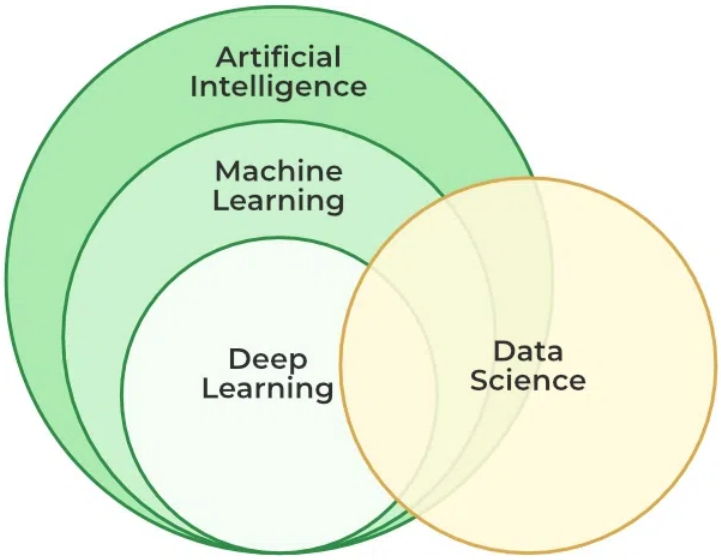
Lecture 1 - Introduction to Machine Learning

Preamble

Machine learning (ML) is a branch of AI that develops algorithms by learning the hidden patterns of a datasets and uses the patterns to make predictions on new similar type data, without being explicitly programmed for each task. **Artificial intelligence (AI)** is a broad term that refers to intelligent systems that act and think like humans. A crucial distinction is that, while all machine learning is AI, not all AI is machine learning.

Machine learning is used to make decisions based on data. By modelling the algorithms on the bases of historical data, Algorithms find the patterns and relationships that are difficult for humans to detect. These patterns are now further used for future references to predict solution of unseen problems.



ML

Where is ML used ?

Machine learning is used in many different applications, from;

a) Image and speech recognition to natural language processing,
b) Recommendation systems,
c) Fraud detection,
d) Portfolio optimization,
e) Automated task,
f) Machine learning models are also used to power autonomous vehicles,
g) Drones, and robots, making them more intelligent and adaptable to changing environments.

How machine learning algorithms work

Machine Learning works in the following manner.

- **Forward Pass:** In the Forward Pass, a machine learning algorithm takes in input data and produces an output. Depending on the model algorithm it computes the predictions.
- **Loss Function:** The loss function, also known as the error or cost function, is used to evaluate the accuracy of the predictions made by the model. The function compares the predicted output of the model to the actual output and calculates the difference between them. This difference is known as error or loss. *The goal of any model is to minimize the error or loss function by adjusting its internal parameters*.
- **Model Optimization Process:** The model optimization process is the iterative process of adjusting the internal parameters of the model to minimize the error or loss function. This is done using an optimization algorithm, such as **gradient descent**. The optimization algorithm calculates the gradient of the error function with respect to the model's parameters and uses this information to adjust the parameters to reduce the error. The algorithm repeats this process until the error is minimized to a satisfactory level.

Once the model has been trained and optimized on the training data, it can be used to make predictions on new, unseen data. The accuracy of the model's predictions can be evaluated using various performance metrics, such as accuracy, precision, recall, and F1-score.

Machine Learning lifecycle:

The lifecycle of a machine learning projects involves a series of steps that include:

1. **Study the Problems:** The first step is to study the problem. This step involves understanding the business problem and defining the objectives of the model.
2. **Data Collection:** When the problem is well-defined, we can collect the relevant data required for the model. The data could come from various sources such as databases like Kaggle, APIs, or web scraping.
3. **Data Preparation:** When our problem-related data is collected. then it is a good idea to check the data properly and make it in the desired format so that it can be used by the model to find the hidden patterns. This can be done in the following steps:
    - Data cleaning
    - Data Transformation
    - Explanatory Data Analysis and Feature Engineering
    - Split the dataset for training and testing.

4. **Model Selection:** The next step is to select the appropriate machine learning algorithm that is suitable for our problem. This step requires knowledge of the strengths and weaknesses of different algorithms. Sometimes we use multiple models and compare their results and select the best model as per our requirements.
5. **Model building and Training:** After selecting the algorithm, we have to build the model. In the case of traditional machine learning building mode is easy it is just a few hyperparameter tunings. In the case of deep learning, we have to define layer-wise architecture along with input and output size, number of nodes in each layer, loss function, gradient descent optimizer, etc.

After that model is trained using the preprocessed dataset.

6. **Model Evaluation:** Once the model is trained, it can be evaluated on the test dataset to determine its accuracy and performance using different techniques like classification report, F1 score, precision, recall, ROC Curve, Mean Square error, absolute error, etc.
7. **Model Tuning:** Based on the evaluation results, the model may need to be tuned or optimized to improve its performance. This involves tweaking the hyperparameters of the model.
8. **Deployment:** Once the model is trained and tuned, it can be deployed in a production environment to make predictions on new data. This step requires integrating the model into an existing software system or creating a new system for the model.
9. **Monitoring and Maintenance:** Finally, it is essential to monitor the model's performance in the production environment and perform maintenance tasks as required. This involves monitoring for data drift, retraining the model as needed, and updating the model as new data becomes available.

Types of Machine Learning

- Supervised Machine Learning
- Unsupervised Machine Learning
- Reinforcement Machine Learning

Supervised Machine Learning:

Supervised learning is a type of machine learning in which the algorithm is trained on the labeled dataset. It learns to map input features to targets based on labeled training data. In supervised learning, the algorithm is provided with input features

and corresponding output labels, and it learns to generalize from this data to make predictions on new, unseen data.

There are two main types of supervised learning:

- **Regression**: Regression is a type of supervised learning where the algorithm learns to predict continuous values based on input features. The output labels in regression are continuous values, such as stock prices, and housing prices. The different regression algorithms in machine learning are: Linear Regression, Polynomial Regression, Ridge Regression, Decision Tree Regression, Random Forest Regression, Support Vector Regression, etc
- **Classification**: Classification is a type of supervised learning where the algorithm learns to assign input data to a specific category or class based on input features. The output labels in classification are discrete values. Classification algorithms can be binary, where the output is one of two possible classes, or multiclass, where the output can be one of several classes. The different Classification algorithms in machine learning are: Logistic Regression, Naive Bayes, Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), etc.

2. Unsupervised Machine Learning:

Unsupervised learning is a type of machine learning where the algorithm learns to recognize patterns in data without being explicitly trained using labeled examples. The goal of unsupervised learning is to discover the underlying structure or distribution in the data.

There are two main types of unsupervised learning:

- **Clustering**: Clustering algorithms group similar data points together based on their characteristics. The goal is to identify groups, or clusters, of data points that are similar to each other, while being distinct from other groups. Some popular clustering algorithms include K-means, Hierarchical clustering, and Density-based spatial clustering like DBSCAN.
- **Dimensionality reduction:** Dimensionality reduction algorithms reduce the number of input variables in a dataset while preserving as much of the original information as possible. This is useful for reducing the complexity of a dataset and making it easier to visualize and analyze. Some popular dimensionality reduction algorithms include Principal Component Analysis (PCA), t-SNE, and Autoencoders.

3. Reinforcement Machine Learning

Reinforcement learning is a type of machine learning where an agent learns to interact with an environment by performing actions and receiving rewards or penalties based on its actions/results. The goal of reinforcement learning is to learn

a policy, which is a mapping function from states to actions, that maximizes the expected cumulative reward over time.

There are two main types of reinforcement learning:

- **Model-based reinforcement learning:** In model-based reinforcement learning, the agent learns a model of the environment, including the transition probabilities between states and the rewards associated with each state-action pair. The agent then uses this model to plan its actions in order to maximize its expected reward. Some popular model-based reinforcement learning algorithms include Value Iteration and Policy Iteration.
- **Model-free reinforcement learning:** In model-free reinforcement learning, the agent learns a policy directly from experience without explicitly building a model of the environment. The agent interacts with the environment and updates its policy based on the rewards it receives. Some popular model-free reinforcement learning algorithms include Q-Learning, SARSA, and Deep Reinforcement Learning.

Common ML Algorithms

A number of machine learning algorithms are commonly used. These include:

- **Neural networks:** Neural networks simulate the way the human brain works, with a huge number of linked processing nodes. Neural networks are good at recognizing patterns and play an important role in applications including natural language translation, image recognition, speech recognition, and image creation.
- **Linear regression:** This algorithm is used to predict numerical values, based on a linear relationship between different values. For example, the technique could be used to predict house prices based on historical data for the area.
- **Logistic regression:** This supervised learning algorithm makes predictions for categorical response variables, such as“yes/no” answers to questions. It can be used for applications such as classifying spam and quality control on a production line.
- **Clustering:** Using unsupervised learning, clustering algorithms can identify patterns in data so that it can be grouped. Computers can help data scientists by identifying differences between data items that humans have overlooked.
- **Decision trees:** Decision trees can be used for both predicting numerical values (regression) and classifying data into categories. Decision trees use a branching sequence of linked decisions that can be represented with a tree diagram. One of the advantages of decision trees is that they are easy to validate and audit, unlike the black box of the neural network.

- **Random forests:** In a random forest, the machine learning algorithm predicts a value or category by combining the results from a number of decision trees.

Where is ML being used today ?

- **Automation**: Machine learning, which works entirely autonomously in any field without the need for any human intervention. For example, robots perform the essential process steps in manufacturing plants.
- **Finance Industry**: Machine learning is growing in popularity in the finance industry. Banks are mainly using ML to find patterns inside the data but also to prevent fraud.
- **Government organization**: The government makes use of ML to manage public safety and utilities. Take the example of China with its massive face recognition. The government uses Artificial intelligence to prevent jaywalking.
- **Healthcare industry**: Healthcare was one of the first industries to use machine learning with image detection.
- **Marketing:** Broad use of AI is done in marketing thanks to abundant access to data. Before the age of mass data, researchers develop advanced mathematical tools like Bayesian analysis to estimate the value of a customer. With the boom of data, the marketing department relies on AI to optimize customer relationships and marketing campaigns.
- **Retail industry**: Machine learning is used in the retail industry to analyze customer behavior, predict demand, and manage inventory. It also helps retailers to personalize the shopping experience for each customer by recommending products based on their past purchases and preferences.
- **Transportation**: Machine learning is used in the transportation industry to optimize routes, reduce fuel consumption, and improve the overall efficiency of transportation systems. It also plays a role in autonomous vehicles, where ML algorithms are used to make decisions about navigation and safety

Challenges and Limitations of Machine Learning

1. The primary challenge of machine learning is the lack of data or the diversity in the dataset.
2. A machine cannot learn if there is no data available. Besides, a dataset with a lack of diversity gives the machine a hard time.

3. A machine needs to have heterogeneity to learn meaningful insight.
4. It is rare that an algorithm can extract information when there are no or few variations.

Others

5. AI surpassing human intelligence in the near future. Technological singularity is also referred to as strong AI or superintelligence
6. AI impact on jobs
7. Instances of bias and discrimination across a number of machine learning systems e.g. dataset gathering, Recruitment system, hotels recommendations
8. There isn't significant legislation to regulate AI practices