Assignment-Introduction to Data Science

Submitted By: Aaditya Sapkota, B.D.Sc Roll No: 37

Solutions:

1.Explain the difference between supervised, unsupervised and reinforcement learning with examples. How do these learning paradigms apply to real-world applications?

= Machine Learning can be defined as the subset of Artificial Intelligence that enables a system to identify patterns and make decisions or predictions based on given data without being explicitly programmed.

Machine Learning can be broadly classified into following three types:

a.Supervised Learning

b.Unsupervised Learning

c.Reinforcement Learning

a.Supervised Learning:

In Supervised Learning, the model is trained on a labeled dataset.It means that every input is associated with its corresponding output i.e input-output pairs.

Example: Suppose we want to predict house price based on different features like location,size,etc.Here,the input shall be the different feature and the output shall be the price of the house.The model once trained on this labeled data can predict even unseen house prices .

Real-World Application:

-Spam Email DetectIion

-Sentiment Analysis

-Image Classification

b.Unsupervised Learning:

In Unsupervised Learning, the model is trained on the unlabeled dataset, and model once trained tried to find the hidden patterns or clusters or structures in the data.

Example: Taking the Customer purchasing behaviour related data, and clustering customers into groups based on their behaviours.

Real World Application:

-Anomaly Detection

-Recommendation System

-Dimensionality Reduction

c.Reinforcement Learning: In Reinforcement Learning , An agent learns to make decisions by interacting with an environment and receives reward and penalty based on the actions taken.

Example: A robot learns to walk by trial and error in a simulated environment.

Real World Application:

-Self Driving Cars

-Game Playing AI

-Robotics

2.Describe the working of an artificial neural network (ANN) with a simple example. How does the backpropagation algorithm help in adjusting the weights of a neural network?

=An artificial Neural Network(ANN) is a machine learning algorithm that identifies patterns from the data by gaining knowledge mimicking humain brain structure. It has three main types of layers:

-Input Layer: It takes raw data as input and passes it to the next layer.

-Hidden Layer:This layer consists of number of layers, which perform computations using weights, biases,and activation function to identify patterns.

-Output Layer:This layer makes predictions or classifications based on the patterns identified by the previous layers.

Example: Suppose we have to predict whether a student will pass or fail based on the study hours and attendance,Here, Study Hours and Attendance are the input data for the input layer, the hidden layer processes these given inputs using weights and biases and the output layer shows the output i.e the probability of passing .

Backpropagation is a key algorithm used to train ANNs by minimizing prediction or classification errors through weight and bias adjustments.

Backpropagation works in several steps . Firstly, there is Forward Pass where the input data flows through several layers and some predictions are generated by the output layer. The next step is the Error Calculation where the difference between the expected and actual value is computed.Now to minimize the error, the next step is the Backward Pass , where using the chain rule new weights are computed and the weights are then updated using gradient descent to minimize errrors and this steps are repeated until the error is minimized.

Backpropagation plays a crucial role for ANNs as it helps in minimizing the model’s prediction error by adjusting weights and thus improving predictions.

3.AI can be classified into narrow AI, general AI and super intelligent AI. Explain each type with examples. What are the ethical and societal implications of AI advancements?

=Artificial Intelligence in simple terms refers to making computer systems and programs able to think ,learn and make decision. Based on their level of capability, AI is classified into three types:

a.Narrow AI

b.General AI

c.SuperIntelligent AI

a.Narrow AI:

These AI systems are designed to perform a specific task or solve a particular problem.It is task specific,cannot generalize knowledge beyond training data and lacks consciousness and awareness.For Example: Virtual Assistants like Siri,Alexa , Google Translate and Facial Recognition Systems.

b.General AI:

These AI systems have the ability to perform any intellectual tasks that human intelligence can do. It can generalize knowledge ,learn new skills and show adaptability without being explicitly programmed and have self awareness and consciousness. For Example: An Advanced Robot that can learn new languages,solve complex problems and engage in creative tasks like humans.

c.SuperIntelligent AI:

These AI systems have intelligence that goes beyond the human intelligence in all aspects including problem solving and decision making.It is purely hypothetical but is assumed to be the ultimate goal of AI systems.For Example: A potential example would be an AI system that can independently discover scientific theories or technologies without human intervention.

The rapid advancements of AI brings both opportunities and challenges for society which may raise various ethical and societal concerns. While Automation is powered by AI systems, this potentially increases the risk of humans losing their jobs.The ML models if trained on biased data can explicitly cause partiality and discrimination in society.With AI system effectively being enforced everywhere, this increase the privacy concerns questioning the user’s data privacy.Furthermore,Economic Inequality between Nations and various ethical issues may also arise.

4.What are the key characteristics of Big Data (Volume, Variety, Velocity, Veracity and Value)? Explain each with real-world examples and discuss the challenges associated with managing big data.

=Big Data refers to extremely large and complex datasets that cannot be processed, stored, or analyzed using traditional data management tools. It is characterized by the 5 Vs:Volume,Variety,Velocity,Veracity and Value:

Key Characteristics of Big Data:

a.Volume: Refers to the size or amount of data generated every second.

Real World Example: Social media platforms like Facebook generate petabytes of data daily from posts, comments, likes, and shares.

Challenges:

-Storage of such large data requires huge infrastructures.

-Processing demands high computational power.

b.Variety: Refers to the diversity of data types and formats.

Real World Example:Structured data in databases, unstructured data like images,videos,social media posts ,etc,

Challenges:

-Integrating different formats into a unified system for analysis.

-Extracting meaningful insights from unstructured data like images or audio.

c.Velocity:Refers to the speed at which data is generated and processed.

Real World Example: Stock Market transactions occur in microseconds, streaming platforms recommend content instantly.

Challenges:

-Real-time processing requires advanced tools.

-Delayed Processing

d.Veracity: Refers to the accuracy, reliability, and quality of the data being collected.

Real World Example: Social media platforms often contain fake news or misinformation which can cause analysis mistakes.

-Challenges: Cleaning and preprocessing data to ensure its reliability for decision-making.

-Identifying and removing outliers or incomplete records.

e.Value:Refers to the ability to extract insights from data that can drive further innovations.

-Real World Example: Healthcare providers analyze patient data to predict diseases and improve treatment outcomes.

Challenges:

-Identifying insights from massive dataset.

-Ensuring that extracted insight align with our business goals.

Furthermore, Some of the challenges associated with managing big data in addition to the ones listed above include: Data Storage and Scalability,Data Processing,Data Security and Processing,Data Integration,Data Quality,etc.