



Machine Learning

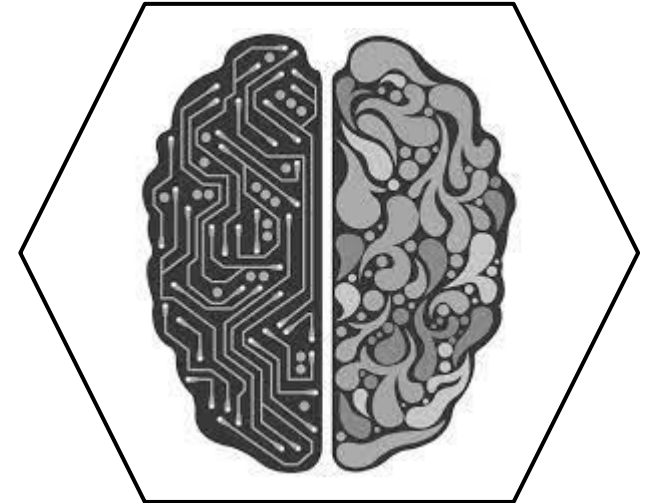
Introduction

Agenda

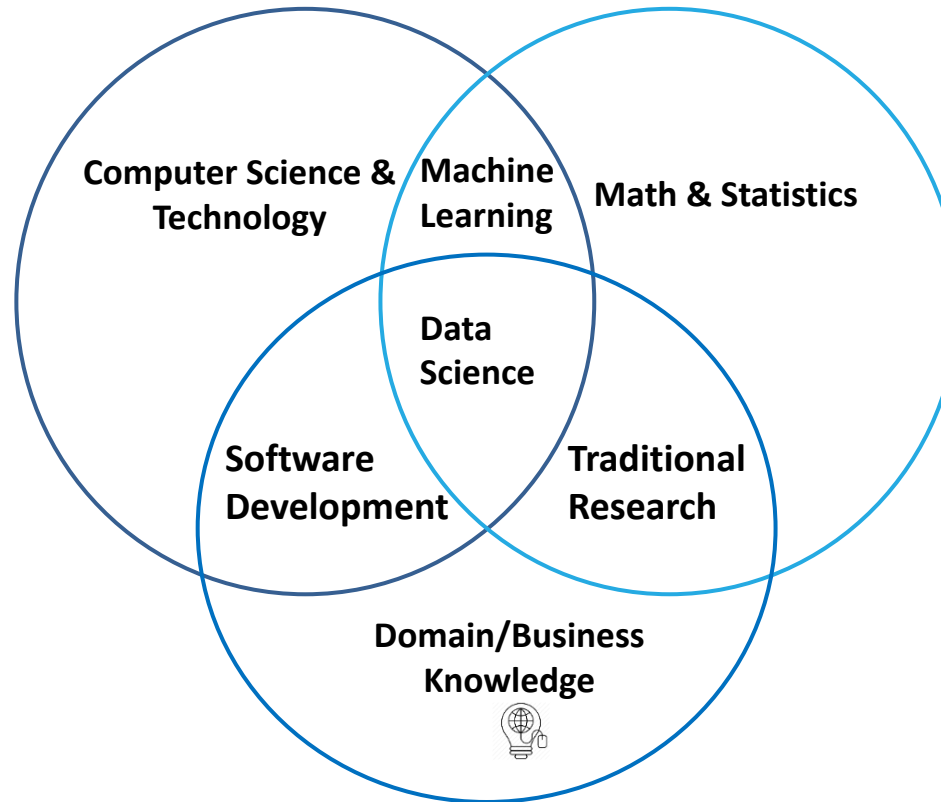
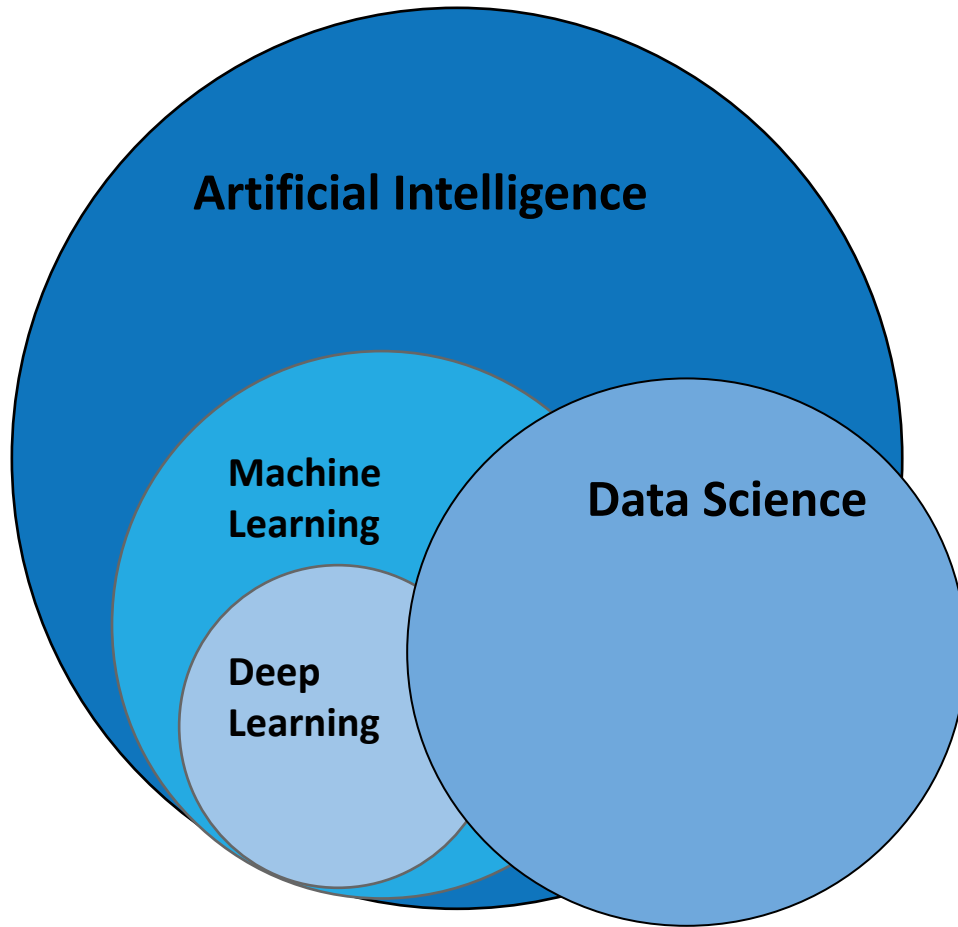
- What is machine learning?
- Machine learning lifecycle
- Data pattern
 - Types of data
 - Types of data for model building
- Types of machine learning
- Supervised learning
 - Regression and classification
- Unsupervised learning
 - Cluster analysis and dimensionality reduction
- Applications of machine learning
- ML libraries and Frameworks
- Limitations of machine learning
- Summary

What is Machine Learning?

Making computers do some task(s) without programming them explicitly.



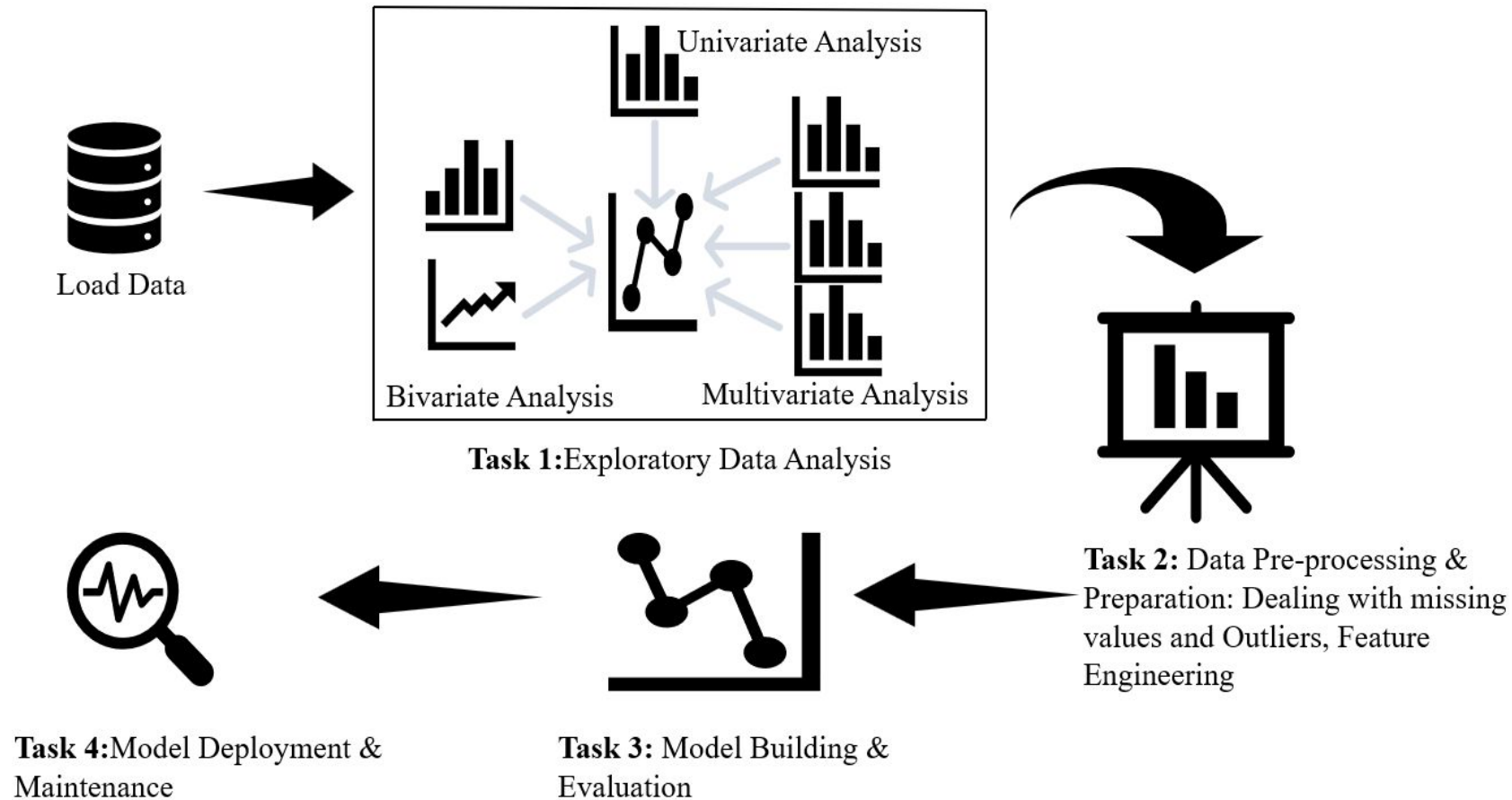
AI, Data Science & Analytics – Big picture



Machine learning

- Machine Learning is the science to make computers learn from data without programming them explicitly and improve their learning over time in an autonomous fashion.
- This learning comes by feeding the data in the form of observations and real-world interactions.
- Machine Learning can also be defined as a tool to predict future events or values using past data.

Machine learning lifecycle



Data pattern

- **Input data can be:**
 - Structured data (ex. Tabular data)
 - Unstructured data (ex. Audio, Video, Text)
- Most of the classical machine learning algorithms are suitable for numerical structured data.
- There are dedicated algorithms for working with image and text data.
- Almost all kinds of data require data pre-processing and preparation for model building.

Type of data

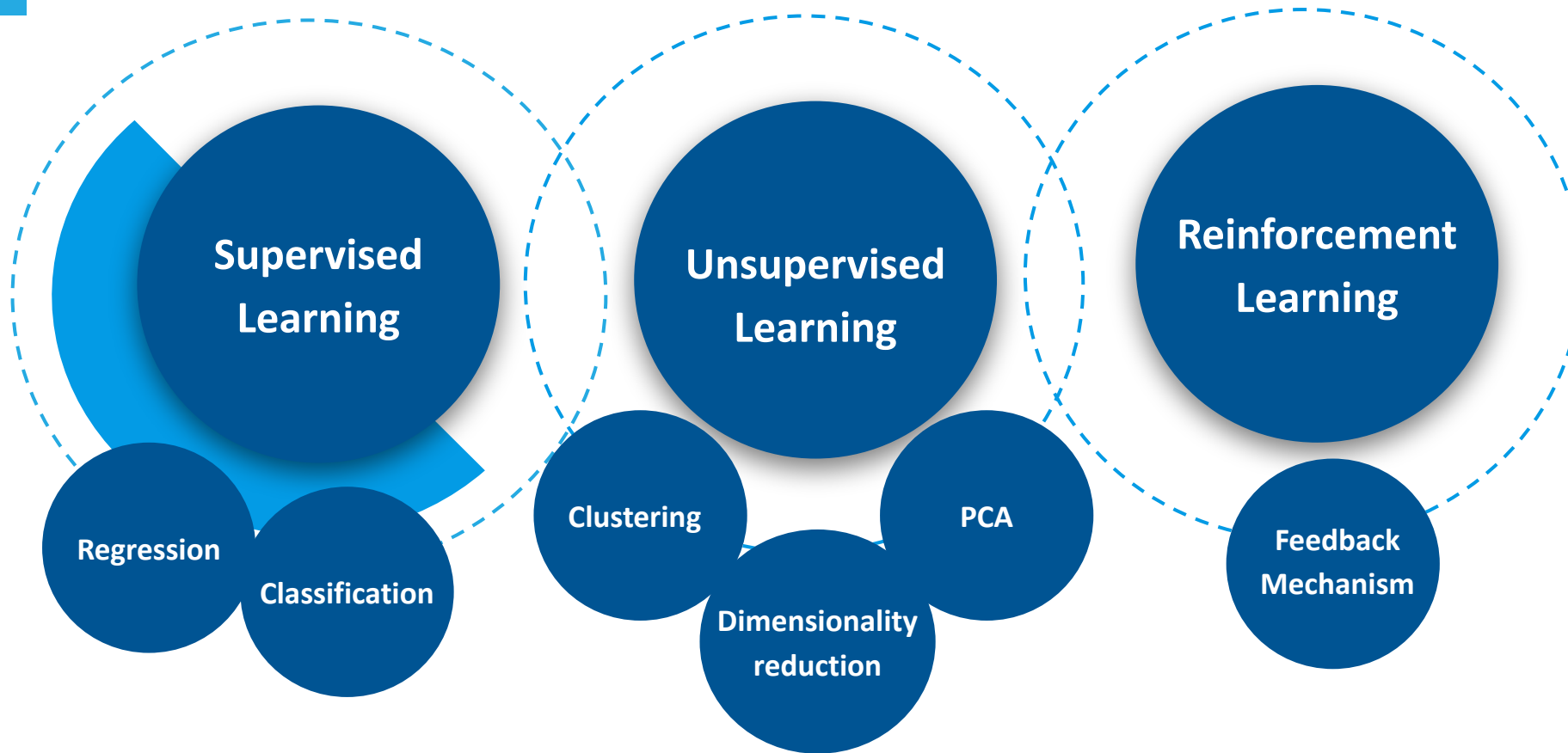
- Data can be divided into two broad categories **based on values**
 - Continuous data (ex. Price-1,2,3,....1000)
 - Categorical data (ex. Gender- Male/Female, Yes/No)
- These categories help to perform suitable EDA analysis.
- They also help to decide which supervised task should be performed.

Type of data for model building

For choosing a machine learning model, available data is categorised into:

- **Labelled data** – consists of input output pair. For every set input features the output/response/label is present in dataset.
 - ex- labelled image as cat's or dog's photo
 - Sample structure: $\{(x_1, y_1), (x_2, y_2), (x_3, y_3) \dots (x_n, y_n)\}$
 - Supervised model can be developed when the label is available
- **Unlabelled data** - There is no output/response/label for the input features in data.
 - ex. news articles, tweets, audio
 - Sample structure: $\{x_1, x_2, x_3, x_4, x_5 \dots x_n\}$
 - Unsupervised models are built on unlabelled data

Types of Machine Learning

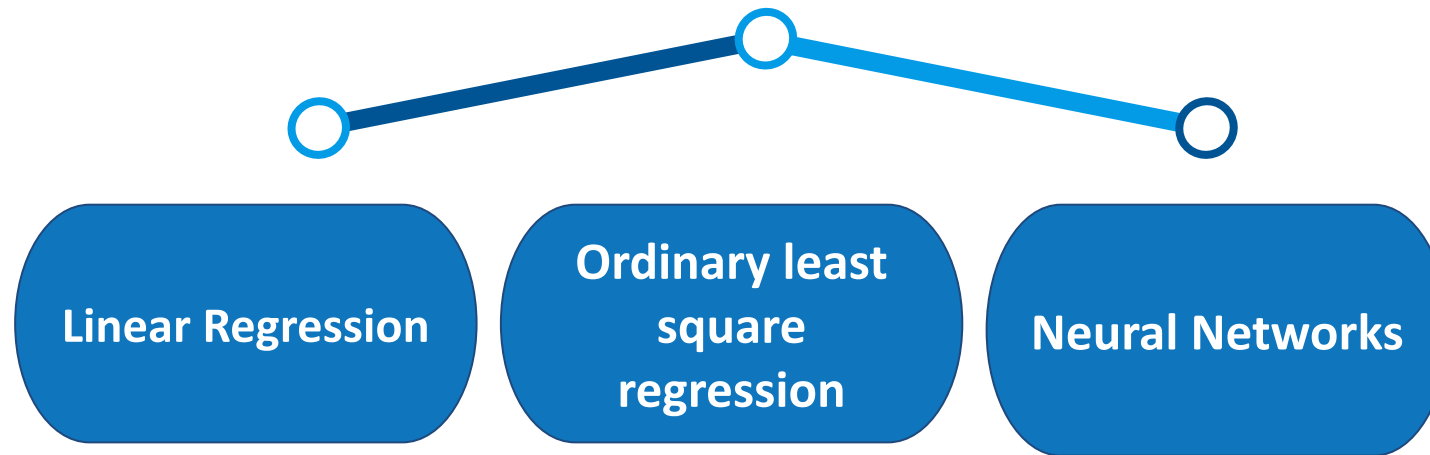


Supervised learning

- Class of machine learning that work on externally supplied instances in form of predictor attributes and **associated target values**.
- The model learns from the training data using these '**target variables**' as reference variables.
 - Ex1 : model to predict the resale value of a car based on its mileage, age, color etc.
- The **target values** are the 'correct answers' for the predictor model which can either be a **regression model** or a **classification model**.

Supervised learning - Regression

Regression



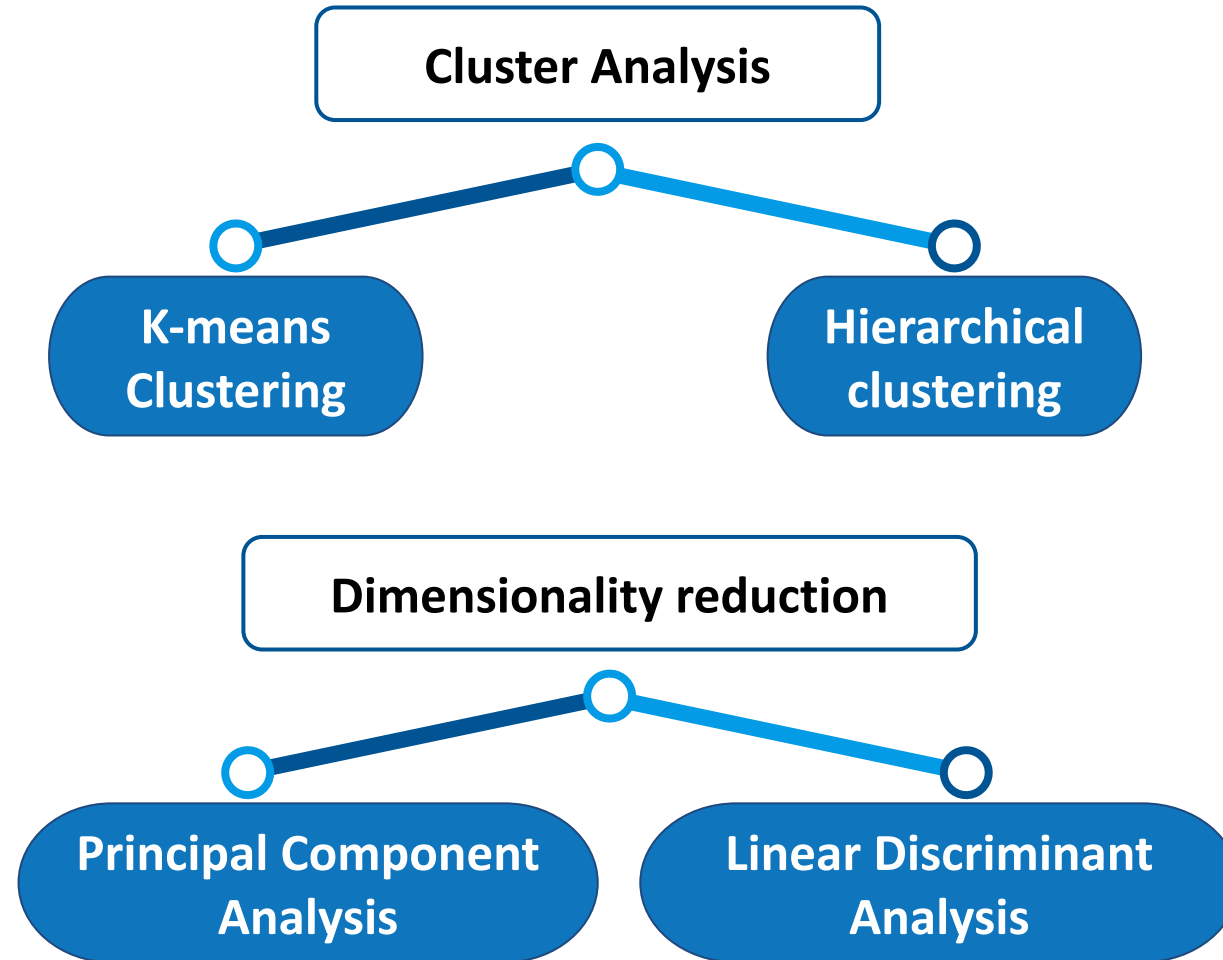
Supervised learning - Classification

Classification



Note:- Most of these algorithms can be used for regression as well.

Unsupervised learning



How to use data for building a ML model?

- Since we work with a sample of data, we cannot use the entire dataset for training process. A small sample is saved for testing the model.
- **Training Data** – Sample data points which are used to train the machine learning model.
- **Test Data** - Sample data points that are used to test the performance of machine learning model.

Note- For modelling, the original dataset is partitioned into the ratio of 70:30 or 75:25 as training data and test data.

Applications of ML

Machine learning has applications in almost all fields wherever we have digital data in any format.

- Medical diagnosis
- Agriculture
- Banking
- Computer vision
- Finance
- Natural language processing
- Recommender systems
- Speech recognition
- Time series forecasting

ML libraries and Frameworks

- Scikit-learn
- TensorFlow
- Keras
- PyTorch
- Theano
- MXNet
- Caffe etc
- All these are free and open source



Limitations of ML

There are some limitations and scope of improvement as well.

- Related to data
 - Lack of suitable data & human bias in the data
 - Data privacy and ethical issues
 - Rapid changes in the data
- Related to models
 - Biased models
 - Poor performance in production
 - Regular training required
 - Black box models
- Related to infrastructure
 - Expensive infrastructure requirement

Summary

In this module we discussed:

- What is machine learning and overview of the machine learning lifecycle.
- Data Patterns and types of data for model building.
- Types of machine learning.
- Supervised learning and Unsupervised learning.
- Applications of machine learning.
- ML libraries and Frameworks.
- Limitations of machine learning.

Thank you!

Happy Learning :)