Ishita Srivastava, 190103027
Final B.Tech Chemical Engineering.

Date 26/01/2023

Assignment-1

Saathi

Unit-1: Introduction to Machine Learning

Machine Learning: Machine Learning is a subfield of artificial intelligence, which is broadly defined as 'the capability of a machine to imitate intelligent human behavior'. It focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, and to uncover key insights in data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. Machine learning algorithms are typically created using frameworks that accelerate solution development.

Why Machine Learning?

Machine learning is concerned with using the right features to build the right models that achieve the right tasks.

Through machine learning, models can be produced quickly and automatically to analyze bigger, more complex data and deliver faster, more accurate results, even on a very large scale.

Following are the benefits of machine learning:

Faster decision making: It allows the user to process and analyze data more quickly, enabling rapid - even split-second - decision making.

Saathi

Date __/__/____

Accurate forecasting: Incorporation of machine learning models into the data analytics can help individuals or businesses to gain far more accurate and powerful capabilities to forecast results, such as product demands.

Boosting efficiency: The use of machine learning allows to accelerate repetitive tasks and reduces human effort.

Applications of Machine Learning

The applications and uses of machine learning are vast and diverse. Some of them are -

Recommendations: The recommendations provided by popular streaming platforms like spotify and Netflix, or by social media platforms such as facebook and instagram, are based on machine learning algorithms. They analyze the content you listen, watch or follow, and suggest additional content you may enjoy.

Fraud detection: Using machine learning models, banks and other financial institutions can identify transactions that fall outside typical parameters - such as purchase amount and user location - and alert you when unusual activity occurs.

Search engine results: Every time you type a

search term into Google, machine learning algorithms analyze your behavior to refine the future delivery of results.

Chatbots: When you chat with an AI-based assistant to resolve an issue online, a trained machine learning model is at work, providing automated appropriate response based on your input.

Spam filters: By analyzing characteristics in subject lines, body content and return addresses, machine learning algorithms use neural networks to help protect your inbox from unwanted emails.

Customer retention: Service providers rely on machine learning models to identify customers who may be ready to take their businesses elsewhere. If you've stopped using a credit card and suddenly received an email offer for a lower APR, your credit card provider is likely attempting to boost customer retention with the help of a machine learning-based platform.

Sentiment analysis: Also called opinion mining or emotion AI, sentiment analysis uses natural language processing and machine learning to understand the underlying sentiment in social media posts. Businesses can use this analysis to discover how people feel about their brand or product.

Date __/__/____

Saathi

Real estate valuation: By analyzing available data on a home's features and the sales of comparable houses in its vicinity, machine learning algorithms estimate the current value of real estate for websites like Zillow and Redfin.

Learning apps: Educational tools like the Duolingo language platform use machine learning models to analyze data gathered from users and adjust the pacing of courses as needed.

Medical image processing: For health care companies, machine learning radiology platforms can be trained to identify potential issues in patient x-rays, flagging them as warranting further attention.

Structure of learning

Following steps may be involved in the structure of machine learning -

1. Data collection - The quantity and quality of data collected will directly determine how accurate the model is. The outcome of this step is generally a representation of data which will be used for training.

2. Data preprocessing - It is used to wrangle data and prepare it for training. Cleaning of data such as removal of duplicates, correction of errors,

dealing with missing values, normalization, data type conversion, etc are done at this stage. Data is visualized to help detect relevant relationships between variables. Data is split training and evaluation sets.

3. Selection of model - Different algorithms are for different tasks, the most appropriate one is chosen on the basis of problem domain.

4. Training of model - The goal of training is to answer a question or make a prediction correctly as often as possible. Each iteration of process is a training step.

5. Model testing - This step uses some metric or combination of metrics to measure objective performance of model. The model is tested against previously unseen data.

6. Model deployment/evaluation - Model performance in the real world is evaluated, and model parameters are tuned to improve performance.

Machine Learning Tasks

A machine learning task is the type of prediction or inference being made, based on the problem or question that is being asked, and the available data. These tasks rely on patterns in the data.

On the basis of application, there are following four types of machine learning tasks -

Saathi

Date __/__/____

→ Recommendation systems
→ Classification tasks
→ Predictive tasks
→ Descriptive tasks

Recommendation systems - It uses big data to recommend additional products to consumers which are relevant on the basis of past purchases, helps websites to improve user engagement, sending personalized content, etc.

Classification tasks - Here, model is used to categorize the input data onto different labels by the help of training data. For example: classifying emails into categories (primary, social, promotions and spam).

Predictive tasks - Predictive analytics is used to observe recent data trends and predict the future event or other data or trends. It is based on the proactive approach. For example: weather forecast, trends in share market, etc.

Descriptive tasks - Descriptive tasks are generally about providing correlation, cross-tabulation, frequency, etc. The techniques or models are used to determine the data regularities and to reveal patterns. Descriptive tasks focus on "what has happened in the past?" and provide useful data. For example: helping companies in recruitment, maintaining inventories, etc.

Date __/__/____

Saathi

Difference between predictive and descriptive tasks:

Predictive tasks	Descriptive tasks
→ These include prediction of future event or other data or trends	→ These involve providing correlation, cross-tabulation, frequency, etc.
→ These are based on the proactive approach	→ These are based on the reactive approach.
→ They require execution of induction over the current and past data so that prediction can happen.	→ They require specification of characteristics of data in a target data set.
→ Outcomes produced do not ensure accuracy.	→ Output data is precise.
→ They need statistics and data forecasting procedures	→ These require data aggregation and data mining

Machine learning models

Models are the central concept in machine learning as they learn from data in order to solve a given tasks. There is a large range of machine learning models that are available due to the omnipresence of tasks that machine learning aims to solve. On the basis of structure of data, machine learning models can be categorized as-

Saathi

Date __/__/____

Geometric models - Geometric models use the technique of combining machine learning and computer vision to solve visual tasks. These models define similarity by considering the geometry of the instant space. Geometric models are basically of two types :

→ Linear models: These use geometric concepts such as lines or planes to segment the instant space.

→ Non-linear / Distance-based models: If the distance between two instances is small then instances are similar in terms of their feature values.

There are different linear models like the least-squares method, support vector machine, etc. Various distance-based models include KNN (K-nearest neighbours), K-means, hierarchial clustering, etc.

Probabilistic models - A probability model / method is based on the theory of probability, or the fact that randomness plays a role in predicting future events. These models take into account the impact of random events or actions in predicting the potential occurence of future outcomes. One of the key benifits of probabilistic models is that they give an idea about the uncertainty ~~the~~ linked with predictions. The best known algorithm in this group is the Naive Bayes algorithm.

Page No.

Saathi

Date __/__/____

Logical models - Logical models use a logical expression to divide the instance space into segments and hence construct grouping models. A logical expression is an expression that returns a boolean value, i.e., a True or false outcome. Once the data is grouped using a logical expression, the data is divided into homogeneous groupings for the problems we are trying to solve. There are mainly two kinds of logical models : Tree models and Rule models.

Learning-based categories in Machine Learning

In general, most machine learning techniques can be classified into supervised learning, unsupervised learning, and reinforcement learning.

Supervised Machine Learning - In supervised machine learning, the algorithm is provided an input dataset, and is rewarded or optimized to meet a set of specific outputs. There are two types of supervised learning -

→ Classification : It is a process of finding a function which helps in dividing the dataset into predetermined classes based on different parameters.
eg - email categories. Here, output variable is a discrete value.

→ Regression : Regression is a process of finding the correlations between dependent and independent variables. It helps in predicting the continuous variables, such

as prediction of market trends, house prices, etc. Here, output variable is of continuous nature or real value.

Unsupervised Machine Learning - In unsupervised machine learning, the algorithm is provided an input dataset, but not rewarded or optimised to specific outputs, and instead trained to group objects by common characteristics. For example, recommendation engines on online stores rely on unsupervised machine learning, specifically a technique called clustering.

Reinforcement Learning - In reinforcement learning, the algorithm is made to train itself using many trial and error experiments. Reinforcement learning happens when the algorithm interacts continually with the environment, rather than relying on training data. One of the most popular examples of reinforcement learning is autonomous driving.