ACADEMIC CITY
UNIVERSITY COLLEGE

MACHINE LEARNING

Instructor: Ishaya, Jeremiah Ayock

Lecture 1: Introduction to Machine Learning

January 27, 2025

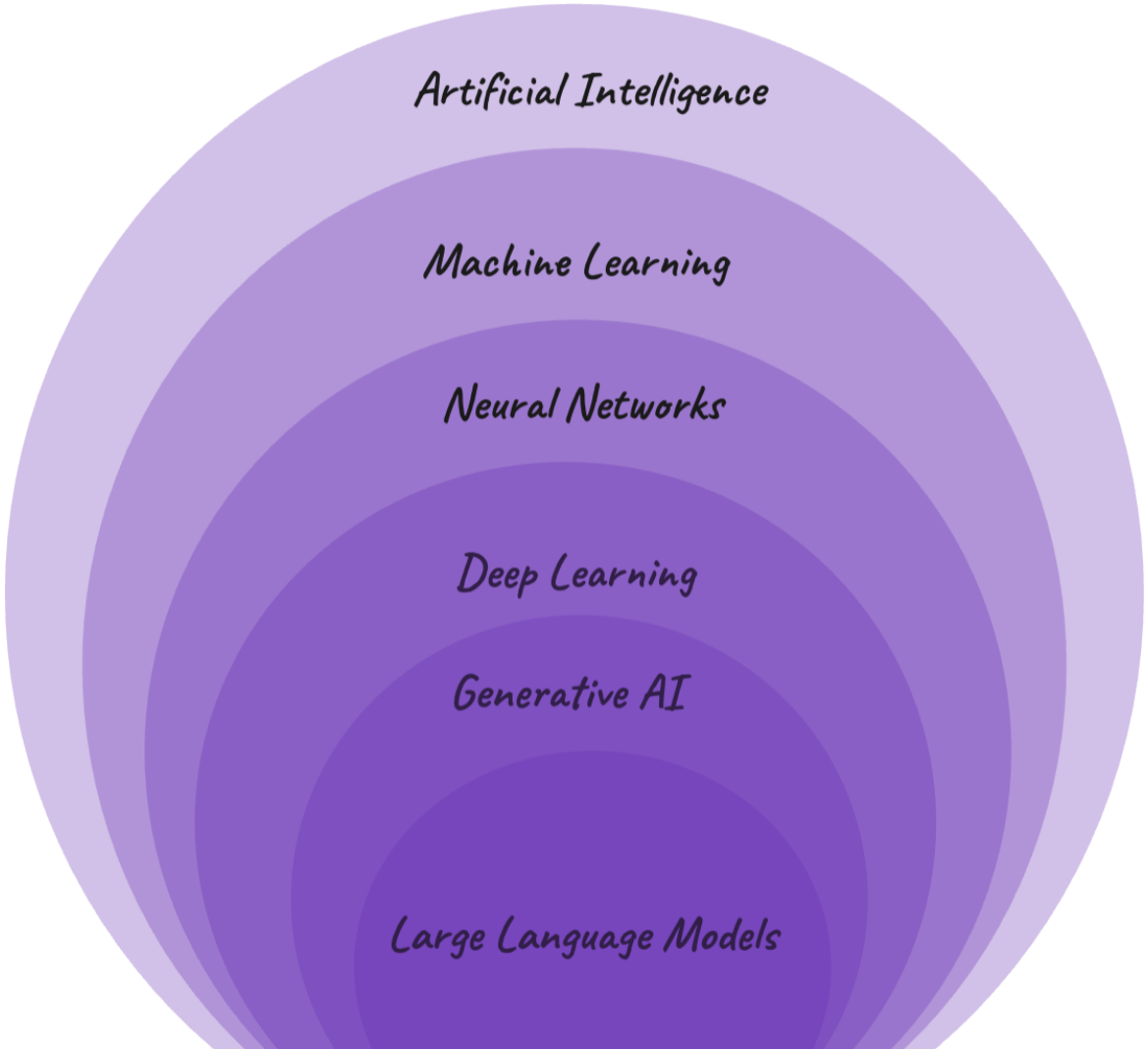
Academic City University College, Agbogba Haatso, Ghana.

OBJECTIVE

-
-
- Explore the scope and applications of Machine Learning

Machine Learning

AI, MACHINE LEARNING, DEEP LEARNING, GENAI



INTRODUCTION

Machine learning is a subfield of **Artificial Intelligence** that involves building **algorithms and models** that can learn from data and make predictions or decisions based on that learning.

Machine learning algorithms are designed to **automatically identify patterns** and **relationships** in **data** without being explicitly programmed.

INTRODUCTION CON'T

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

INTRODUCTION CON'T

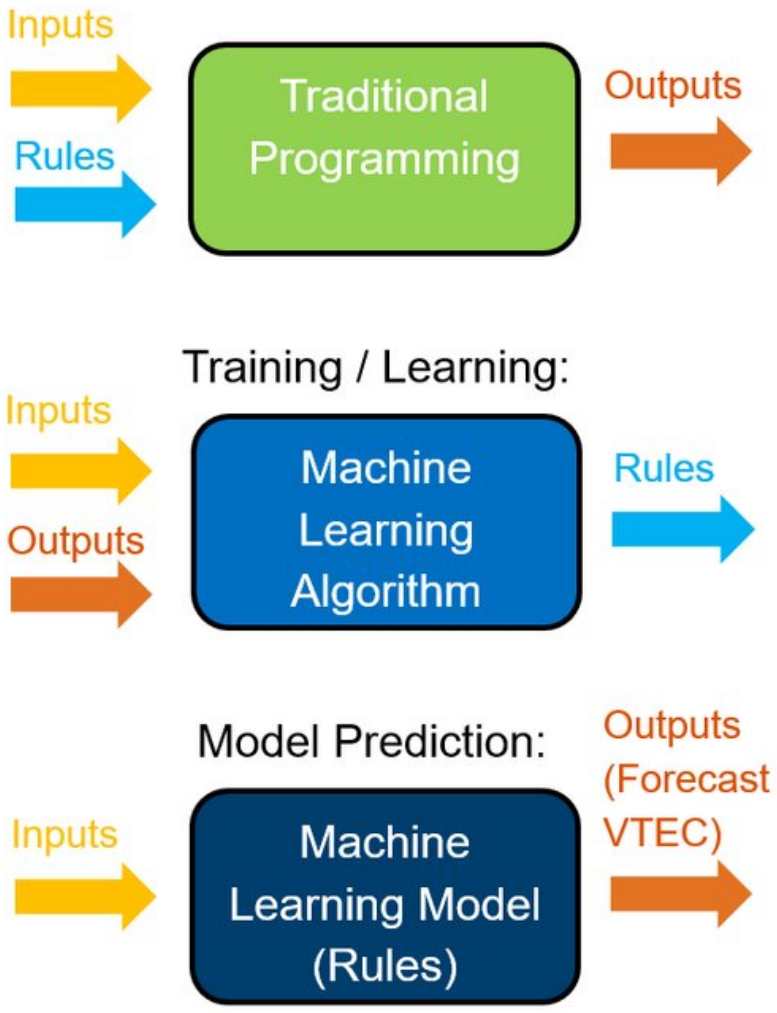


Figure 2: ML and Traditional Programming

MACHINE LEARNING WORKING PRINCIPLE

Machine Learning algorithms use a variety of techniques to identify patterns in data. Some of the most common techniques include **supervised learning, unsupervised learning, and reinforcement learning.**

These techniques involve training algorithms on **large datasets**, using statistical methods to identify patterns, and optimizing algorithms to improve performance.

TYPES OF MACHINE LEARNING

There are three basic approaches, and these are:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

SUPERVISED LEARNING

Supervised Learning is a type of Machine Learning where the algorithm is trained on **labelled data**.

The algorithm is provided with **input data** and corresponding **output data**, and the goal is to **learn a mapping from inputs to outputs.**

The most common types of Supervised Learning are **Regression and Classification.**

UNSUPERVISED LEARNING

Unsupervised learning is a type of machine learning in which the algorithm is trained on **unlabeled data**.

The goal of Unsupervised Learning is to **identify patterns** and **relationships** in **the data without any prior knowledge of what those patterns might be**.

The most common types of Unsupervised Learning are **Clustering and Association.**

REINFORCEMENT LEARNING

Reinforcement Learning is a type of Machine Learning where the algorithm is **trained to make decisions based on rewards and punishments**.

The algorithm learns through **trial and error**, and the goal is to **maximize the cumulative reward over time.**

Supervised Learning

SUPERVISED LEARNING

Supervised Learning is a type of machine learning, where data scientists provide algorithms with labelled training data and define the variables they want the algorithm to assess for correlations.

Both the **input** and the **output** of the algorithm are specified.

SUPERVISED LEARNING

When we use labels in our machine learning model, we deal with the supervised learning problem. It is defined by its use of labelled datasets to train algorithms that classify data or predict outcomes accurately.

For example, in the **USA Housing dataset**, we have been given the prices of houses. As input data is fed into the model, it adjusts its weights until the model has been fitted appropriately.

Supervised learning helps organizations solve a variety of real-world problems at scale, such as **classifying spam** in a separate folder from your inbox.

SUPERVISED LEARNING CON'T

Supervised learning uses a training set to teach models to yield the desired output. This training dataset has both the right inputs and outputs, enabling the model to develop over time.

The **loss function** serves as a measurement for the algorithm's accuracy and iterations are made until the error is sufficiently reduced.

SUPERVISED LEARNING CON'T

Supervised learning algorithms are good for the following tasks:

- **Binary Classification:** Categorising data into two classes.
- **Multi-class Classification:** Categorising data into more than two classes.
- **Regression Modeling:** Predicting continuous values.
- **Ensembling:** Combining the predictions of multiple machine learning models to produce an accurate single prediction prediction.

TYPES OF SUPERVISED MACHINE LEARNING

There are two types of supervised learning algorithms:

- Classification
- Regression

CLASSIFICATION

Classification is the process of organizing a given set of data into **classes**. It can be performed on both structured and unstructured data. The classes are often referred to as **targets, labels, or categories.**

If the output labels have discrete values, it is a classification problem. Classification uses an algorithm to accurately assign test data into specific categories. It recognizes specific entities within the dataset and attempts to draw some conclusions on how those entities should be labelled or defined.

CLASSIFICATION ALGORITHMS

Below are some popular algorithms associated with the classification type of supervised machine learning.

- Logistics Regression
- Decision trees
- Random Forest
- Naive Bayes
- Support Vector Machines (SVM)
- K-Nearest Neighbor

POPULAR CLASSIFICATION PROBLEMS

Below are the most common classification problems:

- Sentiment Analysis
- Disease Classification
- Speech recognition
- Face detection
- Handwriting recognition
- Document classification

Regression

REGRESSION

Regression is a type of supervised learning that maps a predictive relationship between labels and continuous data points.

- Regression is used to understand the relationship between dependent and independent variables.
- If the **output labels have a continuous range of values**, it is a **regression problem** (house prices).
- It is commonly used to make predictions or estimations, such as for **sales revenue or salaries** for a given business.

REGRESSION ALGORITHMS

Common algorithms associated with the type of regression of supervised machine learning.

- Linear regression
- Ridge regression
- Lasso Regression
- Others (classification algorithms also used for regression problems)

COMMON REGRESSION PROBLEMS

Below are the most common Regression problems:

- Predicting the **house price** based on the size of the house, availability of schools in the area, and other essential factors.
- Predicting a company's **sales revenue** based on data such as the previous sales of the company.
- Predicting the **temperature** of any day based on data such as wind speed, humidity, and atmospheric pressure.

CHALLENGES OF SUPERVISED MACHINE LEARNING

Although supervised learning can benefit businesses by improving automation and providing deep data insights, there are significant difficulties in creating supervised learning models

- Supervised learning model training can be very time-consuming.
- Datasets may have a higher chance of human error, which could cause algorithms to learn incorrectly.

Unsupervised Learning

UNSUPERVISED LEARNING

Unsupervised learning is a type of machine learning where the **algorithm** is **trained** on a **dataset without any labelled output or supervision.**

In unsupervised learning, the algorithm tries to **find patterns, structures, or relationships in the data by itself.**

The goal of unsupervised learning is to extract useful information or insights from the data, such as **clusters or groups of similar data points, or low-dimensional representations of the data.**

TYPES OF UNSUPERVISED LEARNING

There are two main types of unsupervised learning:

Clustering:

Clustering algorithms group similar data points together based on some similarity metric.

The goal is to **divide the data into distinct groups, or clusters, such that data points within a cluster are more similar to each other than to data points in other clusters.**

Popular clustering algorithms include **k-means clustering, hierarchical clustering, and density-based clustering.**

Dimensionality reduction algorithms aim to reduce the number of features or variables in the dataset while retaining the most important information.

This is often achieved by **projecting** the **data onto a lower-dimensional space or manifold.**

Popular dimensionality reduction techniques include **Principal Component Analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and Autoencoders.**

Association rule learning is a technique used to discover interesting relations between variables in large databases.

The goal is to identify relationships between items that tend to occur together, such as **frequently purchased products** in a supermarket.

Popular algorithms include **Apriori and FP-growth.**

GENERATIVE MODELS

Generative models are used to model the **probability distribution of the data**, allowing for the generation of new data samples that are similar to the original data.

Popular generative models include **Variational Autoencoders, Generative Adversarial Networks (GANs), and Boltzmann machines.**

ANOMALY DETECTION

Anomaly Detection algorithms are used to identify unusual or unexpected data points in a dataset.

This can be useful for detecting fraud, detecting faults in manufacturing processes, or identifying outliers in scientific data.

Common algorithms include a **k-nearest neighbour, isolation forest, and one-class SVM.**

UNSUPERVISED LEARNING APPLICATIONS

Unsupervised learning has many applications in various fields, such as **image and speech recognition, anomaly detection, recommendation systems, and natural language processing.**

However, unsupervised learning can be more challenging than supervised learning because there is no clear objective or metric to optimize, and the resulting output can be highly dependent on the choice of algorithm and parameters.

Therefore, it is important to carefully evaluate and interpret the results of unsupervised learning algorithms.

UNSUPERVISED LEARNING APPLICATIONS CON'T

Unsupervised learning can be used alone or in combination with supervised learning techniques to improve the accuracy and interpretability of models.

For example, *unsupervised pre-training* can be used to *initialize the weights of a deep neural network*, which can then be *fine-tuned using supervised learning* on labelled data.

Unsupervised learning can also be used for feature extraction or selection, which can improve the efficiency and effectiveness of supervised learning algorithms.

In general, unsupervised learning is an important and powerful technique for extracting insights from large and complex data sets and has many practical applications in various domains.

Applications of Unsupervised Learning

APPLICATIONS OF UNSUPERVISED LEARNING

Unsupervised learning has many applications in various fields, including:

- **Clustering:** Clustering algorithms can be used for *customer segmentation in marketing*, *identifying groups of patients with similar health characteristics in healthcare*, and *grouping together similar news articles for recommendation systems in news media.*
- **Dimensionality Reduction:** Dimensionality reduction techniques can be used for *visualizing high-dimensional data*, *compressing data to reduce storage requirements*, and improving the performance of machine learning algorithms by *reducing the number of input features.*

- **Association Rule Learning:** Association rule learning can be used for the *analysis of the market basket in retail*, *identifying common patterns in customer buying behavior and recommending related products.*
- **Generative Models:** Generative models can be used for *data augmentation in image and speech recognition*, *generating new samples for art and music creation*, and *generating synthetic data for training machine learning models.*

- **Anomaly Detection:** Anomaly detection algorithms can be used for fraud detection in finance, fault detection in manufacturing processes, and detecting unusual patterns in scientific data.
- **Natural Language Processing:** Unsupervised learning algorithms can be used for topic modelling in text analysis, identifying hidden patterns and relationships between words and phrases in large corpora of text.

Challenges of Machine Learning

CHALLENGES OF MACHINE LEARNING

Data Quality: Machine learning algorithms require large amounts of **high-quality data** in order to be effective. If the data is **noisy**, **incomplete**, or **biased**, the algorithm may not be able to identify the patterns and relationships in the data.

Overfitting: Overfitting occurs when the machine learning algorithm becomes too complex and begins to fit the training data too closely.

This can lead to poor performance on new, unseen data.

Interpretability: Machine learning algorithms can be difficult to interpret, which can be a problem in fields such as healthcare and finance where decisions must be explainable.

Algorithm Selection: Choosing the **right machine learning algorithm** for a given task can be challenging, as different algorithms have different strengths and weaknesses.

JOB PLACEMENT (MACHINE LEARNING BACKGROUND)

Data Scientist: Data scientists use machine learning algorithms to analyze large data sets and extract insights that can be used to inform business decisions.

Machine Learning Engineer: Machine Learning Engineers design and implement machine learning algorithms, working closely with data scientists to develop and deploy models.

Software Engineer: Software engineers with machine learning experience can work on developing and optimizing machine learning algorithms for use in a wide range of applications.

Research Scientist: Research scientists with a background in machine learning can work in academia or industry to develop new algorithms and techniques and advance the field of machine learning as a whole.

AI Ethicist: As machine learning becomes more prevalent in society, there is a growing need for professionals who can assess the ethical implications of AI and ensure that algorithms are being developed and deployed in a responsible and ethical manner.

Consultant: Machine learning consultants work with businesses and organizations to identify opportunities for the use of machine learning and develop and deploy customized machine learning solutions to meet their specific needs.

Further Readings/Assignments:

1.
2. **Assignment:** Write a short essay on the potential impact of machine learning in healthcare.

END OF PRESENTATION

THANK YOU