**Q.1) Explain the following with an example -**

**A) Artificial Intelligence**

**B) Machine Learning**

**C) Deep Learning**

**ANS -**

A) Artificial Intelligence:

Artificial Intelligence (AI) refers to the development of computer systems that can perform tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, and natural language processing. AI systems can be designed to learn and adapt from experience, allowing them to improve their performance over time.

Example: An example of AI is a virtual assistant like Siri or Alexa.

B) Machine Learning:

Machine Learning (ML) is a subfield of AI that involves developing algorithms that can learn patterns and insights from data without being explicitly programmed. ML algorithms can identify patterns and make predictions based on large datasets, and can continue to improve their accuracy with more data and feedback.

Example: An example of ML is a spam filter for email.

C) Deep Learning:

Deep Learning (DL) is a type of ML that uses neural networks with many layers to learn complex representations of data. DL algorithms can identify patterns and features in large datasets, and are often used in applications such as computer vision and speech recognition.

Example: An example of DL is a self-driving car.

**Q.2) What is supervised learning? List some examples of supervised learning.**

**ANS -**

Supervised learning is a type of machine learning algorithm where the model learns to map inputs to outputs based on labeled training data. In supervised learning, the algorithm is given input data along with the correct output or target variable, and the algorithm learns to make predictions based on that input.

Examples of supervised learning include:

1. Image classification

2. Spam filtering

3. Regression analysis

4. Natural Language Processing

5. Credit risk assessment

**Q.3) What is unsupervised learning? List some examples of unsupervised learning.**

**ANS –**

Unsupervised learning is a type of machine learning where the algorithm learns to identify patterns and relationships in unlabeled data without the need for explicit target labels. Unlike supervised learning, unsupervised learning algorithms are given only the input data and no corresponding output data, and the algorithm must find patterns and structure in the data on its own.

Examples of unsupervised learning include:

1. Clustering

2. Anomaly detection

3. Dimensionality reduction

4. Generative models

5. Association rule mining

**Q4- What is the difference between AI, ML, DL, and DS?**

**ANS -**

AI (Artificial Intelligence), ML (Machine Learning), DL (Deep Learning), and DS (Data Science) are related fields, but they differ in their focus and scope.

AI is a broad field that refers to the development of computer systems that can perform tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, and natural language processing.

ML is a subfield of AI that involves developing algorithms that can learn patterns and insights from data without being explicitly programmed. ML algorithms can identify patterns and make predictions based on large datasets, and can continue to improve their accuracy with more data and feedback.

DL is a type of ML that uses neural networks with many layers to learn complex representations of data. DL algorithms can identify patterns and features in large datasets and are often used in applications such as computer vision and speech recognition.

DS is a broader field that encompasses the entire data lifecycle, from data collection and cleaning to analysis and visualization. It includes techniques from statistics, machine learning, and computer science to extract insights and knowledge from data.

**Q5- What are the main differences between supervised, unsupervised, and semi-supervised learning?**

**ANS –**

The main differences between supervised, unsupervised, and semi-supervised learning are:

1. Supervised learning involves using labeled data to train a model to predict a target variable. The algorithm is given input data along with the correct output or target variable, and the algorithm learns to make predictions based on that input. Unsupervised learning involves finding patterns and structure in unlabeled data, and the algorithm learns to identify relationships and similarities among the input data without the need for explicit target labels.

2. In supervised learning, the algorithm is given both input and output data, while in unsupervised learning, the algorithm is given only the input data. Semi-supervised learning is a combination of both, where the algorithm is given some labeled data along with a larger set of unlabeled data. The algorithm can use the labeled data to guide its learning on the unlabeled data.

3. Supervised learning is generally used for classification and regression problems, where the goal is to predict a discrete or continuous target variable, while unsupervised learning is used for clustering, dimensionality reduction, and anomaly detection, where the goal is to find structure in the data. Semi-supervised learning can be used in cases where labeled data is expensive or time-consuming to obtain.

4. Supervised learning algorithms typically require larger amounts of labeled data to achieve high accuracy, while unsupervised learning algorithms can be effective with smaller amounts of unlabeled data. Semi-supervised learning can be effective in cases where only a small amount of labeled data is available.

**Q6- What is train, test, and validation split? Explain the importance of each term.**

**ANS –**

Train, test, and validation split is a common technique used in machine learning to divide a dataset into three subsets: a training set, a validation set, and a testing set. The purpose of this technique is to evaluate the performance of a machine learning model and to prevent overfitting.

The training set is used to train the machine learning model, which involves adjusting the model's parameters using a set of input data and known output data. The model's goal is to minimize the error between the predicted output and the actual output. By using a training set, the model can learn to recognize patterns in the data and improve its accuracy over time.

The validation set is used to evaluate the model's performance during the training process. After each training iteration, the model is tested against the validation set to see how well it performs on new, unseen data. The validation set is crucial in preventing overfitting, which occurs when the model becomes too complex and begins to memorize the training data rather than generalizing to new data.

The testing set is used to evaluate the final performance of the model. Once the training process is complete, the model is tested against the testing set to measure its accuracy on new, unseen data. The testing set is essential in assessing the model's ability to generalize to new data and is a critical step in determining the model's overall effectiveness.

**Q7- How can unsupervised learning be used in anomaly detection?**

**ANS –**

Unsupervised learning can be used in anomaly detection by identifying data points that deviate significantly from the norm. Anomaly detection is the process of identifying rare events, observations, or data points that are significantly different from the rest of the dataset. In unsupervised learning, the model is trained on a dataset without any labeled output, meaning the model does not have any prior knowledge of what the "normal" or "anomalous" data points are.

Clustering algorithms, such as K-means, can be used to group similar data points together. Once the data points have been clustered, anomalies can be identified as those that do not belong to any of the clusters or belong to a cluster with very few data points. Density-based algorithms, such as DBSCAN, can also be used to identify anomalies by detecting regions of the data space with low-density regions where anomalous points are located.

Another approach to unsupervised anomaly detection is to use autoencoders, which are neural networks that are trained to learn a low-dimensional representation of the data. The autoencoder learns to encode the data into a smaller set of latent features, and then reconstructs the original data from these features. Anomalies can be identified as data points that are difficult to reconstruct from the learned latent features.

**Q8- List down some commonly used supervised learning algorithms and unsupervised learning algorithms.**

**ANS –**

Supervised Learning Algorithms:

1. Linear Regression

2. Logistic Regression

3. Decision Trees

4. Random Forest

5. Support Vector Machines (SVM)

Unsupervised Learning Algorithms:

1. K-Means Clustering

2. Hierarchical Clustering

3. Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

4. Autoencoders

5. Generative Adversarial Networks (GANs)