Course: Artificial Intelligence and Machine Learning Code: 20CS51I

WEEK- 2 Machine Learning

- **Fundamentals**
- **Machine learning types**
- **-Machine learning workflow**
- **Machine learning applications**
- **Challenges in ML**
- **Building a model – steps involved**

Session No. 1

Machine Learning: Fundamentals

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

(OR)

Machine learning is a subset of AI, which enables the machine to automatically learn from data, improve performance from past experiences, and make predictions.

(OR)

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that makes it more similar to humans: The ability to learn. Machine learning is actively being used today, perhaps in many more places than one would expect.

Machine learning types

1. Supervised Machine Learning

As its name suggests, Supervised machine learning is based on supervision. It means in the supervised learning technique, we train the machines using the "labelled" dataset, and based on the training, the machine predicts the output.

Supervised learning works on labelled data. Each input data has a corresponding labelled output. The goal of supervised machine learning is to learn a mapping from the input to the output. The input data is called attributes, features or predictors. This output variable is also called response variable or target variable.

For example, the problem of building a utility for predicting the selling price of the car. The dataset is shown below:

Brand	Car Model	Car Registration Year	Miles driven	Price($)
Nissan	Sunny	2010	90000	10000
Toyota	Corolla	2009	45000	25000
Honda	Brio	2015	70000	18000
Chevrolet	Cruze	2013	60000	30000
Nissan	Terrano	2018	12000	40000
Honda	City	2017	25000	45000
.....				
.....				

Input Variables (X) Output Variable (Y)

From the given dataset, the machine learning algorithm learns the mapping from input variables to output variable. This learning is represented in the form of a model. When a new instance is given to the model as shown below, it can predict its output value.

Brand	Car Model	Car Registration Year	Miles driven	Price($)
Honda	City	2012	15000	?

Some of the other examples of supervised learning:

Given an email defined by its collection of phrases(X), predict if the mail is a spam(Y).

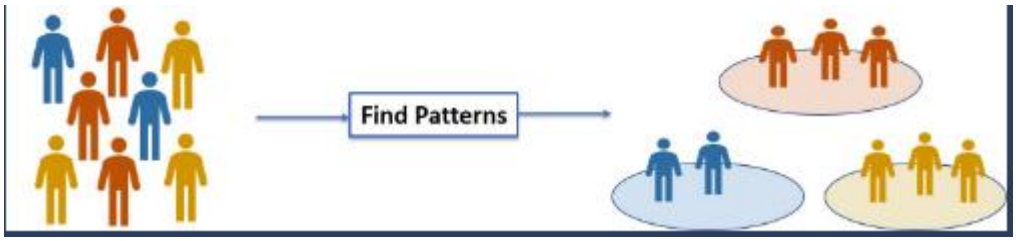
Given a medical brain scan image (X), predict if the patient has tumour(Y).

2. Unsupervised Machine Learning

unsupervised learning algorithm is to group or categories the unsorted dataset according to the similarities, patterns, and differences. Machines are instructed to find the hidden patterns from the input dataset.

Unsupervised machine learning has no explicitly defined output. The idea is to discover knowledge or structure in the data.

For example, an online retailer will have data about all items that the customers purchased. Unsupervised learning algorithms can be applied on this data to group customers based on their buying patterns.

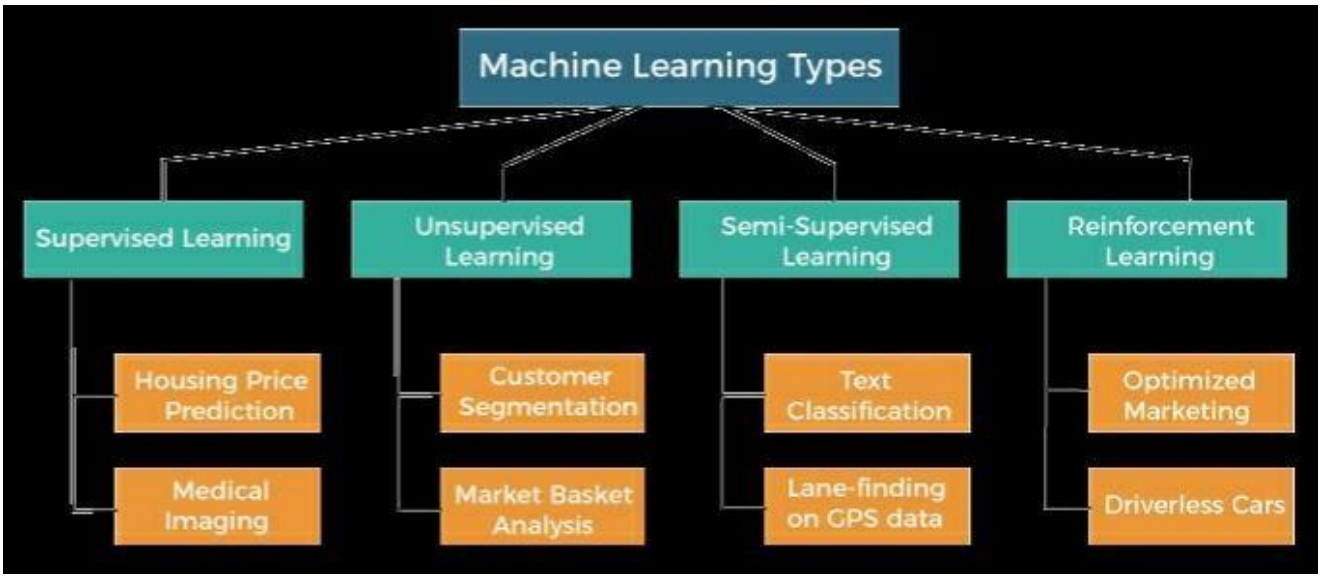


Grouping new articles based on topics like sports, politics, business etc. is another example of unsupervised learning.

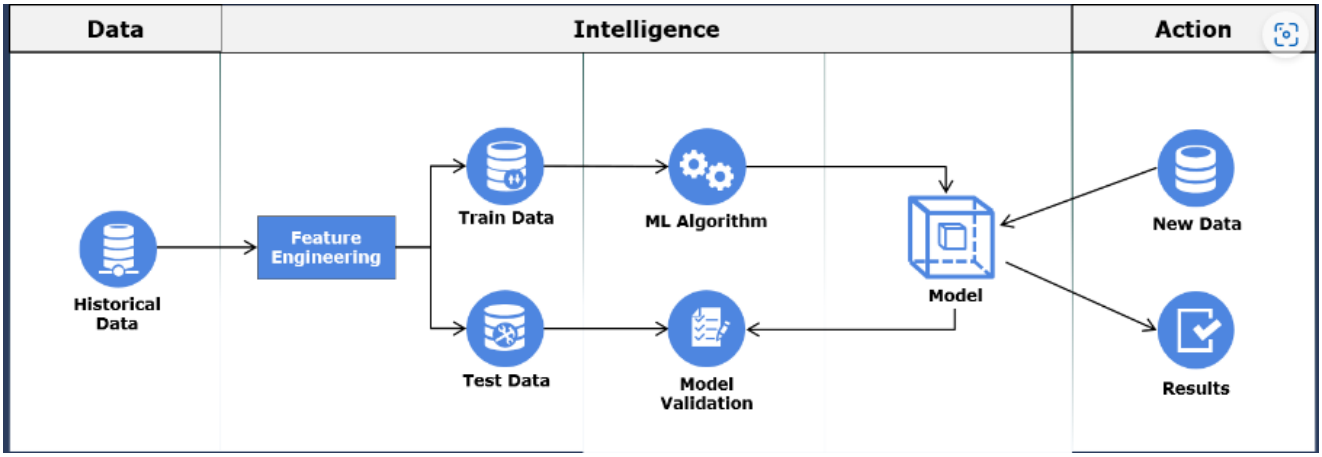
This task of discovering inherent clusters or groups in the data is known as Clustering

3. Reinforcement Learning

Reinforcement learning works on a feedback-based process, in which an AI agent (A software component) automatically explore its surrounding by hitting & trail, taking action, learning from experiences, and improving its performance.



Machine Learning Process



The diagram above illustrates the Machine Learning process.

In this process, first relevant data is gathered then is cleaned and transformed through a process called Feature Engineering. During the process of Feature Engineering, handling missing value, handling outliers, creating new features out of existing ones are some of the common tasks performed.

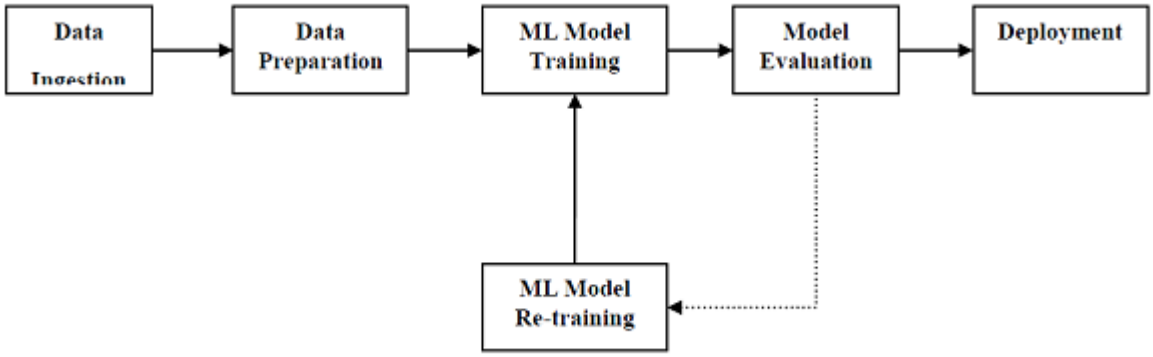
After feature engineering, the data is split into Train Data and Test Data. The Train Data is used for training the machine learning model. Once the model is built, it is validated against the Test Data for accuracy. This accuracy helps us in estimating the performance on previously unseen data. If the model performance on both Train and Test Data is satisfactory, the model may be deployed.

Once deployed, the model makes predictions on new data ; these predictions/insights are used to take business decisions

Machine learning workflow

In order to execute and produce results successfully, a machine learning model must automate some standard workflows. The process of automate these standard workflows can be done with the help of Scikit-learn Pipelines. From a data scientist's perspective, pipeline is a generalized, but very important concept. It basically allows data flow from its raw format to some useful information.

The working of pipelines can be understood with the help of following diagram −



The blocks of ML pipelines are as follows −

Data ingestion − As the name suggests, it is the process of importing the data for use in ML project. The data can be extracted in real time or batches from single or multiple systems. It is one of the most challenging steps because the quality of data can affect the whole ML model.

Data Preparation − After importing the data, we need to prepare data to be used for our ML model. Data preprocessing is one of the most important technique of data preparation.

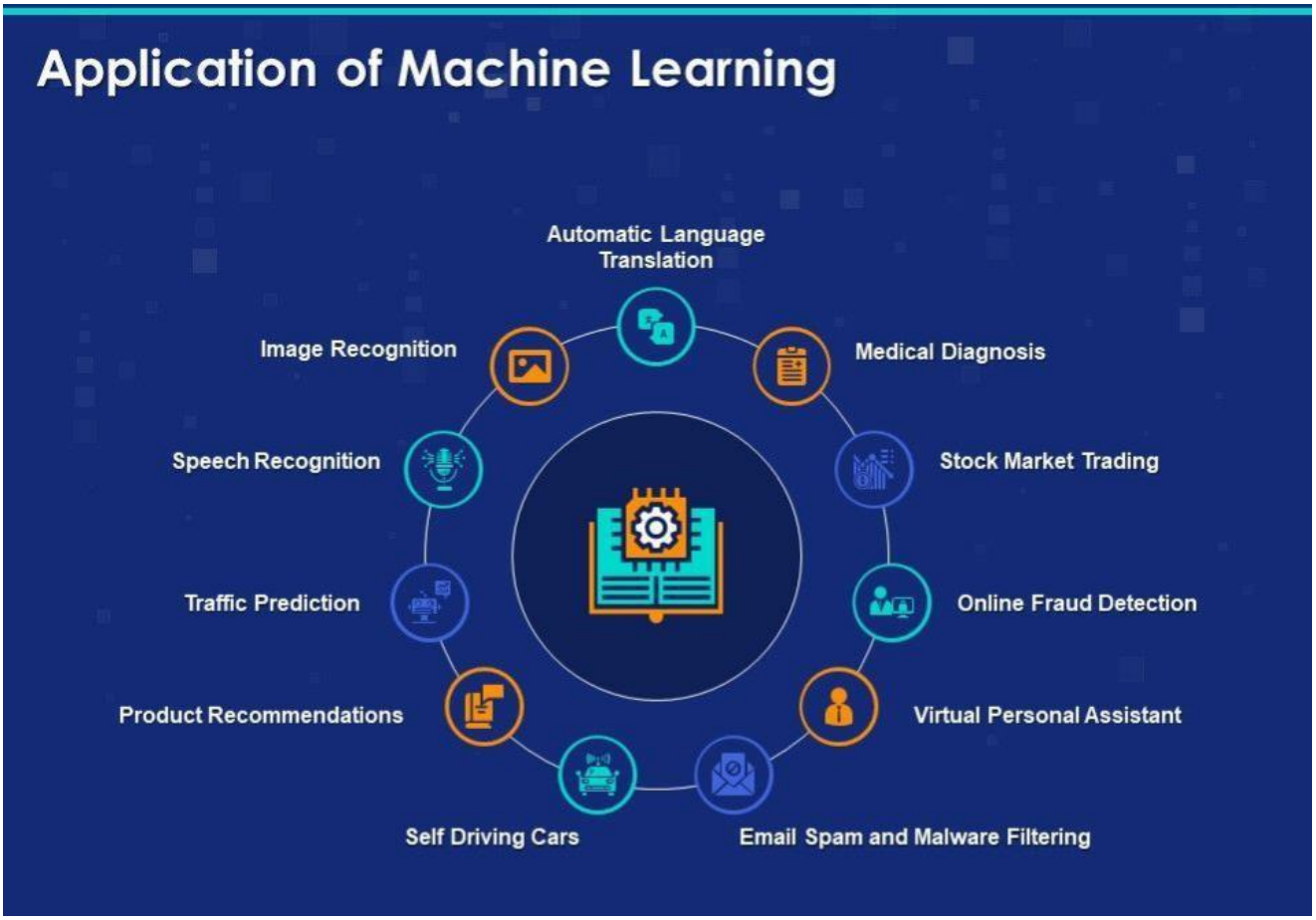
ML Model Training − Next step is to train our ML model. We have various ML algorithms like supervised, unsupervised, reinforcement to extract the features from data, and make predictions.

Model Evaluation − Next, we need to evaluate the ML model. In case of AutoML pipeline, ML model can be evaluated with the help of various statistical methods and business rules.

ML Model retraining − In case of AutoML pipeline, it is not necessary that the first model is best one. The first model is considered as a baseline model and we can train it repeatably to increase model's accuracy.

Deployment − At last, we need to deploy the model. This step involves applying and migrating the model to business operations for their use.

Machine learning Applications



4. Challenges in ML

- ✓ Inadequate Training Data
- ✓ Poor quality of data
- ✓ Monitoring and maintenance
- ✓ Getting bad recommendations
- ✓ Lack of skilled resources
- ✓ Process Complexity of Machine Learning
- ✓ Data Bias
- ✓ Slow implementations and results

Building model steps involved

Step 1: Collect Data

Given the problem you want to solve, you will have to investigate and obtain data that you will use to feed your machine.

Step 2: Prepare the data

This is a good time to visualize your data and check if there are correlations between the different characteristics that we obtained.

You must also separate the data into two groups: one for training and the other for model evaluation which can be divided approximately in a ratio of 80/20 but it can vary depending on the case and the volume of data we have.

At this stage, you can also pre-process your data by normalizing, eliminating duplicates, and making error corrections.

Step 3: Choose the model

you will use algorithms of classification, prediction, linear regression, clustering, i.e. k-means or K-Nearest Neighbor, Deep Learning, i.e Neural Networks, Bayesian, etc.

There are various models to be used depending on the data you are going to process such as images, sound, text, and numerical values.

In the following table, we will see some models and their applications that you can apply in your projects:

Model	**Applications**
Logistic Regression	Price prediction
Fully connected networks	Classification
Convolutional Neural Networks	Image processing
Recurrent Neural Networks	Voice recognition
Random Forest	Fraud Detection
Reinforcement Learning	Learning by trial and error
Generative Models	Image creation
K-means	Segmentation

Model	Applications
k-Nearest Neighbors	Recommendation systems
Bayesian Classifiers	Spam and noise filtering

Step 4 Train your machine model

You will need to train the datasets to run smoothly and see an incremental improvement in the prediction rate.

Step 5: Evaluation

You will have to check the machine created against your evaluation data set of your already trained model.

If the accuracy is less than or equal to 50%, that model will not be useful.

If you reach 90% or more, you can have good confidence in the results that the model gives you.

Step 6: Parameter Tuning

If during the evaluation you did not obtain good predictions, you must return to the training step before making a new configuration of parameters in your model.

Step 7: Prediction or Inference

You are now ready to use your Machine Learning model inferring results in real-life scenarios.

Examples of ML in Daily life