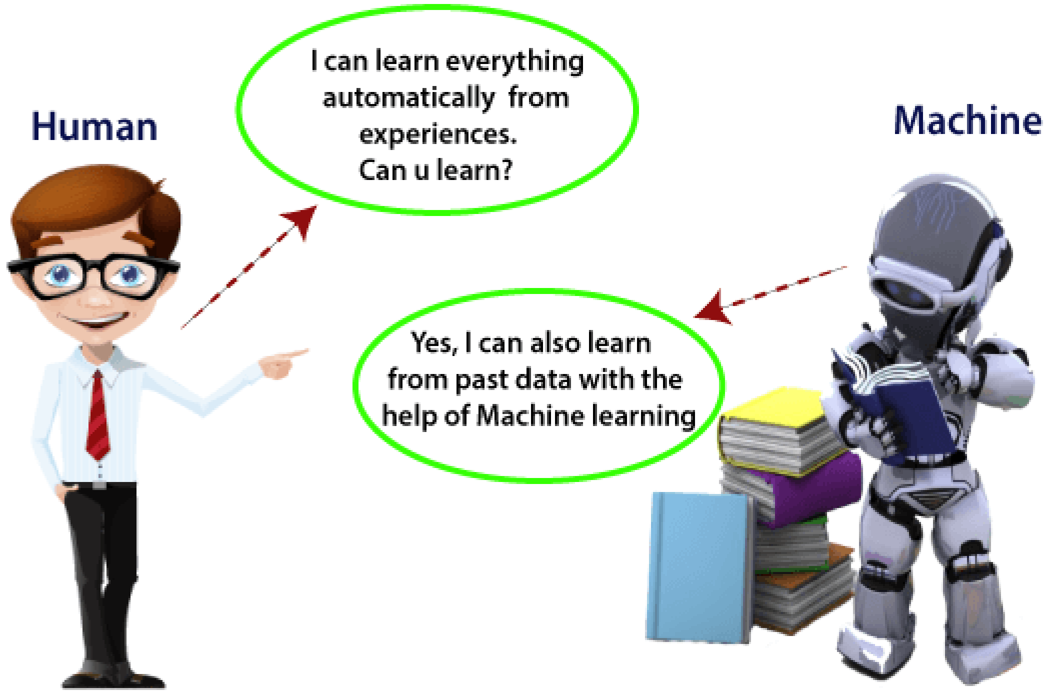
Overview of Machine Learning

- A rapidly developing field of technology, machine learning allows computers to automatically learn from previous data.
- For building mathematical models and making predictions based on historical data or information, machine learning employs a variety of algorithms.
- It is currently being used for a variety of tasks, including speech recognition, email filtering, auto-tagging on Facebook, a recommender system, and image recognition.

Introduction to Machine Learning (Contd.)

What is Machine Learning?

- In the real world, we are surrounded by humans who can learn everything from their experiences with their learning capability, and we have computers or machines which work on our instructions.
- But can a machine also learn from experiences or past data like a human does? So here comes the role of **Machine Learning**.



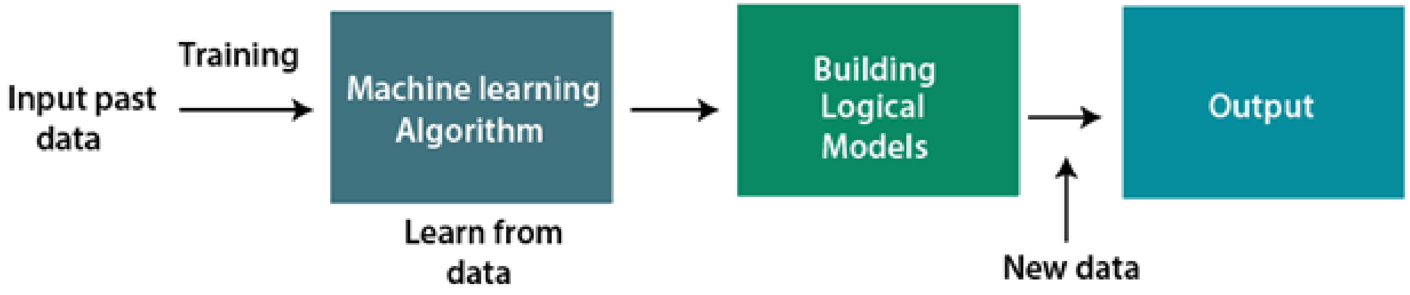
Introduction to Machine Learning (Contd.)

- A subset of artificial intelligence known as machine learning focuses primarily on the creation of algorithms that enable a computer to independently learn from data and previous experiences.
- Arthur Samuel first used the term "machine learning" in 1959. It could be summarized as follows:
- Without being explicitly programmed, machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things.
- Machine learning algorithms create a mathematical model that, without being explicitly programmed, aids in making predictions or decisions with the assistance of sample historical data, or training data.
- For the purpose of developing predictive models, machine learning brings together statistics and computer science.
- Algorithms that learn from historical data are either constructed or utilized in machine learning. The performance will rise in proportion to the quantity of information we provide.
- **A machine can learn if it can gain more data to improve its performance.**

Introduction to Machine Learning (Contd.)

How does Machine Learning work?

- A machine learning system builds prediction models, learns from previous data, and predicts the output of new data whenever it receives it.
- The amount of data helps to build a better model that accurately predicts the output, which in turn affects the accuracy of the predicted output.
- Let's say we have a complex problem in which we need to make predictions. Instead of writing code, we just need to feed the data to generic algorithms, which build the logic based on the data and predict the output. Our perspective on the issue has changed as a result of machine learning.
- The Machine Learning algorithm's operation is depicted in the following block diagram:



Introduction to Machine Learning (Contd.)

Features of Machine Learning:

- Machine learning uses data to detect various patterns in a given dataset.
- It can learn from past data and improve automatically.
- It is a data-driven technology.
- Machine learning is much similar to data mining as it also deals with the huge amount of the data.

Need for Machine Learning

- The demand for machine learning is steadily rising. Because it is able to perform tasks that are too complex for a person to directly implement, machine learning is required. Humans are constrained by our inability to manually access vast amounts of data; as a result, we require computer systems, which is where machine learning comes in to simplify our lives.
- By providing them with a large amount of data and allowing them to automatically explore the data, build models, and predict the required output, we can train machine learning algorithms. The cost function can be used to determine the amount of data and the machine learning algorithm's performance. We can save both time and money by using machine learning.
- The significance of AI can be handily perceived by its utilization's cases, Presently, AI is utilized in self-driving vehicles, digital misrepresentation identification, face acknowledgment, and companion idea by Facebook, and so on. Different top organizations, for example, Netflix and Amazon have constructed AI models that are utilizing an immense measure of information to examine the client interest and suggest item likewise.

Introduction to Machine Learning (Contd.)

Following are some key points which show the importance of Machine Learning:

- Rapid increment in the production of data
- Solving complex problems, which are difficult for a human
- Decision making in various sector including finance
- Finding hidden patterns and extracting useful information from data.

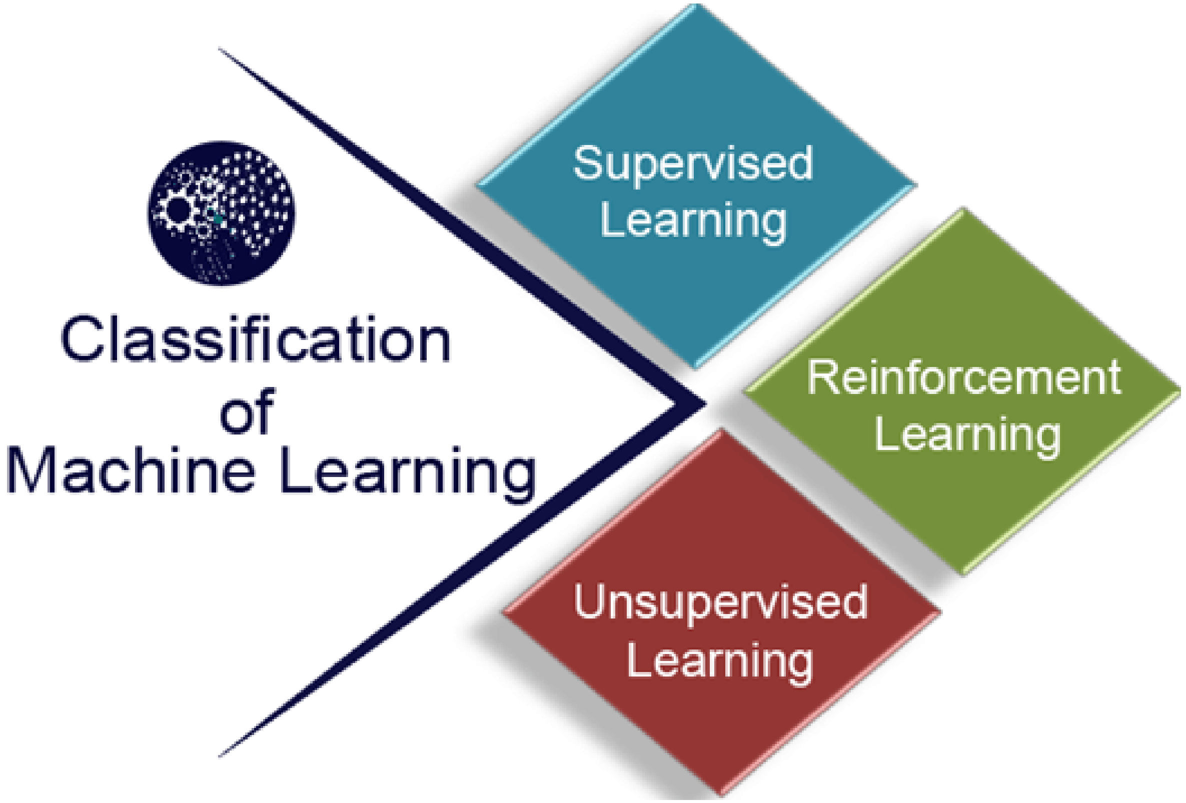
Machine Learning at present

- The field of machine learning has made significant strides in recent years, and its applications are numerous, including self-driving cars, Amazon Alexa, Catboats, and the recommender system.
- It incorporates clustering, classification, decision tree, SVM algorithms, and reinforcement learning, as well as unsupervised and supervised learning.
- Present day AI models can be utilized for making different expectations, including climate expectation, sickness forecast, financial exchange examination, and so on.

Types of Machine Learning

Classification of Machine Learning

- At a broad level, machine learning can be classified into three types:
  1. **Supervised learning**
  2. **Unsupervised learning**
  3. **Reinforcement learning**



Types of Machine Learning (Contd.)

1. Supervised Learning

- In supervised learning, sample labeled data are provided to the machine learning system for training, and the system then predicts the output based on the training data.
- The system uses labeled data to build a model that understands the datasets and learns about each one.
- After the training and processing are done, we test the model with sample data to see if it can accurately predict the output.
- The mapping of the input data to the output data is the objective of supervised learning.
- The managed learning depends on oversight, and it is equivalent to when an understudy learns things in the management of the educator.
- Spam filtering is an example of supervised learning.
- Supervised learning can be grouped further in two categories of algorithms:
  - **Classification**
  - **Regression**

Types of Machine Learning (Contd.)

2. Unsupervised Learning

- Unsupervised learning is a learning method in which a machine learns without any supervision.
- The training is provided to the machine with the set of data that has not been labeled, classified, or categorized, and the algorithm needs to act on that data without any supervision.
- The goal of unsupervised learning is to restructure the input data into new features or a group of objects with similar patterns.
- In unsupervised learning, we don't have a predetermined result. The machine tries to find useful insights from the huge amount of data.
- It can be further classifieds into two categories of algorithms:
  - **Clustering**
  - **Association**

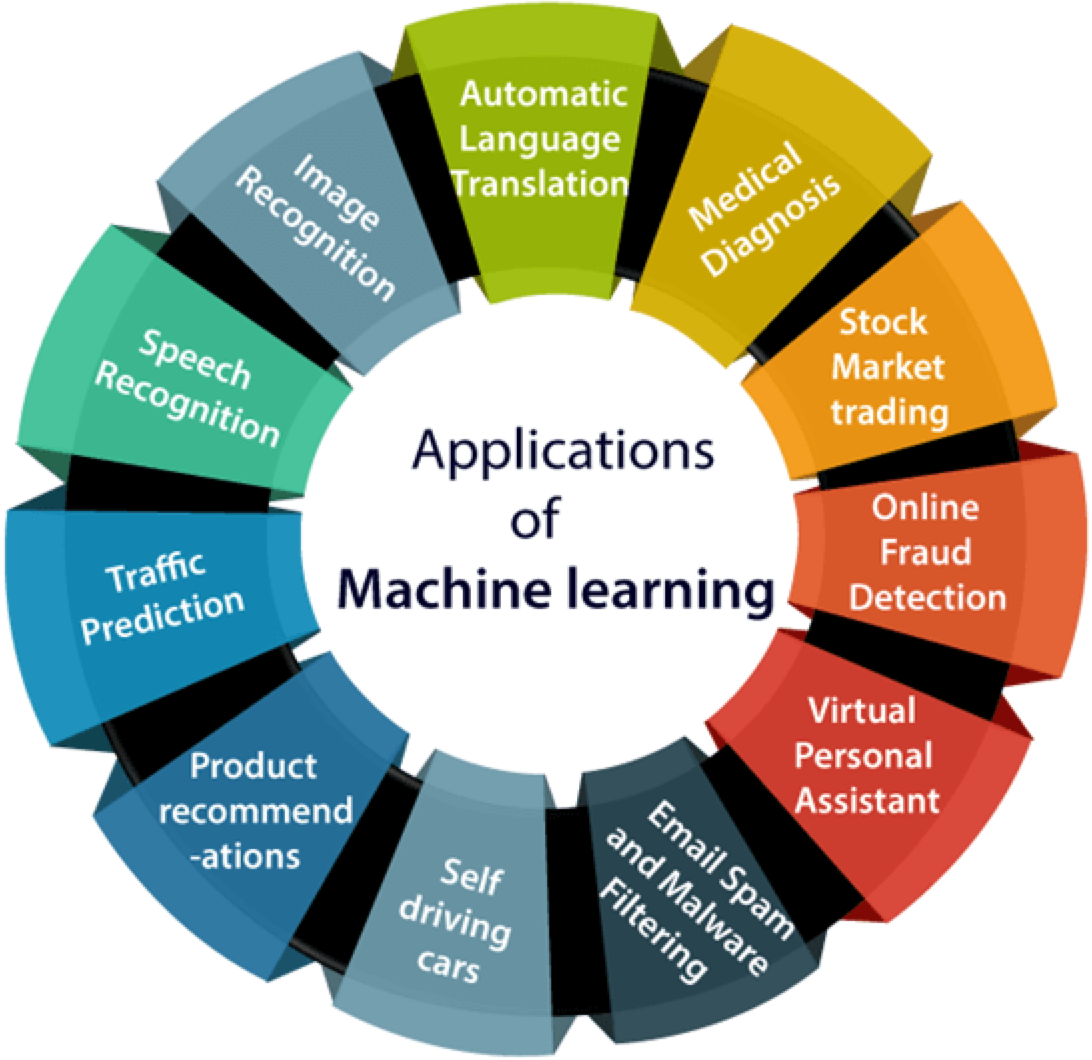
Types of Machine Learning (Contd.)

3. Reinforcement Learning

- Reinforcement learning is a feedback-based learning method, in which a learning agent gets a reward for each right action and gets a penalty for each wrong action.
- The agent learns automatically with these feedbacks and improves its performance.
- In reinforcement learning, the agent interacts with the environment and explores it.
- The goal of an agent is to get the most reward points, and hence, it improves its performance.
- The robotic dog, which automatically learns the movement of his arms, is an example of Reinforcement learning.

Applications of Machine Learning

- Machine learning is a buzzword for today's technology, and it is growing very rapidly day by day.
- We are using machine learning in our daily life even without knowing it such as Google Maps, Google assistant, Alexa, etc.
- Some of the most trending real-world applications of Machine Learning are as follows:



Applications of Machine Learning

1. Image Recognition

- Image recognition is one of the most common applications of machine learning. It is used to identify objects, persons, places, digital images, etc. The popular use case of image recognition and face detection is, **Automatic friend tagging suggestion**
- Facebook provides us a feature of auto friend tagging suggestion. Whenever we upload a photo with our Facebook friends, then we automatically get a tagging suggestion with name, and the technology behind this is machine learning's **face detection** and **recognition algorithm**.
- It is based on the Facebook project named "**Deep Face**," which is responsible for face recognition and person identification in the picture.

2. Speech Recognition

- While using Google, we get an option of "**Search by voice**," it comes under speech recognition, and it's a popular application of machine learning.
- Speech recognition is a process of converting voice instructions into text, and it is also known as "**Speech to text**", or "**Computer speech recognition**." At present, machine learning algorithms are widely used by various applications of speech recognition. **Google assistant**, **Siri**, **Cortana**, and **Alexa** are using speech recognition technology to follow the voice instructions.

Applications of Machine Learning

3. Traffic prediction

- If we want to visit a new place, we take help of Google Maps, which shows us the correct path with the shortest route and predicts the traffic conditions.
- It predicts the traffic conditions such as whether traffic is cleared, slow-moving, or heavily congested with the help of two ways:
- **Real Time location** of the vehicle form Google Map app and sensors
- **Average time has taken** on past days at the same time.
- Everyone who is using Google Map is helping this app to make it better. It takes information from the user and sends back to its database to improve the performance.

4. Product recommendations

- Machine learning is widely used by various e-commerce and entertainment companies such as **Amazon**, **Netflix**, etc., for product recommendation to the user. Whenever we search for some product on Amazon, then we started getting an advertisement for the same product while internet surfing on the same browser and this is because of machine learning.
- Google understands the user interest using various machine learning algorithms and suggests the product as per customer interest.
- As similar, when we use Netflix, we find some recommendations for entertainment series, movies, etc., and this is also done with the help of machine learning.

Applications of Machine Learning

5. Self-driving cars

- One of the most exciting applications of machine learning is self-driving cars. Machine learning plays a significant role in self-driving cars. Tesla, the most popular car manufacturing company is working on self-driving car. It is using unsupervised learning method to train the car models to detect people and objects while driving.

6. Email Spam and Malware Filtering

- Whenever we receive a new email, it is filtered automatically as important, normal, and spam. We always receive an important mail in our inbox with the important symbol and spam emails in our spam box, and the technology behind this is Machine learning. Below are some spam filters used by Gmail:
- Content Filter
- Header filter
- General blacklists filter
- Rules-based filters
- Permission filters
- Some machine learning algorithms such as **Multi-Layer Perceptron**, **Decision tree**, and **Naïve Bayes classifier** are used for email spam filtering and malware detection.

Applications of Machine Learning

7. Virtual Personal Assistant

- We have various virtual personal assistants such as **Google assistant**, **Alexa**, **Cortana**, **Siri**. As the name suggests, they help us in finding the information using our voice instruction. These assistants can help us in various ways just by our voice instructions such as Play music, call someone, Open an email, Scheduling an appointment, etc.
- These virtual assistants use machine learning algorithms as an important part.
- These assistant record our voice instructions, send it over the server on a cloud, and decode it using ML algorithms and act accordingly.

8. Online Fraud Detection

- Machine learning is making our online transaction safe and secure by detecting fraud transaction. Whenever we perform some online transaction, there may be various ways that a fraudulent transaction can take place such as **fake accounts**, **fake ids**, and **steal money** in the middle of a transaction. So to detect this, **Feed Forward Neural network** helps us by checking whether it is a genuine transaction or a fraud transaction.
- For each genuine transaction, the output is converted into some hash values, and these values become the input for the next round. For each genuine transaction, there is a specific pattern which gets change for the fraud transaction hence, it detects it and makes our online transactions more secure.

Applications of Machine Learning

9. Stock Market trading

- Machine learning is widely used in stock market trading. In the stock market, there is always a risk of up and downs in shares, so for this machine learning's **long short term memory neural network** is used for the prediction of stock market trends.

10. Medical Diagnosis

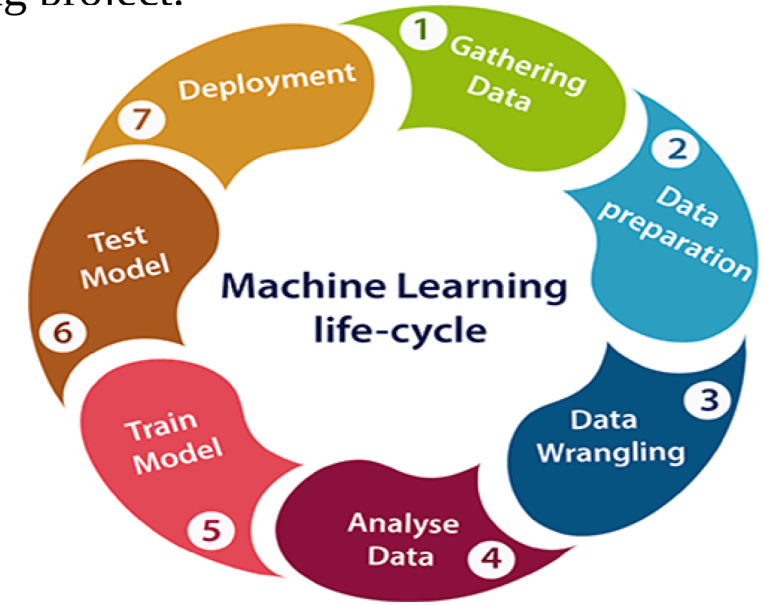
- In medical science, machine learning is used for diseases diagnoses. With this, medical technology is growing very fast and able to build 3D models that can predict the exact position of lesions in the brain.
- It helps in finding brain tumors and other brain-related diseases easily.

11. Automatic Language Translation

- Nowadays, if we visit a new place and we are not aware of the language then it is not a problem at all, as for this also machine learning helps us by converting the text into our known languages. Google's GNMT (Google Neural Machine Translation) provide this feature, which is a Neural Machine Learning that translates the text into our familiar language, and it called as automatic translation.
- The technology behind the automatic translation is a sequence to sequence learning algorithm, which is used with image recognition and translates the text from one language to another language.

Machine Learning Life Cycle

- Machine learning has given the computer systems the abilities to automatically learn without being explicitly programmed.
- But how does a machine learning system work? So, it can be described using the life cycle of machine learning.
- Machine learning life cycle is a cyclic process to build an efficient machine learning project.
- The main purpose of the life cycle is to find a solution to the problem or project.



Machine learning life cycle involves **seven major steps**, which are given below:

1. **Gathering Data**
2. **Data preparation**
3. **Data Wrangling**
4. **Analyse Data**
5. **Train the model**
6. **Test the model**
7. **Deployment**

- The most important thing in the complete process is to understand the problem and to know the purpose of the problem. Therefore, before starting the life cycle, we need to understand the problem because the good result depends on the better understanding of the problem.
- In the complete life cycle process, to solve a problem, we create a machine learning system called "model", and this model is created by providing "training". But to train a model, we need data, hence, life cycle starts by collecting data.

Machine Learning Life Cycle

1. Gathering Data

- Data Gathering is the first step of the machine learning life cycle. The goal of this step is to identify and obtain all data-related problems.
- In this step, we need to identify the different data sources, as data can be collected from various sources such as **files**, **database**, **internet**, or **mobile devices**. It is one of the most important steps of the life cycle. The quantity and quality of the collected data will determine the efficiency of the output. The more will be the data, the more accurate will be the prediction.
- This step includes the below tasks:
  - **Identify various data sources**
  - **Collect data**
  - **Integrate the data obtained from different sources**
- By performing the above task, we get a coherent set of data, also called as a **dataset**. It will be used in further steps.

Machine Learning Life Cycle

2. Data preparation

- After collecting the data, we need to prepare it for further steps. Data preparation is a step where we put our data into a suitable place and prepare it to use in our machine learning training.
- In this step, first, we put all data together, and then randomize the ordering of data.
- This step can be further divided into two processes:
- **Data exploration:** It is used to understand the nature of data that we have to work with. We need to understand the characteristics, format, and quality of data. A better understanding of data leads to an effective outcome. In this, we find Correlations, general trends, and outliers.
- **Data pre-processing:** Now the next step is preprocessing of data for its analysis.

Machine Learning Life Cycle

3. Data Wrangling

- Data wrangling is the process of cleaning and converting raw data into a useable format. It is the process of cleaning the data, selecting the variable to use, and transforming the data in a proper format to make it more suitable for analysis in the next step. It is one of the most important steps of the complete process. Cleaning of data is required to address the quality issues.
- It is not necessary that data we have collected is always of our use as some of the data may not be useful. In real-world applications, collected data may have various issues, including:
  - **Missing Values**
  - **Duplicate data**
  - **Invalid data**
  - **Noise**
- So, we use various filtering techniques to clean the data.
- It is mandatory to detect and remove the above issues because it can negatively affect the quality of the outcome.

Machine Learning Life Cycle

4. Data Analysis

- Now the cleaned and prepared data is passed on to the analysis step. This step involves:
  - **Selection of analytical techniques**
  - **Building models**
  - **Review the result**
- The aim of this step is to build a machine learning model to analyze the data using various analytical techniques and review the outcome. It starts with the determination of the type of the problems, where we select the machine learning techniques such as **Classification**, **Regression**, **Cluster analysis**, **Association**, etc. then build the model using prepared data, and evaluate the model.
- Hence, in this step, we take the data and use machine learning algorithms to build the model.

Machine Learning Life Cycle

5. Train Model

- Now the next step is to train the model, in this step we train our model to improve its performance for better outcome of the problem.
- We use datasets to train the model using various machine learning algorithms. Training a model is required so that it can understand the various patterns, rules, and, features.

6. Test Model

- Once our machine learning model has been trained on a given dataset, then we test the model. In this step, we check for the accuracy of our model by providing a test dataset to it.
- Testing the model determines the percentage accuracy of the model as per the requirement of project or problem.

7. Deployment

- The last step of machine learning life cycle is deployment, where we deploy the model in the real-world system.
- If the above-prepared model is producing an accurate result as per our requirement with acceptable speed, then we deploy the model in the real system. But before deploying the project, we will check whether it is improving its performance using available data or not. The deployment phase is similar to making the final report for a project.

Difference between Artificial Intelligence and Machine Learning

ARTIFICIAL INTELLIGENCE	MACHINE LEARNING
Artificial intelligence is a technology which enables a machine to simulate human behavior.	Machine learning is a subset of AI which allows a machine to automatically learn from past data without programming explicitly.
The goal of AI is to make a smart computer system like humans to solve complex problems.	The goal of ML is to allow machines to learn from data so that they can give accurate output.
In AI, we make intelligent systems to perform any task like a human.	In ML, we teach machines with data to perform a particular task and give an accurate result.
Machine learning and deep learning are the two main subsets of AI.	Deep learning is a main subset of machine learning.
AI has a very wide range of scope.	Machine learning has a limited scope.
AI is working to create an intelligent system which can perform various complex tasks.	Machine learning is working to create machines that can perform only those specific tasks for which they are trained.
AI system is concerned about maximizing the chances of success.	Machine learning is mainly concerned about accuracy and patterns.
The main applications of AI are Siri, customer support using catboats, Expert System, Online game playing, intelligent humanoid robot, etc.	The main applications of machine learning are Online recommender system, Google search algorithms, Facebook auto friend tagging suggestions, etc.
On the basis of capabilities, AI can be divided into three types, which are, Weak AI, General AI, and Strong AI.	Machine learning can also be divided into mainly three types that are Supervised learning, Unsupervised learning, and Reinforcement learning.
It includes learning, reasoning, and self-correction.	It includes learning and self-correction when introduced with new data.
AI completely deals with Structured, semi-structured, and unstructured data.	Machine learning deals with Structured and semi-structured data.

How to get datasets for Machine Learning?

- The field of ML depends vigorously on datasets for preparing models and making precise predictions.
- Datasets assume a vital part in the progress of AIML projects and are fundamental for turning into a gifted information researcher.

What is a dataset?

- **A dataset** is a collection of data in which data is arranged in some order.
- A dataset can contain any data from a series of an array to a database table.

Country	Age	Salary	Purchased
India	38	48000	No
France	43	45000	Yes
Germany	30	54000	No
France	48	65000	No
Germany	40		Yes
India	35	58000	Yes

Example of a Dataset

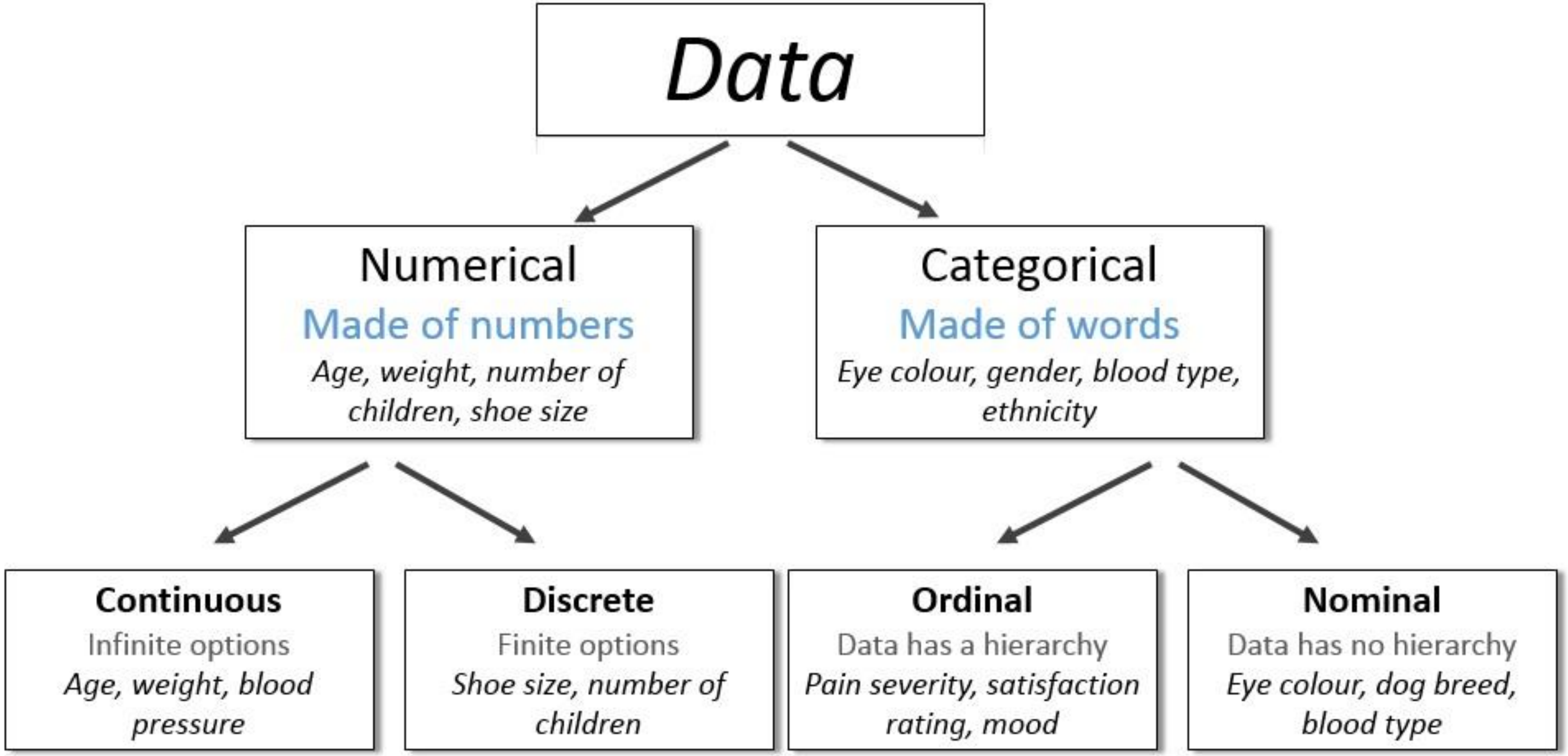
- A tabular dataset can be understood as a database table or matrix, where each column corresponds to a **particular variable,** and each row corresponds to the **fields of the dataset.** The most supported file type for a tabular dataset is **"Comma Separated File,"** or **CSV.** But to store a "tree-like data," we can use the JSON file more efficiently.

Types of data in datasets

- **Numerical data:** Such as house price, temperature, etc.
- **Categorical data:** Such as Yes/No, True/False, Blue/green, etc.
- **Ordinal data:** These data are similar to categorical data but can be measured on the basis of comparison.

Note: A real-world dataset is of huge size, which is difficult to manage and process at the initial level. Therefore, to practice machine learning algorithms, we can use any dummy dataset.

Types of Data



Types of Datasets

- Machine learning incorporates different domains, each requiring explicit sorts of datasets. A few normal sorts of datasets utilized in machine learning include:

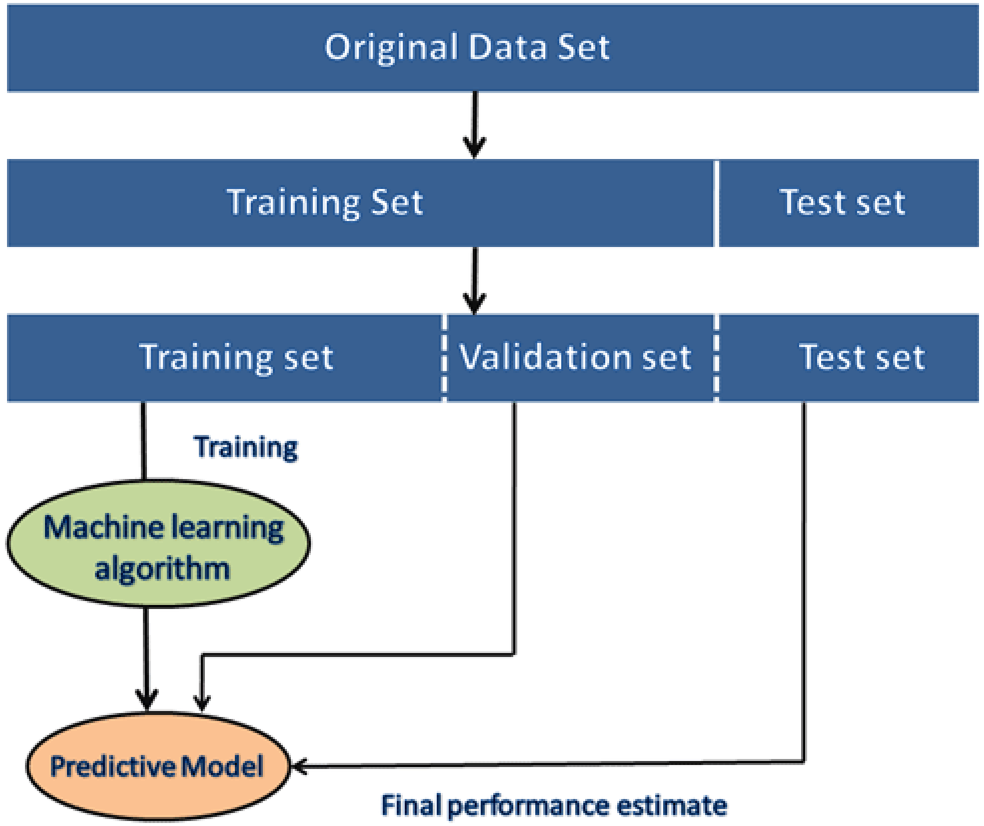
Image Datasets	Text Datasets	Time Series Datasets	Tabular Datasets
Image datasets contain an assortment of images and are normally utilized in computer vision tasks such as image classification, object detection, and image segmentation.	Text datasets comprise textual information, like articles, books, or virtual entertainment posts. These datasets are utilized in NLP techniques like sentiment analysis, text classification, and machine translation.	Time series datasets include information focuses gathered after some time. They are generally utilized in determining, abnormality location, and pattern examination.	Tabular datasets are organized information coordinated in tables or calculation sheets. They contain lines addressing examples or tests and segments addressing highlights or qualities. Tabular datasets are utilized for undertakings like relapse and arrangement.
Example: ImageNet CIFAR-10 MNIST	Example: Gutenberg Task dataset IMDb film reviews dataset	Example: Securities exchange information Climate information Sensor readings.	Example: Tabulated form

Need of Dataset

- Completely ready and pre-handled datasets are significant for machine learning projects.
- They give the establishment to prepare exact and solid models. Notwithstanding, working with enormous datasets can introduce difficulties regarding the board and handling.
- To address these difficulties, productive information executive strategies are expected to handle calculations.

Data Pre-processing

- Data pre-processing is a fundamental stage in preparing datasets for machine learning. It includes changing raw data into a configuration reasonable for model training.
- Normal pre-processing procedures incorporate data cleaning to eliminate irregularities or blunders, standardization to scale data inside a particular reach, highlight scaling to guarantee highlights have comparative ranges, and taking care of missing qualities through ascription or evacuation.
- During the development of the ML project, the developers completely rely on the datasets.
- In building ML applications, datasets are divided into two parts:
  - **Training dataset**
  - **Test Dataset**



Training Dataset and Test Dataset

- In machine learning, datasets are ordinarily partitioned into two sections: the training dataset and the test dataset.
- The training dataset is utilized to prepare the machine learning model, while the test dataset is utilized to assess the model's exhibition.
- This division surveys the model's capacity, to sum up to inconspicuous data.
- It is fundamental to guarantee that the datasets are representative of the issue space and appropriately split to stay away from inclination or overfitting.

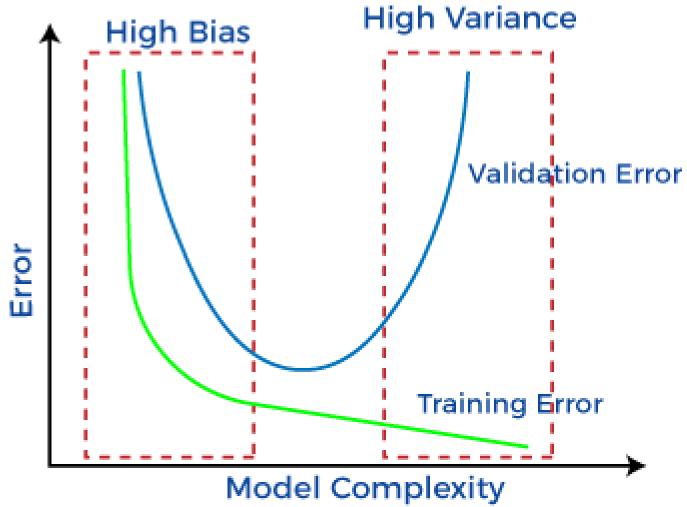
Popular sources for Machine Learning datasets

Below is the list of datasets which are freely available for the public to work on it:

1. Kaggle Datasets
2. UCI Machine Learning Repository
3. Datasets via AWS
4. Google's Dataset Search Engine
5. Microsoft Datasets
6. Awesome Public Dataset Collection
7. Government Datasets
8. Scikit-learn dataset
9. Computer Vision Datasets

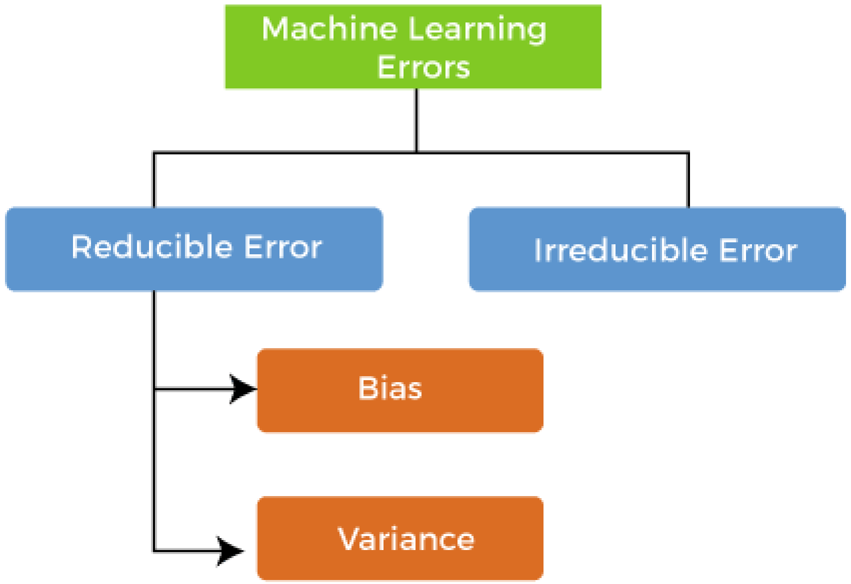
Bias and Variance in Machine Learning

- Machine learning is a branch of Artificial Intelligence, which allows machines to perform data analysis and make predictions.
- However, if the machine learning model is not accurate, it can make prediction errors, and these prediction errors are usually known as **Bias and Variance**.
- In machine learning, these errors will always be present as there is always a slight difference between the **model predictions and actual predictions**.
- The main aim of ML/data science analysts is to reduce these errors in order to get more accurate results.



Errors in Machine Learning?

- In machine learning, an error is a measure of how accurately an algorithm can make predictions for the previously unknown dataset. On the basis of these errors, the machine learning model is selected that can perform best on the particular dataset. There are mainly two types of errors in machine learning, which are:
  - **Reducible errors:** These errors can be reduced to improve the model accuracy. Such errors can further be classified into **Bias and Variance**.
  - **Irreducible errors:** These errors will always be present in the model regardless of which algorithm has been used. The cause of these errors is unknown variables whose value can't be reduced.



What is Bias?

- In general, a machine learning model analyses the data, find patterns in it and make predictions.
- While training, the model learns these patterns in the dataset and applies them to test data for prediction.
- ***While making predictions, a difference occurs between prediction values made by the model and actual values/expected values**, **and this difference is known as bias errors or Errors due to bias**.*
- It can be defined as an inability of machine learning algorithms such as Linear Regression to capture the true relationship between the data points. Each algorithm begins with some amount of bias because bias occurs from assumptions in the model, which makes the target function simple to learn.
- A model has either:
  - **Low Bias:** A low bias model will make fewer assumptions about the form of the target function.
  - **High Bias:** A model with a high bias makes more assumptions, and the model becomes unable to capture the important features of our dataset. **A high bias model also cannot perform well on new data.**
- Generally, a linear algorithm has a high bias, as it makes them learn fast. The simpler the algorithm, the higher the bias it has likely to be introduced. Whereas a nonlinear algorithm often has low bias.
- Some examples of machine learning algorithms with low bias **are Decision Trees, k-Nearest Neighbours and Support Vector Machines**.
- At the same time, an algorithm with high bias is **Linear Regression, Linear Discriminant Analysis and Logistic Regression.**

Ways to reduce High Bias:

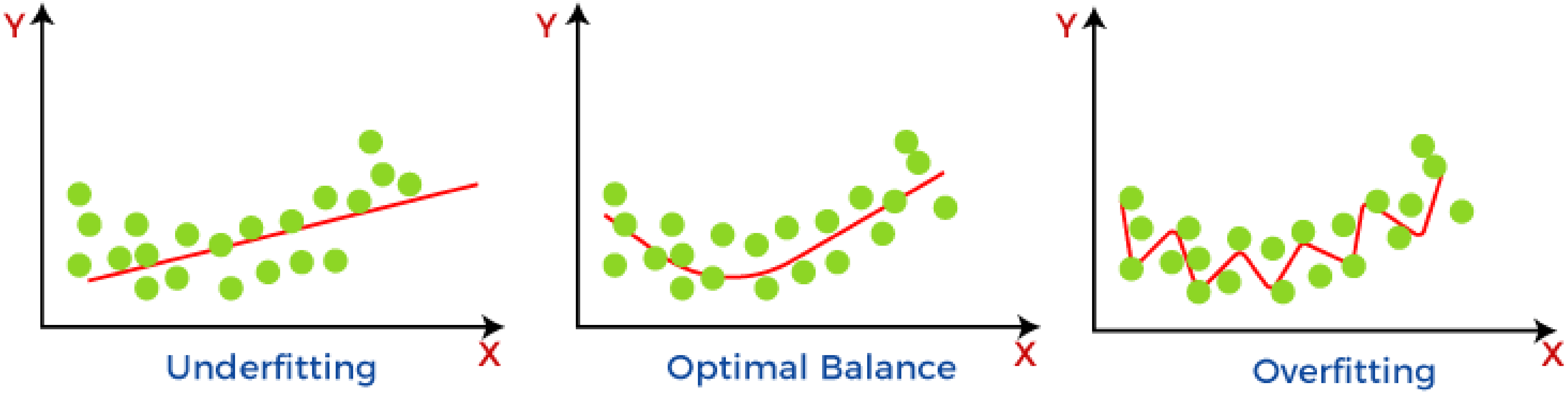
High bias mainly occurs due to a much simple model. Below are some ways to reduce the high bias:

- Increase the input features as the model is underfitted.
- Decrease the regularization term.
- Use more complex models, such as including some polynomial features.

What is a Variance Error?

- The variance would specify the amount of variation in the prediction if the different training data was used.
- In simple words, ***variance tells that how much a random variable is different from its expected value.***
- Ideally, a model should not vary too much from one training dataset to another, which means the algorithm should be good in understanding the hidden mapping between inputs and output variables.
- Variance errors are either of **low variance or high variance.**
- **Low variance** means there is a small variation in the prediction of the target function with changes in the training data set.
- At the same time, **High variance** shows a large variation in the prediction of the target function with changes in the training dataset.
- A model that shows high variance learns a lot and perform well with the training dataset, and does not generalize well with the unseen dataset. As a result, such a model gives good results with the training dataset but shows high error rates on the test dataset. Since, with high variance, the model learns too much from the dataset, it leads to **overfitting** of the model.
- A model with high variance has the below problems:
  - A high variance model leads to overfitting.
  - Increase model complexities.

- Usually, nonlinear algorithms have a lot of flexibility to fit the model, have high variance.



- Some examples of machine learning algorithms with low variance are, **Linear Regression, Logistic Regression, and Linear discriminant analysis**.
- At the same time, algorithms with high variance are **decision tree, Support Vector Machine, and K-nearest neighbours.**

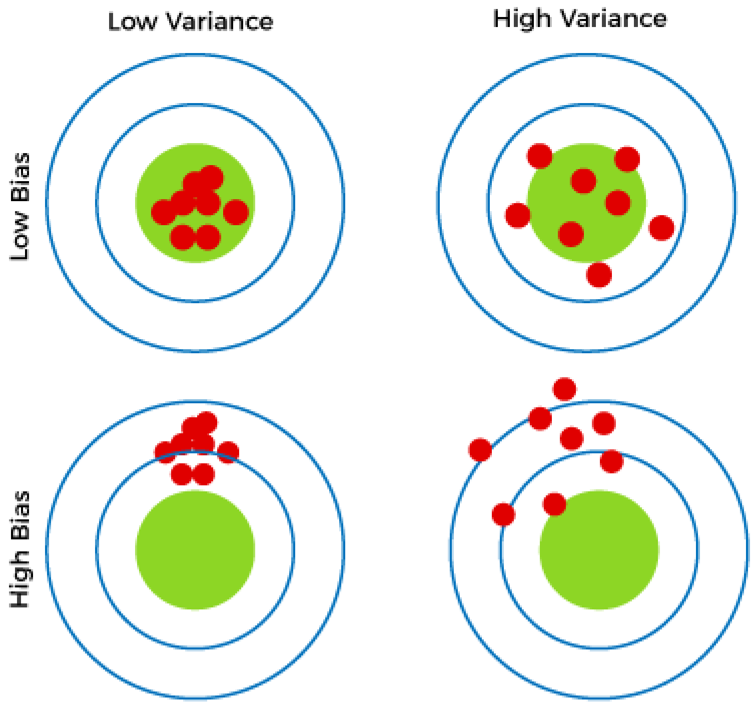
Ways to Reduce High Variance:

- Reduce the input features or number of parameters as a model is overfitted.
- Do not use a much complex model.
- Increase the training data.
- Increase the Regularization term.

Different Combinations of Bias-Variance

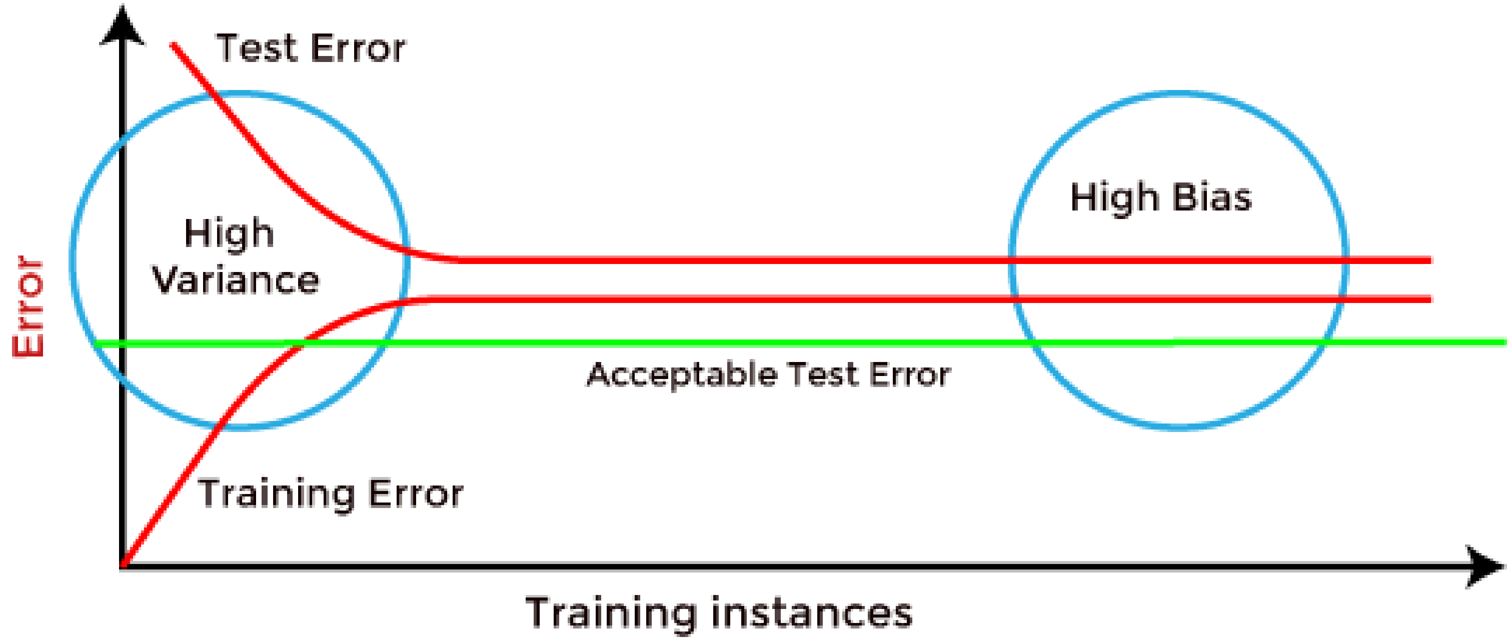
There are four possible combinations of bias and variances, which are represented by the below diagram:

1. **Low-Bias, Low-Variance:** The combination of low bias and low variance shows an ideal machine learning model. However, it is not possible practically.
2. **Low-Bias, High-Variance:** With low bias and high variance, model predictions are inconsistent and accurate on average. This case occurs when the model learns with a large number of parameters and hence leads to an **overfitting**
3. **High-Bias, Low-Variance:** With High bias and low variance, predictions are consistent but inaccurate on average. This case occurs when a model does not learn well with the training dataset or uses few numbers of the parameter. It leads to **underfitting** problems in the model.
4. **High-Bias, High-Variance:** With high bias and high variance, predictions are inconsistent and also inaccurate on average.



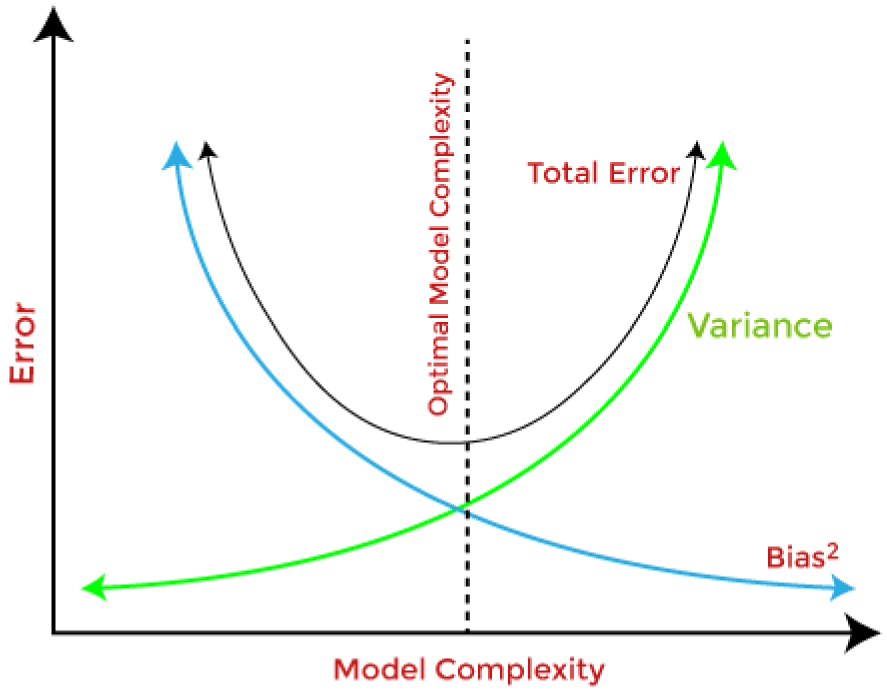
How to identify High variance or High Bias?

- High variance can be identified if the model has Low training error and high test error.
- High Bias can be identified if the model has High training error and the test error is almost similar to training error.



Bias-Variance Trade-Off

- While building the machine learning model, it is really important to take care of bias and variance in order to avoid overfitting and underfitting in the model.
- If the model is very simple with fewer parameters, it may have low variance and high bias.
- Whereas, if the model has a large number of parameters, it will have high variance and low bias.
- So, it is required to make a balance between bias and variance errors, and this balance between the bias error and variance error is known as **the** **Bias-Variance trade-off.**



Bias-Variance Trade-Off

- For an accurate prediction of the model, algorithms need a low variance and low bias. But this is not possible because bias and variance are related to each other:
  - If we decrease the variance, it will increase the bias.
  - If we decrease the bias, it will increase the variance.
- **Bias-Variance trade-off** is a central issue in supervised learning.
- Ideally, we need a model that accurately captures the regularities in training data and simultaneously generalizes well with the unseen dataset. Unfortunately, doing this is not possible simultaneously.
- Because a high variance algorithm may perform well with training data, but it may lead to overfitting to noisy data.
- Whereas, high bias algorithm generates a much simple model that may not even capture important regularities in the data. So, we need to find a sweet spot between bias and variance to make an optimal model.
- Hence, the ***Bias-Variance trade-off is about finding the sweet spot to make a balance between bias and variance errors.***

Model Evaluation & Performance Metrics in Machine Learning

- Evaluating the performance of a Machine learning model is one of the important steps while building an effective ML model.
- ***To evaluate the performance or quality of the model, different metrics are used, and these metrics are known as performance metrics or evaluation metrics.***
- These performance metrics help us understand how well our model has performed for the given data. In this way, we can improve the model's performance by tuning the hyper-parameters.
- Each ML model aims to generalize well on unseen/new data, and performance metrics help determine how well the model generalizes on the new dataset.
- In machine learning, each task or problem is divided into **classification** and **Regression**.
- Not all metrics can be used for all types of problems; hence, it is important to know and understand which metrics should be used.
- Different evaluation metrics are used for both Regression and Classification tasks.

Model Evaluation & Performance Metrics in Machine Learning

Performance Metrics for Classification

- In a classification problem, the category or classes of data is identified based on training data.
- The model learns from the given dataset and then classifies the new data into classes or groups based on the training.
- It predicts class labels as the output, such as *Yes or No*, *0 or 1*, *Spam or Not Spam*, etc.
- To evaluate the performance of a classification model, different metrics are used, and some of them are as follows:
  - **Accuracy**
  - **Confusion Matrix**
  - **Precision**
  - **Recall**
  - **F-Score**
  - **AUC(Area Under the Curve)-ROC**

I. Accuracy

- The accuracy metric is one of the simplest Classification metrics to implement, and it can be determined as the number of correct predictions to the total number of predictions.
- It can be formulated as:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ number\ of\ predictions}$$

- **When to Use Accuracy?**
- It is good to use the Accuracy metric when the target variable classes in data are approximately balanced. For example, if 60% of classes in a fruit image dataset are of Apple, 40% are Mango. In this case, if the model is asked to predict whether the image is of Apple or Mango, it will give a prediction with 97% of accuracy.
- **When not to use Accuracy?**
- It is recommended not to use the Accuracy measure when the target variable majorly belongs to one class. For example, Suppose there is a model for a disease prediction in which, out of 100 people, only five people have a disease, and 95 people don't have one. In this case, if our model predicts every person with no disease (which means a bad prediction), the Accuracy measure will be 95%, which is not correct.

II. Confusion Matrix

- A confusion matrix is a tabular representation of prediction outcomes of any binary classifier, which is used to describe the performance of the classification model on a set of test data when true values are known.
- The confusion matrix is simple to implement, but the terminologies used in this matrix might be confusing for beginners.
- A typical confusion matrix for a binary classifier looks like the below image

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

We can determine the following from the above matrix:

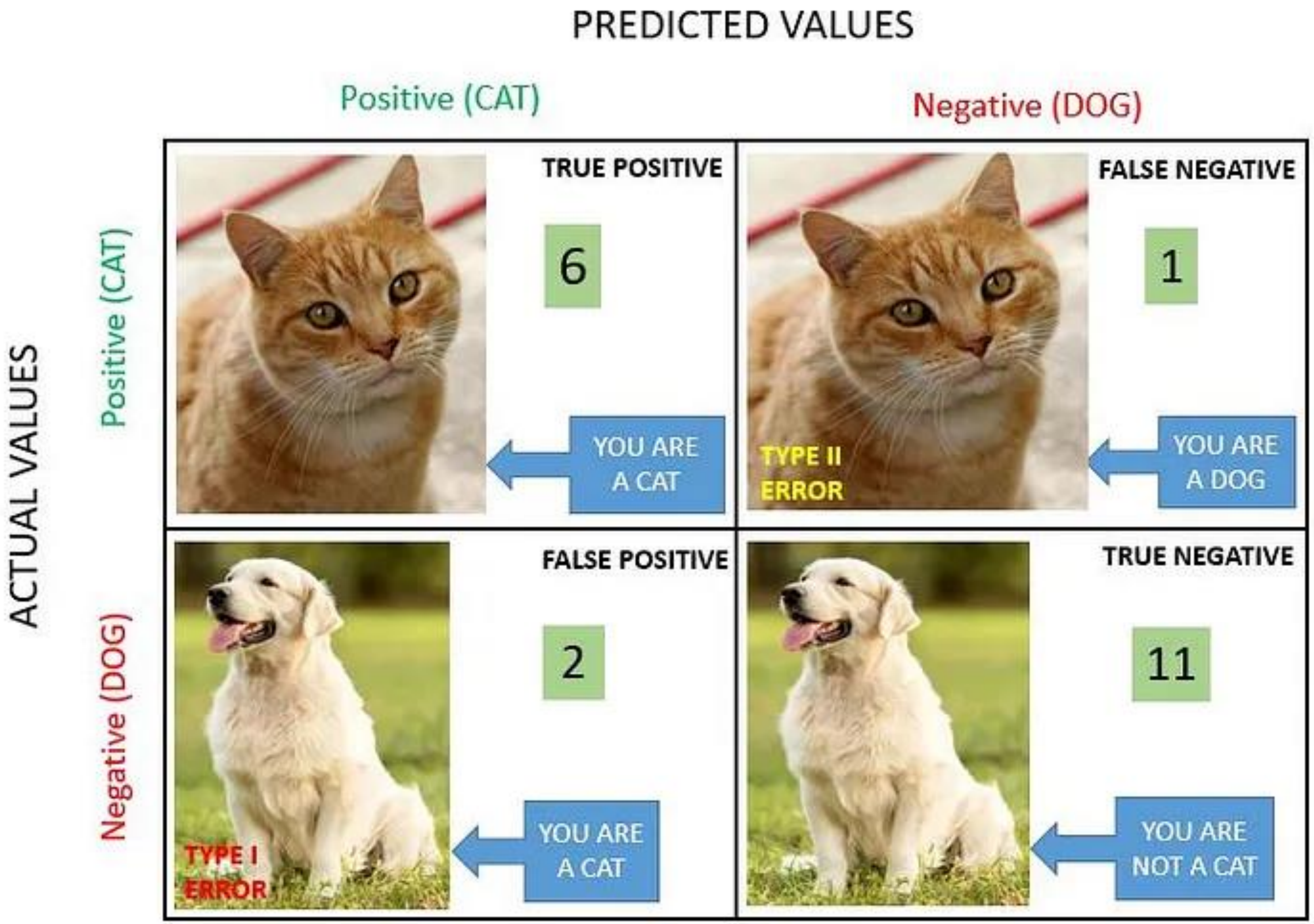
- In the matrix, columns are for the prediction values, and rows specify the Actual values. Here Actual and prediction give two possible classes, Yes or No. So, if we are predicting the presence of a disease in a patient, the Prediction column with Yes means, Patient has the disease, and for NO, the Patient doesn't have the disease.
- In this example, the total number of predictions are 165, out of which 110 time predicted yes, whereas 55 times predicted No.
- However, in reality, 60 cases in which patients don't have the disease, whereas 105 cases in which patients have the disease.

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

In general, the table is divided into four terminologies, which are as follows:

- **True Positive(TP):** In this case, the prediction outcome is true, and it is true in reality, also.
- **True Negative(TN):** in this case, the prediction outcome is false, and it is false in reality, also.
- **False Positive(FP):** In this case, prediction outcomes are true, but they are false in actuality.
- **False Negative(FN):** In this case, predictions are false, and they are true in actuality.

		Predicted	
		Negative (N) -	Positive (P) +
Actual	Negative -	True Negative (TN)	**False Positive (FP) Type I Error**
	Positive +	**False Negative (FN) Type II Error**	True Positive (TP)


PREDICTED VALUES
Positive (CAT)
Negative (DOG)
ACTUAL VALUES
Positive (CAT)
Negative (DOG)
TRUE POSITIVE
6
YOU ARE A CAT
FALSE NEGATIVE
1
TYPE II ERROR
YOU ARE A DOG
FALSE POSITIVE
2
TYPE I ERROR
YOU ARE A CAT
TRUE NEGATIVE
11
YOU ARE NOT A CAT

III. Precision

- The precision metric is used to overcome the limitation of Accuracy.
- The precision determines the proportion of positive prediction that was actually correct.
- It can be calculated as the True Positive or predictions that are actually true to the total positive predictions (True Positive and False Positive).

$$Precision = \frac{TP}{(TP + FP)}$$

IV. Recall or Sensitivity

- It is also similar to the Precision metric; however, it aims to calculate the proportion of actual positive that was identified incorrectly.
- It can be calculated as True Positive or predictions that are actually true to the total number of positives, either correctly predicted as positive or incorrectly predicted as negative (true Positive and false negative).
- The formula for calculating Recall is given below:

$$\text{Recall} = \frac{TP}{TP+FN}$$

When to use Precision and Recall?

- From the above definitions of Precision and Recall, we can say that recall determines the performance of a classifier with respect to a false negative, whereas precision gives information about the performance of a classifier with respect to a false positive.
- So, if we want to minimize the false negative, then, Recall should be as near to 100%, and if we want to minimize the false positive, then precision should be close to 100% as possible.
- In simple words, *if we maximize precision, it will minimize the FP errors, and if we maximize recall, it will minimize the FN error.*

V. F-Scores

- F-score or F1 Score is a metric to evaluate a binary classification model on the basis of predictions that are made for the positive class.
- It is calculated with the help of Precision and Recall.
- It is a type of single score that represents both Precision and Recall. So, ***the F1 Score can be calculated as the harmonic mean of both precision and Recall, assigning equal weight to each of them.***
- The formula for calculating the F1 score is given as:

$$F1 - score = 2 * \frac{precision * recall}{precision + recall}$$

- **When to use F-Score?**
- As F-score make use of both precision and recall, so it should be used if both of them are important for evaluation, but one (precision or recall) is slightly more important to consider than the other. For example, when False negatives are comparatively more important than false positives, or vice versa.

VI. AUC-ROC

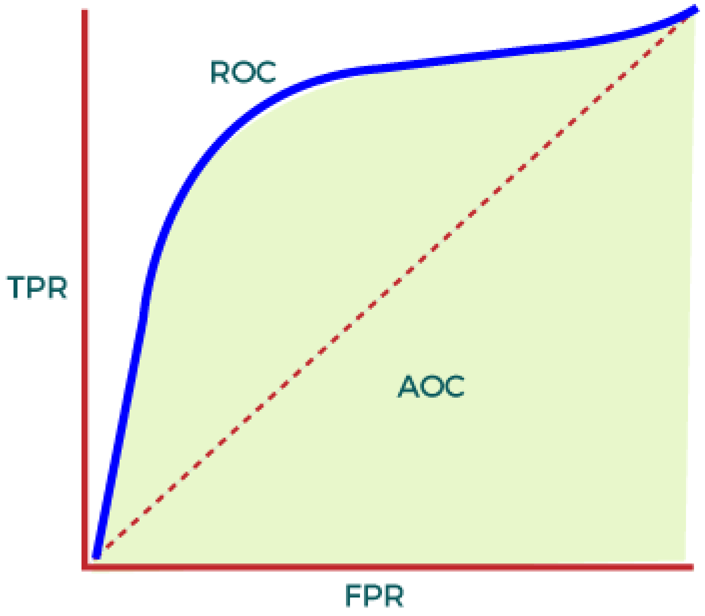
- Sometimes we need to visualize the performance of the classification model on charts; then, we can use the AUC-ROC curve. It is one of the popular and important metrics for evaluating the performance of the classification model.
- ***ROC*** (Receiver Operating Characteristic curve) curve ***represents a graph to show the performance of a classification model at different threshold levels***. The curve is plotted between two parameters, which are:
  - **True Positive Rate**
  - **False Positive Rate**

- TPR or true Positive rate is a synonym for Recall, hence can be calculated as: $TPR = \frac{TP}{TP+FN}$
- FPR or False Positive Rate can be calculated as: $FPR = \frac{FP}{FP+TN}$
- To calculate value at any point in a ROC curve, we can evaluate a logistic regression model multiple times with different classification thresholds, but this would not be much efficient.
- So, for this, one efficient method is used, which is known as AUC.

AUC: Area Under the ROC curve

- AUC is known for **Area Under the ROC curve**. As its name suggests, AUC calculates the two-dimensional area under the entire ROC curve, as shown:

- AUC calculates the performance across all the thresholds and provides an aggregate measure. The value of AUC ranges from 0 to 1. It means a model with 100% wrong prediction will have an AUC of 0.0, whereas models with 100% correct predictions will have an AUC of 1.0.



When to Use AUC

- AUC should be used to measure how well the predictions are ranked rather than their absolute values. Moreover, it measures the quality of predictions of the model without considering the classification threshold.

When not to use AUC

- As AUC is scale-invariant, which is not always desirable, and we need calibrating probability outputs, then AUC is not preferable.
- Further, AUC is not a useful metric when there are wide disparities in the cost of false negatives vs. false positives, and it is difficult to minimize one type of classification error.

Example-1

Index	1	2	3	4	5	6	7	8	9	10
Actual	Dog	Dog	Dog	Not Dog	Dog	Not Dog	Dog	Dog	Not Dog	Not Dog
Predicted	Dog	Not Dog	Dog	Not Dog	Dog	Dog	Dog	Dog	Not Dog	Not Dog
Result	TP	FN	TP	TN	TP	FP	TP	TP	TN	TN

Actual Dog Counts = 6
Actual Not Dog Counts = 4
True Positive Counts = 5
False Positive Counts = 1
True Negative Counts = 3
False Negative Counts = 1

		Actual	
		Dog	**Not Dog**
Predicted	**Dog**	True Positive (TP =5)	False Positive (FP=1)
	Not Dog	False Negative (FN =1)	True Negative (TN=3)

Confusion Matrix

Example-1

- For the previous example:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Accuracy = (5+3)/(5+3+1+1) = 8/10 = 0.8

$$\text{Precision} = \frac{TP}{TP+FP}$$

Precision = 5/(5+1) =5/6 = 0.8333

$$\text{Recall} = \frac{TP}{TP+FN}$$

Recall = 5/(5+1) =5/6 = 0.8333

$$\text{F1-Score} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

F1-Score: = (2* 0.8333* 0.8333)/(0.8333+ 0.8333) = 0.8333

Example-2

	Predicted without the variant	Predicted with the variant
Actual number without the variant = 195	True negative = 45	False positive = 150
Actual number with the variant = 305	False negative = 105	True positive = 200
	150	350

Recall rate = (True positive value) / (Actual positive value) = (200) / (305) = 0.66 = 66%

Specificity rate = (True negative value) / (Actual negative value) = (45) / (195) = 0.23 = 23%

Accuracy rate = (True positive value + True negative value) / (Total number of samples) = (200 + 45) / (500) = (245) / (500) = 0.49 = 49%

Misclassification (error) rate = (False positive value + False negative value) / (Total number of samples) = (150 + 105) / (500) = (255) / (500) = 0.51 = 51%

Example-3

- Suppose we are trying to create a model that can predict the result for the disease that is either a person has that disease or not. So, the confusion matrix for this is given as:

n = 100	Actual: No	Actual: Yes	
Predicted: No	TN: 65	FP: 3	68
Predicted: Yes	FN: 8	TP: 24	32
	73	27	

Solution:

From the above example, we can conclude that:

The table is given for the two-class classifier, which has two predictions "Yes" and "NO." Here, Yes defines that patient has the disease, and No defines that patient does not has that disease.

The classifier has made a total of **100 predictions**. Out of 100 predictions, **89 are true predictions**, and **11 are incorrect predictions**.

The model has given prediction "yes" for 32 times, and "No" for 68 times. Whereas the actual "Yes" was 27, and actual "No" was 73 times.

Using these values of Confusion matrix, find Accuracy, Precision, Recall, F-score, Misclassification rate etc.

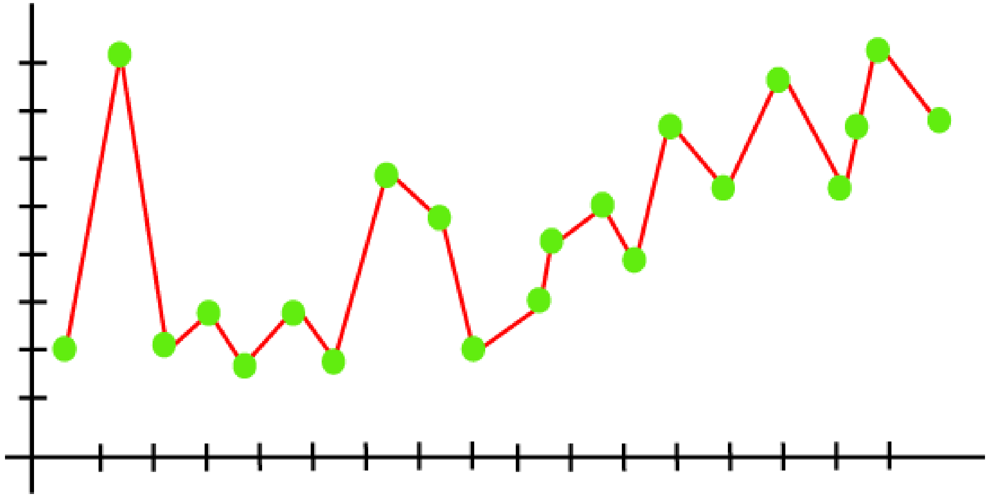
Overfitting and Underfitting in Machine Learning

- Overfitting and Underfitting are the two main problems that occur in machine learning and degrade the performance of the machine learning models.
- The main goal of each machine learning model is **to generalize well**. Here **generalization** defines the ability of an ML model to provide a suitable output by adapting the given set of unknown input.
- It means after providing training on the dataset, it can produce reliable and accurate output. Hence, the underfitting and overfitting are the two terms that need to be checked for the performance of the model and whether the model is generalizing well or not.
- Before understanding the overfitting and underfitting, let's understand some basic terms:
  - **Signal:** It refers to the true underlying pattern of the data that helps the machine learning model to learn from the data.
  - **Noise:** Noise is unnecessary and irrelevant data that reduces the performance of the model.
  - **Bias:** Bias is a prediction error that is introduced in the model due to oversimplifying the machine learning algorithms. Or it is the difference between the predicted values and the actual values.
  - **Variance:** If the machine learning model performs well with the training dataset, but does not perform well with the test dataset, then variance occurs.

Overfitting

- Overfitting occurs when our machine learning model tries to cover all the data points or more than the required data points present in the given dataset. Because of this, the model starts caching noise and inaccurate values present in the dataset, and all these factors reduce the efficiency and accuracy of the model. The overfitted model has **low bias** and **high variance.**
- The chances of occurrence of overfitting increase as much as we provide training to our model. It means the more we train our model, the more chances of occurring the overfitted model.
- Overfitting is the main problem that occurs in supervised learning.

Example: The concept of the overfitting can be understood by the below graph of the linear regression output:



- As we can see from the above graph, the model tries to cover all the data points present in the scatter plot. It may look efficient, but in reality, it is not so. Because the goal of the regression model to find the best fit line, but here we have not got any best fit, so, it will generate the prediction errors.

How to avoid the Overfitting in Model?

- Both overfitting and underfitting cause the degraded performance of the machine learning model. But the main cause is overfitting, so there are some ways by which we can reduce the occurrence of overfitting in our model.
  - **Cross-Validation**
  - **Training with more data**
  - **Removing features**
  - **Early stopping the training**
  - **Regularization**
  - **Ensembling**

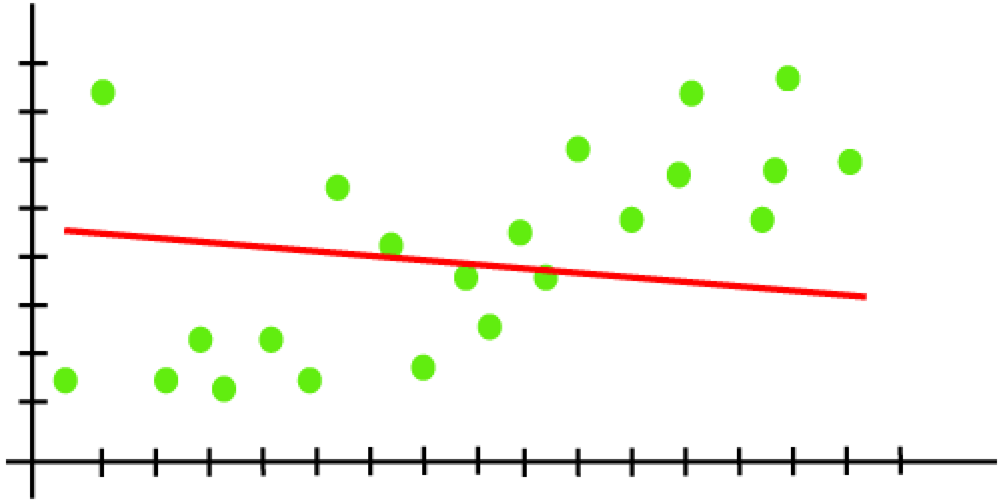
Underfitting

- Underfitting occurs when our machine learning model is not able to capture the underlying trend of the data.
- To avoid the overfitting in the model, the fed of training data can be stopped at an early stage, due to which the model may not learn enough from the training data. As a result, it may fail to find the best fit of the dominant trend in the data.
- In the case of underfitting, the model is not able to learn enough from the training data, and hence it reduces the accuracy and produces unreliable predictions.
- An underfitted model has high bias and low variance.

- **Example:** We can understand the underfitting using below output of the linear regression model:
- As we can see from the above diagram, the model is unable to capture the data points present in the plot.

How to avoid underfitting?

- By increasing the training time of the model.
- By increasing the number of features.



Goodness of Fit

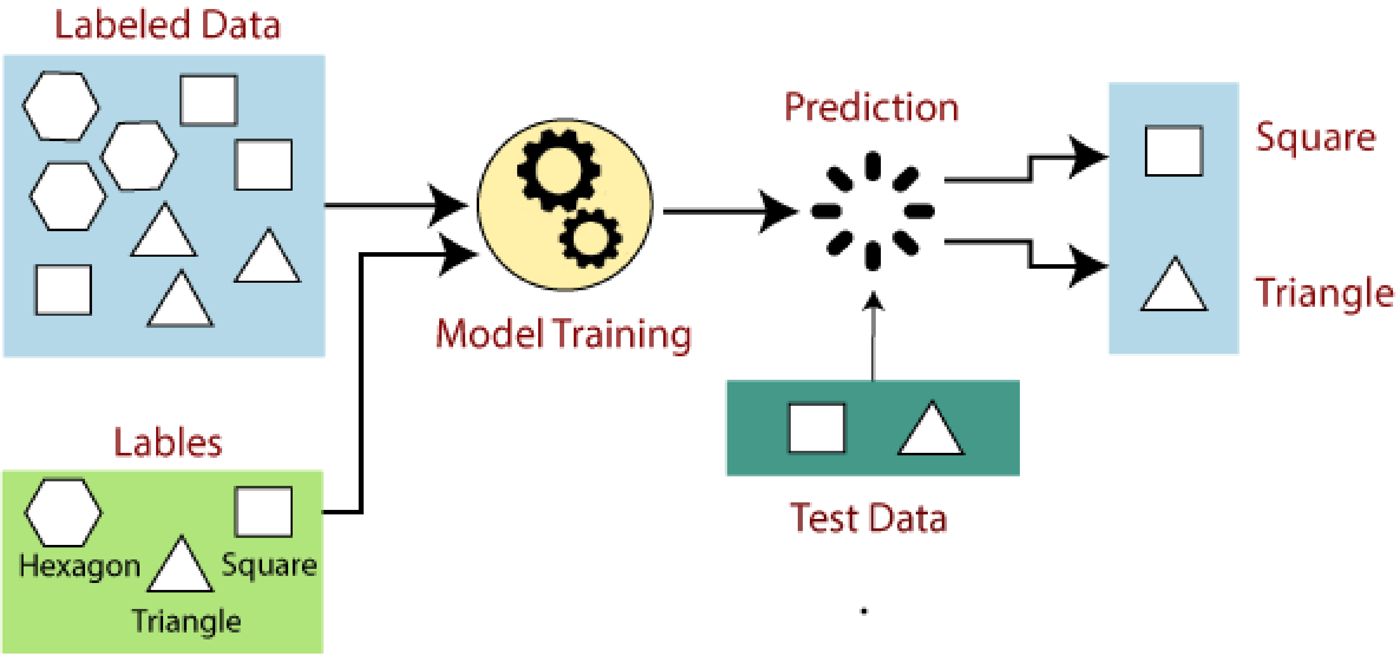
- The "Goodness of fit" term is taken from the statistics, and the goal of the machine learning models to achieve the goodness of fit.
- In statistics modeling, *it defines how closely the result or predicted values match the true values of the dataset.*
- The model with a good fit is between the underfitted and overfitted model, and ideally, it makes predictions with 0 errors, but in practice, it is difficult to achieve it.
- As and when we train our model for a time, the errors in the training data go down, and the same happens with test data.
- But if we train the model for a long duration, then the performance of the model may decrease due to the overfitting, as the model also learn the noise present in the dataset.
- The errors in the test dataset start increasing, *so the point, just before the raising of errors, is the good point, and we can stop here for achieving a good model.*
- There are two other methods by which we can get a good point for our model, which are the **resampling method** to estimate model accuracy and **validation dataset**.

Supervised Machine Learning

- Supervised learning is the type 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:



- 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:

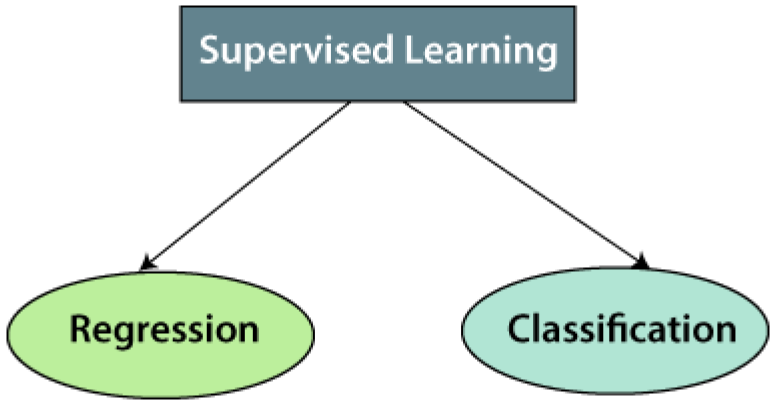
- First Determine the type of training dataset
- Collect/Gather the labelled training data.
- Split the training dataset into **training dataset, test dataset, and validation dataset**.
- Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately 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 forecasting, Market Trends, etc. Below are some popular Regression algorithms which come under supervised learning:
  - 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

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 tasks.
- Supervised learning cannot predict the correct output if the test data is different from the training dataset.
- Training required lots of computation times.
- In supervised learning, we need enough knowledge about the classes of object.

Linear Regression in Machine Learning

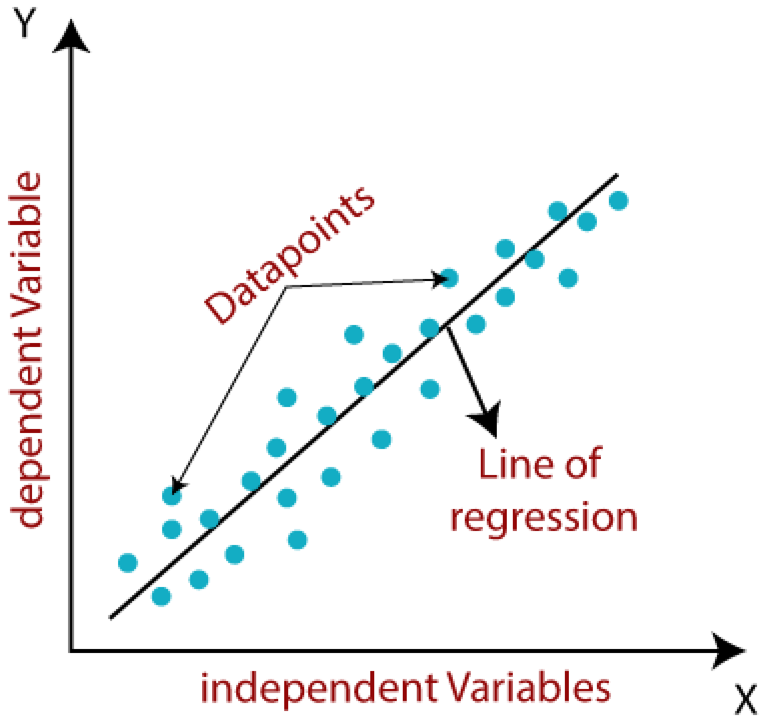
- Linear regression is one of the easiest and most popular Machine Learning algorithms.
- It is a statistical method that is used for **predictive analysis**.
- Linear regression makes predictions for **continuous/real or numeric variables** such as **sales, salary, age, product price,** etc.
- Linear regression algorithm shows a between a dependent (y) and one or more independent (x) variables, **linear relationship** hence called as linear regression.
- Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.
- The linear regression model provides a sloped straight line representing the relationship between the variables.

- Consider the below image:
- Mathematically, we can represent a linear regression as:

$$y= a_0+a_1x+ \varepsilon$$

where,

Y = Dependent Variable (Target Variable)
X= Independent Variable (predictor Variable)
a0= intercept of the line (Gives an additional degree of freedom)
a1 = Linear regression coefficient (scale factor to each input value)
ε = random error



- The values for x and y variables are training datasets for Linear Regression model representation.

Types of Linear Regression

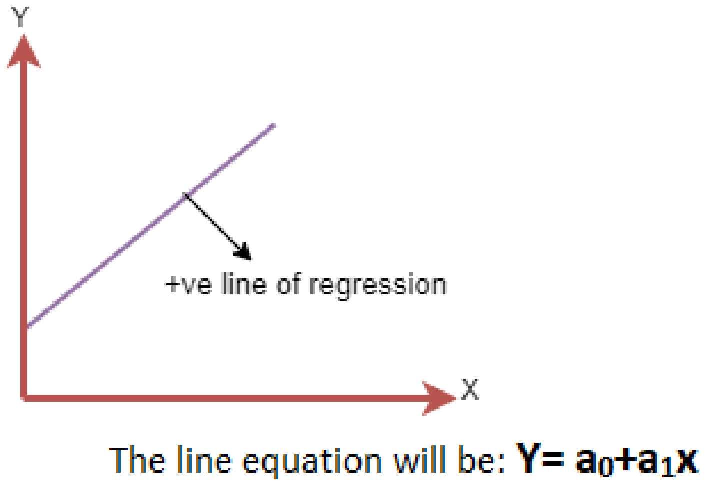
Linear regression can be further divided into two types of the algorithm:

- **Simple Linear Regression:**
  If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.
- **Multiple Linear regression:**
  If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression.

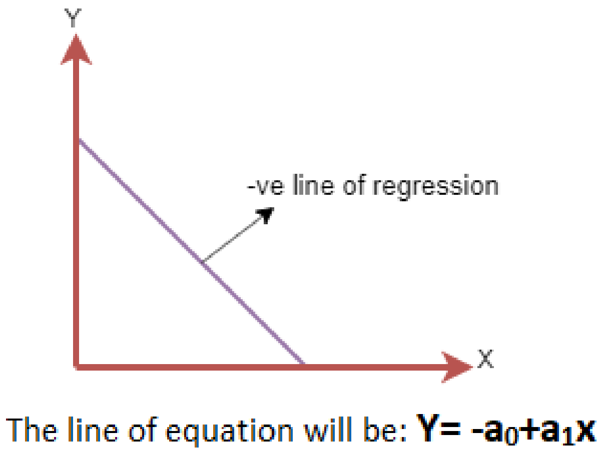
Linear Regression Line

- A linear line showing the relationship between the dependent and independent variables is called a **regression line**. A regression line can show two types of relationship:

Positive Linear Relationship:
If the dependent variable increases on the Y-axis and independent variable increases on X-axis, then such a relationship is termed as a Positive linear relationship.



Negative Linear Relationship:
If the dependent variable decreases on the Y-axis and independent variable increases on the X-axis, then such a relationship is called a negative linear relationship.



Finding the best fit line:

- When working with linear regression, our main goal is to find the best fit line that means the error between predicted values and actual values should be **minimized**. The best fit line will have the least error.
- The different values for weights or the coefficient of lines (a_0, a_1) gives a different line of regression, so we need to calculate the best values for a_0 and a_1 to find the best fit line, so to calculate this we use cost function.

Cost function

- The different values for weights or coefficient of lines (a_0, a_1) gives the different line of regression, and the cost function is used to estimate the values of the coefficient for the best fit line.
- Cost function optimizes the regression coefficients or weights. It measures how a linear regression model is performing.
- We can use the cost function to find the accuracy of the **mapping function**, which maps the input variable to the output variable. This mapping function is also known as **Hypothesis function**.
- For Linear Regression, we use the **Mean Squared Error (MSE)** cost function, which is the average of squared error occurred between the predicted values and actual values.

- For the above linear equation, MSE can be calculated as:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Where,
MSE is Mean Squared Error
N = Number of Data points
Y_i = Actual Values

$\hat{Y}_i$ = Predicted Values

Residuals: The distance between the actual value and predicted values is called residual. If the observed points are far from the regression line, then the residual will be high, and so cost function will high. If the scatter points are close to the regression line, then the residual will be small and hence the cost function.

Gradient Descent:

- Gradient descent is used to minimize the MSE by calculating the gradient of the cost function.
- A regression model uses gradient descent to update the coefficients of the line by reducing the cost function.
- It is done by a random selection of values of coefficient and then iteratively update the values to reach the minimum cost function.

Classification Algorithm in Machine Learning

- As we know, the Supervised Machine Learning algorithm can be broadly classified into Regression and Classification Algorithms. In Regression algorithms, we have predicted the output for continuous values, but to predict the categorical values, we need Classification algorithms.

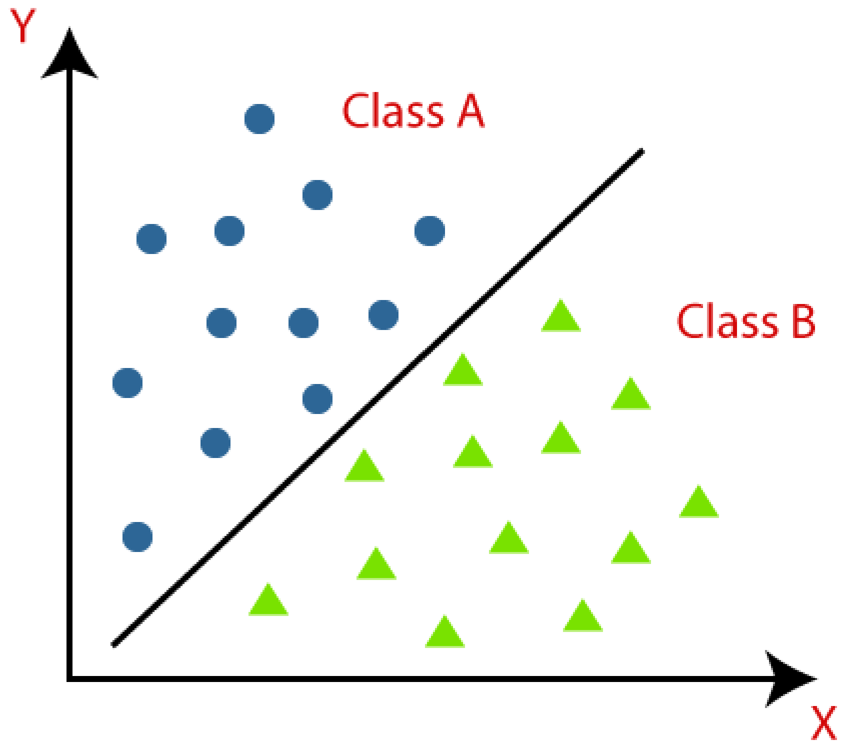
What is the Classification Algorithm?

- The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training data.
- In Classification, a program learns from the given dataset or observations and then classifies new observation into a number of classes or groups. Such as, **Yes or No, 0 or 1, Spam or Not Spam, cat or dog,** etc. Classes can be called as targets/labels or categories.
- Unlike regression, the output variable of Classification is a category, not a value, such as "Green or Blue", "fruit or animal", etc. Since the Classification algorithm is a Supervised learning technique, hence it takes labeled input data, which means it contains input with the corresponding output.
- In classification algorithm, a discrete output function(y) is mapped to input variable(x).

y = f(x), where y = categorical output

- The best example of an ML classification algorithm is **Email Spam Detector**.
- The main goal of the Classification algorithm is to identify the category of a given dataset, and these algorithms are mainly used to predict the output for the categorical data.

- Classification algorithms can be better understood using the below diagram. In the below diagram, there are two classes, class A and Class B. These classes have features that are similar to each other and dissimilar to other classes.
- The algorithm which implements the classification on a dataset is known as a **classifier**.
- There are two types of Classifications:



1. **Binary Classifier:** If the classification problem has only two possible outcomes, then it is called as Binary Classifier.
   Examples: YES or NO, MALE or FEMALE, SPAM or NOT SPAM, CAT or DOG, etc.
2. **Multi-class Classifier:** If a classification problem has more than two outcomes, then it is called as Multi-class Classifier.
   Example: Classifications of types of crops, Classification of types of music.

Learners in Classification Problems:

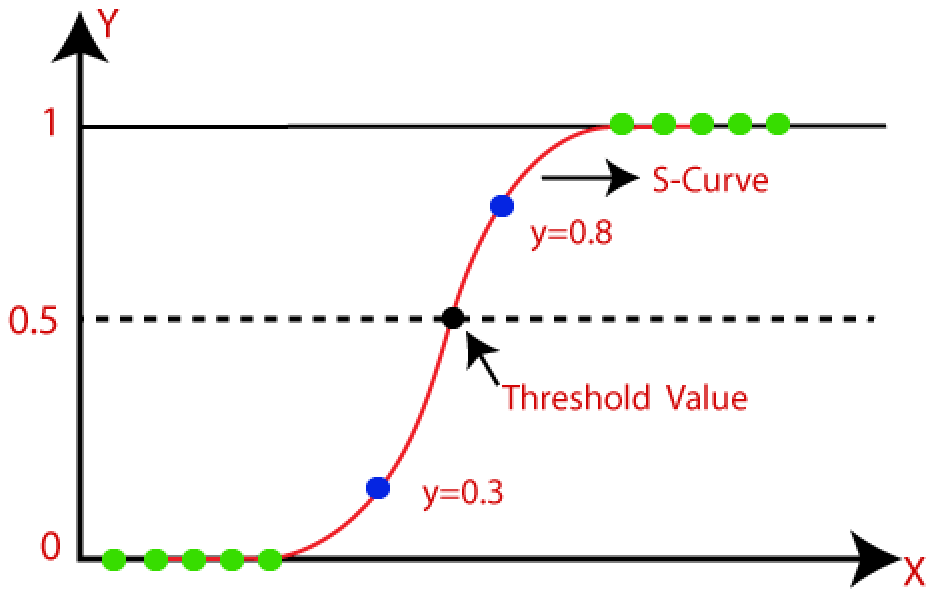
In the classification problems, there are two types of learners:

- **Lazy Learners:** Lazy Learner firstly stores the training dataset and wait until it receives the test dataset. In Lazy learner case, classification is done on the basis of the most related data stored in the training dataset. It takes less time in training but more time for predictions.
  Example: K-NN algorithm, Case-based reasoning
- **Eager Learners:**Eager Learners develop a classification model based on a training dataset before receiving a test dataset. Opposite to Lazy learners, Eager Learner takes more time in learning, and less time in prediction. **Example:** Decision Trees, Naïve Bayes, ANN.

Logistic Regression in Machine Learning

- Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the **categorical** dependent variable using a given set of independent variables.
- Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, **it gives the probabilistic values which lie between 0 and 1**.
- Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas **Logistic regression is used for solving the classification problems**.
- In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).
- The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.
- Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.

- Logistic Regression can be used to classify the observations using different types of data and can easily determine the most effective variables used for the classification.
- The below image is showing the logistic function:
- Logistic regression uses the concept of predictive modeling as regression; therefore, it is called logistic regression, but is used to classify samples; Therefore, it falls under the classification algorithm.



Logistic Function (Sigmoid Function):

- The sigmoid function is a mathematical function used to map the predicted values to probabilities.
- It maps any real value into another value within a range of 0 and 1.
- The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit, so it forms a curve like the "S" form. The S-form curve is called the Sigmoid function or the logistic function.
- In logistic regression, we use the concept of the threshold value, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0.

Assumptions for Logistic Regression:

- The dependent variable must be categorical in nature.
- The independent variable should not have multi-collinearity.

Logistic Regression Equation:

- The Logistic regression equation can be obtained from the Linear Regression equation. The mathematical steps to get Logistic Regression equations are given below:
- We know the equation of the straight line can be written as:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

- In Logistic Regression y can be between 0 and 1 only, so for this let's divide the above equation by (1-y):

$$\frac{y}{1-y} \text{; 0 for y= 0, and infinity for y=1}$$

- But we need range between -[infinity] to +[infinity], then take logarithm of the equation it will become:

$$log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

The above equation is the final equation for Logistic Regression.

Types of Logistic Regression:

- On the basis of the categories, Logistic Regression can be classified into three types:

1. **Binomial:** In binomial Logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, etc.
2. **Multinomial:** In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable, such as "cat", "dogs", or "sheep"
3. **Ordinal:** In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables, such as "low", "Medium", or "High".

Naïve Bayes Classifier Algorithm

- Naïve Bayes algorithm is a supervised learning algorithm, which is based on **Bayes theorem** and used for solving classification problems.
- It is mainly used in *text classification* that includes a high-dimensional training dataset.
- Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
- **It is a probabilistic classifier, which means it predicts on the basis of the probability of an object**.
- Some popular examples of Naïve Bayes Algorithm are **spam filtration, Sentimental analysis, and classifying articles**.

Why is it called Naïve Bayes?

- The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, which can be described as:
- **Naïve**: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the basis of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.
- **Bayes**: It is called Bayes because it depends on the principle of Bayes' Theorem.

Bayes' Theorem

- Bayes' theorem is also known as **Bayes' Rule** or **Bayes' law**, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.
- The formula for Bayes' theorem is given as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- Where,
- **P(A|B) is Posterior probability**: Probability of hypothesis A on the observed event B.
- **P(B|A) is Likelihood probability or Conditional probability**: Probability of the evidence given that the probability of a hypothesis is true.
- **P(A) is Prior Probability**: Probability of hypothesis before observing the evidence.
- **P(B) is Marginal Probability**: Probability of Evidence.

Working of Naïve Bayes' Classifier

- Suppose we have a dataset of **weather conditions** and corresponding target variable "**Play**". So using this dataset we need to decide that whether we should play or not on a particular day according to the weather conditions.
- So to solve this problem, we need to follow the below steps:
  - Convert the given dataset into frequency tables.
  - Generate Likelihood table by finding the probabilities of given features.
  - Now, use Bayes theorem to calculate the posterior probability.

- **Problem**: If the weather is sunny, then the Player should play or not?
- **Solution**: To solve this, first consider the dataset

Outlook	Play	
0	Rainy	Yes
1	Sunny	Yes
2	Overcast	Yes
3	Overcast	Yes
4	Sunny	No
5	Rainy	Yes
6	Sunny	Yes
7	Overcast	Yes
8	Rainy	No
9	Sunny	No
10	Sunny	Yes
11	Rainy	No
12	Overcast	Yes
13	Overcast	Yes

Frequency table for the Weather Conditions:

Weather	Yes	No
Overcast	5	0
Rainy	2	2
Sunny	3	2
Total	10	5

Likelihood table weather condition:

Weather	No	Yes	
Overcast	0	5	5/14= 0.35
Rainy	2	2	4/14=0.29
Sunny	2	3	5/14=0.35
All	4/14=0.29	10/14=0.71	

Applying Bayes'theorem:

P(Yes|Sunny)= P(Sunny|Yes)*P(Yes)/P(Sunny)

P(Sunny|Yes)= 3/10= 0.3

P(Sunny)= 0.35

P(Yes)=0.71

So P(Yes|Sunny) = 0.3*0.71/0.35= **0.60**

P(No|Sunny)= P(Sunny|No)*P(No)/P(Sunny)

P(Sunny|NO)= 2/4=0.5

P(No)= 0.29

P(Sunny)= 0.35

So P(No|Sunny)= 0.5*0.29/0.35 = **0.41**

So as we can see from the above calculation that **P(Yes|Sunny)>P(No|Sunny)**

Hence on a Sunny day, Player can play the game.

Advantages of Naïve Bayes Classifier

- Naïve Bayes is one of the fast and easy ML algorithms to predict a class of datasets.
- It can be used for Binary as well as Multi-class Classifications.
- It performs well in Multi-class predictions as compared to the other Algorithms.
- It is the most popular choice for **text classification problems**.

Disadvantages of Naïve Bayes Classifier

- Naive Bayes assumes that all features are independent or unrelated, so it cannot learn the relationship between features.

Applications of Naïve Bayes Classifier

- It is used for **Credit Scoring**.
- It is used in **medical data classification**.
- It can be used in **real-time predictions** because Naïve Bayes Classifier is an eager learner.
- It is used in Text classification such as **Spam filtering** and **Sentiment analysis**.

Types of Naïve Bayes Model

- There are three types of Naive Bayes Model, which are given below:
- **Gaussian**: The Gaussian model assumes that features follow a normal distribution. This means if predictors take continuous values instead of discrete, then the model assumes that these values are sampled from the Gaussian distribution.
- **Multinomial**: The Multinomial Naïve Bayes classifier is used when the data is multinomial distributed. It is primarily used for document classification problems, it means a particular document belongs to which category such as Sports, Politics, education, etc. The classifier uses the frequency of words for the predictors.
- **Bernoulli**: The Bernoulli classifier works similar to the Multinomial classifier, but the predictor variables are the independent Booleans variables. Such as if a particular word is present or not in a document. This model is also famous for document classification tasks.

Example-1

Consider the given Dataset ,Apply Naive Baye's Algorithm and Predict that if a fruit has the following properties then which type of the fruit it is
Fruit = {Yellow , Sweet ,long}

Frequency Table:

Fruit	Yellow	Sweet	Long	Total
Mango	350	450	0	650
Banana	400	300	350	400
Others	50	100	50	150
Total	800	850	400	1200

Solution: https://medium.com/@balajicena1995/naive-bayes-numerical-example-afcfa2433f95

Examples

- https://codinginfinite.com/naive-bayes-classification-numerical-example/
- https://www.geeksforgeeks.org/naive-bayes-classifiers/
- https://www.machinelearningplus.com/predictive-modeling/how-naive-bayes-algorithm-works-with-example-and-full-code/
- https://www.knowledgehut.com/blog/data-science/naive-bayes-in-machine-learning#what-are-the-applications-of-naive-bayes?%C2%A0

Comparison

LINEAR REGRESSION VERSUS LOGISTIC REGRESSION

LINEAR REGRESSION	LOGISTIC REGRESSION
A linear approach that models the relationship between a dependent variable and one or more independent variables	A statistical model that predicts the probability of an outcome that can only have two values
Used to solve regression problems	Used to solve classification problems (binary classification)
Estimates the dependent variable when there is a change in the independent variable	Calculates the possibility of an event occurring
Output value is continuous	Output value is discrete
Uses a straight line	Uses an S curve or sigmoid function
Ex: predicting the GDP of a country, predicting product price, predicting the house selling price, score prediction	Ex: predicting whether an email is spam or not, predicting whether the credit card transaction is fraud or not, predicting whether a customer will take a loan or not

Visit www.PEDIAA.com

Thank You