Study Of Machine Learning

Kaif Sarfaraz Pathan[1], Riya Sujit Pawar[2], Shravya Shreeyash Pawar[3]
[1, 2] Dept of Computer Science
[3]Asst. Professor, Dept of Computer Science
[1, 2, 3] DUBSS College Dapoli.

Abstract- *This paper represents a novel design and control architecture of the continuous stirred tank reactor (CSTR) based on its mathematical equivalent modeling of the physical system. The plant is formed analytically for the normal operating condition of CSTR. Then the transfer function model is obtained from the process. The analysis is made for the given process for the design of controller with Convectional PID (trial and error method), Ziegler Nichols method, Fuzzy logic method and Model Reference Adaptive method. The simulation is done using MATLAB software and the output of above four different methods was compared so that the Model Reference Adaptive Controller has given better result. This thesis also compares the various time domain specifications of different controllers.*

Keywords- Machine Learning, Supervised Learning, Unsupervised Learning, Regression, Classification, Overfitting, Underfitting, Model Evaluation, Artificial Intelligence.

I. INTRODUCTION

Machine learning focuses on creating a system that is capable of learning and improving from experience without being straightforward planned. It involves using Algorithms to allows computers to analyse and interpret data, identify patterns and make decisions. The key is, Machine learning copies human learning. Just as people learn by observing their neighbourhood or by gathering information and making decisions, Machine learning system depends on data. The more data has a system the better it becomes at solving problems and predicting outcomes. Machine learning is widely used in our daily lives. There are so many examples that are present in today like Alexa, Siri also Netflix, Amazonself-driving cars also part of Machine Learning. It also plays a key role in healthcare for finding the diseases so machine learning isthat who transform everyday and changes the way of life and work. It makes system smarter and more well- organized. The technology is advancing everyday and it is able to solve world's biggest problem it is important to use it responsibility.

II. HISTORY

A Brief History of Machine Learning

1943: The first idea for neural networks, which are the foundation of machine learning, was introduced by Walter Pitts and Warren McCulloch.
1950s: Alan Turing proposed the Turing Test, a way to measure if a computer can think like a human.
1952: Arthur Samuel created the first learning program for playing checkers, where the program got better at the game the more it played.
1957: Frank Rosenblatt developed the first neural network called the perceptron; a model inspired by how human brains work.
1960s-1970s: Early work was done to help machines recognize patterns, like figuring out the shortest route for a salesman visiting different cities.
Major Advances in Machine Learning
1997: IBM’s Deep Blue famously beat the world chess champion.
2006: The term deep learning was coined, referring to advanced algorithms that allow machines to recognize objects and text in images.
2011: Google’s Google Brain used deep learning to teach a computer to recognize cats in videos.
2014: Facebook launched Deep Face, which could recognize people in photos just like humans can.
Machine Learning Today
2015 and Beyond: Big companies like Amazon and Microsoft developed tools to make machine learning easier to use. Researchers also warned about the dangers of autonomous weapons.
2016: AlphaGo, an AI by Google DeepMind, defeated a world champion at the complex game Go.
2017: Waymo (a part of Google) started testing self-driving cars.
2020: GPT-3, a breakthrough AI by OpenAI, was released. It can generate human-like text and is one of the largest language models.

The Future of Machine Learning

Unsupervised Learning: More efforts will be made to improve algorithms that can learn from unlabelleddata. This helps AI find hidden patterns and trends in data, useful for businesses to understand customers or markets better.

III. TYPES OF MACHINE LEARNING

1. Supervised Machine Learning: -

Supervised learning is defined as when a model gets trained on a "Labelled Dataset". Labelled datasets have both input and output parameters. In Supervised Learning algorithms learn to map points between inputs and correct outputs. It has both training and validation datasets labelled.

Example: Consider a scenario where you have to build an image classifier to differentiate between cats and dogs. If you feed the datasets of dogs and cats labelled images to the algorithm, the machine will learn to classify between a dog or a cat from these labelled images. When we input new dog or cat images that it has never seen before, it will use the learned algorithms and predict whether it is a dog or a cat. This is how supervised learning works, and this is particularly an image classification.

There are two main categories of supervised learning that are mentioned below:

- Classification
- Regression

Classification and Regression in Supervised Learning

Classification:

Classification involves predicting categorical target variables, which represent discrete classes or labels.

Examples:

-Classifying emails as spam or not spam
- Predicting whether a patient has a high risk of heart disease

Some common classification algorithms include:

- Logistic Regression
- Support Vector Machine (SVM)
- Random Forest
- Decision Tree
- K-Nearest Neighbours (KNN)

Regression:

Regression involves predicting continuous target variables, which represent numerical values. Examples:

- Predicting the price of a house based on its size, location, and amenities
- Forecasting the sales of a product

Some common regression algorithms:

- Linear Regression
- Polynomial Regression
- Ridge Regression
- Lasso Regression
- Decision Tree

Machine Learning:

Machine learning involves teaching computers to make decisions or predictions by learning from data. The workflow consists of three main steps:

1) Gathering Data: Collecting relevant and accurate information for the model to learn.
2) Data Pre-processing: Cleaning and organizing raw data to make it usable for the model.
3) Data Pre-processing:
- Handling missing values
- Detecting outliers
- Converting categorical and ordinal values into numbers.

1. Gathering Data:

What it is: Collecting the information the model will use to learn. The type of data depends on your project. For example, if you're working on a real-time project, you might use sensors to gather live data.

Sources: Data can come from files, databases, sensors, or online repositories. Websites like Kaggle and the UCI Machine Learning Repository offer free datasets.

Challenges: The raw data you collect often isn't ready for analysis. It may contain missing values, errors, or be unstructured. This is why we need data preparation.

2. Data Pre-processing:

What it is: Cleaning and organizing raw data to make it usable for the model.

Why it's important: Most real-world data is messy. Issues include:
Missing data: Gaps in the dataset due to errors or incomplete records.

Noisy data: Extreme or unusual values caused by mistakes.

Inconsistent data: Errors like duplicated or mismatched entries.

Types of data:

1. Numeric (e.g., age, income)
2. Categorical (e.g., gender, nationality)
3. Ordinal (e.g., low/medium/high)

Steps to clean data:

1. Convert data: Turn categorical and ordinal values into numbers so the model can understand them.
2. Handle missing values: Either remove rows/columns with missing data or fill them in using techniques like averages.
3. Detect outliers: Identify and fix unusually high or low values (e.g., a recorded weight of 800 kg due to a typo).

3. Choosing the Best Model:

Supervised learning is a type of machine learning where the model learns from labelled data. There are two main types of supervised learning:

1) Classification: This is where the model sorts data into groups or categories. For example, classifying emails as spam or not spam.

2)Regression: This is where the model predicts continuous values. For example, predicting house prices.

Some common algorithms used for classification are:

- K-Nearest Neighbours (KNN)
- Naive Bayes
- Decision Trees
- Support Vector Machine (SVM)

Some common algorithms used for regression are:

- Linear Regression
- Support Vector Regression (SVR)
- Decision Trees
- Gaussian Process Regression

Training and Testing a Model

To train a model, you need to split your data into three parts:

1. Training Data: This is the data that the model learns from.
2. Validation Data: This is the data that the model uses to fine-tune its performance during training.
3. Testing Data: This is the data that the model uses to evaluate its final performance.

Model Evaluation

To evaluate a model's performance, you can use metrics such as:

- Accuracy: This measures the number of correct predictions made by the model.
- Confusion Matrix: This shows the number of true positives, true negatives, false positives, and false negatives.
- Precision, Recall, and F1-Score: These metrics are useful for evaluating models on imbalanced datasets.

IV. OVERFITTING AND UNDERFITTING

Overfitting and Underfitting: -

What are they?

Underfitting: This happens when a model is too simple and can't capture the important patterns in the data. It's like trying to draw a picture with only a few lines.

Overfitting: This happens when a model is too complex and learns unnecessary details and noise in the data. It's like trying to draw a picture with too many lines, so it becomes messy.

Causes

Underfitting:

1) Using a model that's too simple
2) Not having enough data
1) Using too much regularization (which limits the model's ability to learn)

Overfitting:

1) Using a model that's too complex
2) Having too small or non-diverse training data

3) Not using regularization techniques (which help prevent overfitting)

Effects

Underfitting: The model performs poorly on both the training data and the testing data.

Overfitting: The model performs well on the training data, but poorly on the testing data.

Solutions:

Underfitting:

1) Use a more complex model
2) Add more relevant features to the data
3) Reduce regularization

Overfitting:

1) Use a simpler model
2) Add more training data
3) Use regularization techniques, such as dropout or L1/L2 regularization

Goal

The goal is to find a balance between underfitting and overfitting. We want a model that's complex enough to learn important patterns, but not so complex that it learns unnecessary details.

V. CONCLUSION

Machine learning has transformed the way computers learn and improve. This paper covered its history, types, and applications, including supervised and unsupervised learning. We discussed common challenges like overfitting and underfitting, and strategies to address them. By understanding these concepts, practitioners can develop effective machine learning models.

As machine learning advances, it's essential to use it responsibly and ethically. The future holds promise, with ongoing research in unsupervised learning, explainability, and transfer learning. By addressing challenges and limitations, we can ensure machine learning benefits society as a whole.

REFERENCES

[1] Types of machine learning https://www.lightsondata.com/the-history-of-machine-learning/#:~:text=Machine%20learning%20history%20starts%20in,by%20Donald%20Hebb%20is%20published.
[2] History https://www.geeksforgeeks.org/types-of-machine-learning/