Q.1. Define machine learning and describe the key steps in the machine learning process?

→ 1. This will itself address the first question i.e. if machines really learn. There are multiple ways to define machine learning.

'A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experince E.'

2. What this essentially means is that a machine can be considered to learn if it is able to gather experience by doing a certain task and improve its performance in doing the similar tasks in the future.

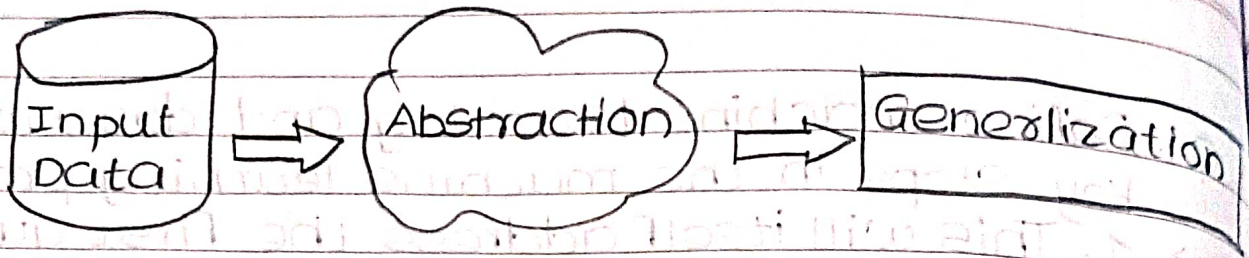
3. When we talk about past experience, it means past data related to the task. This data is an input to the machine from some source.

4. The basic machine learning process can be divided into three parts

A. Data Input

B. Abstraction

C. Generalization



The key steps in the machine learning process are as follows:

1. Problem Definition:
   Clearly define the problem you want to solve using machine learning, such as classification, regression or clustering.

2. Data collection:
   Gather relevant data that will be used to train and test your machine learning model.

3. Data preprocessing
   Clean, transform, and preprocess the data to handle missing values, outliers, and ensure it's in a suitable format for training.

4. Feature Engineering
   Select or create meaningful features from the raw data to help the model learn patterns effectively.

5. Data Splitting:
   Divide the dataset into training, validation, and testing subsets to train, and evalute the model's performance.

6. Model selection
choose an appropriate machine learning algorithm based on the problem and data characteristics.

7. Model training
Train the selected model using the training data, adjusting its parameters to optimize performance.

8. Model Evaluation
Assess the model's performance using the validation dataset, employing metrics like accuracy, precision & recall.

Q.2. Explain supervised machine learning with an example.

→ 1. In supervised learning, the algorithm learns a mapping betⁿ the input and output data.
2. This mapping is learned from a labeled dataset, which consists of pairs of input and output data.
3. The labeled dataset used in supervised learning consists of input features and corresponding output labels.
4. The input features are the attributes or characteristics of the data that are used to make predictions, while the output labels are the desired outcomes that the algorithm

tries to predict.

5. One of the primary advantages of supervised learning is that it allows for the creation of complex models that can make accurate predictions on new data.

7. However, supervised learning requires large amounts of labeled training data to be effective.

8. Supervised learning can be further classified into two categories:

A. Regression:

In Linear Regression, the target variable is a continuous value. The goal of regression is to predict the value of the target variable based on the input variables.

B. Classification

The target variable is a continuous value. The goal of regression is to predict the value of the target variable based on the input variables.

9. Example: House prices

1st we need data about the houses: Square footage, no. of rooms, features whether a house has a garden or not.

We then need to know the prices of these houses i.e. the corresponding labels.

Q13. Explain unsupervised machine learning with example

→ 1. Unsupervised learning is the training of a machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance.

2. Here the task of the machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data.

3. Unsupervised learning is classified into two categories of algorithms.

A. clustering:

A clustering problem is where you want to discover the inherent groupings in the data; such as grouping customers by purchasing behaviour

B. Association

An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

4. Example:

Find a face in an image

Q4. List different types of machine learning and briefly compare supervised learning VS unsupervised learning.

→ Supervised v/s Unsupervised ML

Parameters	Supervised Machine Learning	Unsupervised Machine Learning
1. Input Data	Algorithms are trained using labeled data	Algorithms are used against data that is not labeled
2. Computational Complexity	Simpler method	Computationally complex
3. Accuracy	Highly accurate	Less accurate
4. No. of classes	No. of classes is known	No. of classes is not known
5. Data Analysis	Uses offline analysis	Uses real-time analysis of data
6. Algorithm used	Linear & Logistics, regression, Random Forest, Support vector machine, Neural Network	K-Means clustering, Hierarchical clustering, Apriori Algorithm
7. Training data	Use training data to infer model	No training data is used
8. complex model	It is not possible to learn larger and more complex models than with supervised learning	It is possible to learn larger and more complex models with unsupervised learning
9. Example	optical character recognition	Find a face in an image

Types of Learning

1. Supervised 2. Unsupervised 3. semi supervised 4. Reinforcement

3. Semi Supervised:
It is a type of ML algorithm that lies betn supervised and unsupervised ML

4. Reinforcement
It works on a feedback-based process, in which an AI agent automatically explore its surrounding by hitting & trail, taking action, learning from experience and improving its performance.

Q.6. Write a note on
1. Data preprocessing in machine learning

→ 1. Data preprocessing is an important step in the data mining process.
2. It refers to the cleaning, transforming, and integrating of data in order to make it ready for analysis.
3. The goal of data preprocessing is to improving the quality of the data and to make it more suitable for the specific data mining tas
4. Some common steps in data preprocessing data:

a. Data cleaning
This involves identifying and correcting errors or inconsistencies in the data, such as missing values, outliers, and duplicates.

b. Data Integration
1. This involves combining data from multiple sources to create a unified dataset.
2. Data integration can be challenging as it requires handling data with different formats, structures and semantics.

C. Data Transformation
1. This involves converting the data into a suitable format for analysis.
2. common techniques used in data transformat -n include normalization, standardization and discretization.

d. Data Reduction

1. This involves reducing the size of the dataset while preserving the important information.
2. Data reduction can be acheived through techniques such as feature selection and feature extraction.

e. Data Discretization

1. This involves dividing continuous data into discrete categories or intervals.
2. Discretization is often used in data mining and machine learning algorithms that require categorical data.

2. Data visualization in ML

1. Data visualization is a crucial aspect of machine learning that enables analysts to understand and make sense of data patterns, relationships and trends.
2. Through data visualization, insights and communicated to a wider audience, making it a critical component of machine learning.
3. Data visualization helps machine learning analysts to better understand and analyze complex data sets by presenting them in an easily understandable format.
4. Data visualization is an essential step in data preparation and analysis as it helps to identify outliers, trends and patterns in the data that may be missed by other forms of

analysis.

5. With the increasing availability of big data, it has become more important than ever to use data visualization techniques to explore and understand the data.
6. Machine Learning algorithms work best when they have high-quality and clean data, and data visualization can help to identify and remove any inconsistencies in the data.

3. Real-world applications of machine Learning

→ 1. Banking and finance

In the banking industry, fraudulent transactions, especially the ones related to credit cards, are extremely high, high performance machine learning solutions are implemented by almost all leading banks across the globe.

The models work on a real-time basis, i.e. the fraudulent transactions are spotted and prevented right at the time of occurance.

2. Insurance

Insurance industry is extremely data intensive. For that reason, machine learning is extensively used in the insurance industry.

Two major areas in the insurance industry where machine learning is used are risk prediction during new coustomer onboarding and claims management.