1. What does one mean by the term "machine learning"?

Machine learning is a subset of artificial intelligence that involves the development of algorithms and statistical models that enable computers to perform tasks without explicit programming. It focuses on the development of techniques that allow systems to learn and improve from experience.

2.Can you think of 4 distinct types of issues where it shines?

Image Recognition: Machine learning excels in recognizing patterns and objects in images, making it suitable for applications like facial recognition and object detection.

Natural Language Processing (NLP): ML is widely used in language-related tasks such as sentiment analysis, chatbots, and language translation.

Recommendation Systems: ML algorithms analyze user preferences and behaviors to provide personalized recommendations, as seen in platforms like Netflix and Amazon.

Medical Diagnosis: ML can assist in diagnosing diseases and predicting patient outcomes based on medical data.

3.What is a labelled training set, and how does it work?

A labelled training set is a dataset used in supervised learning, where each example is paired with its corresponding label or output. It consists of input-output pairs, allowing the algorithm to learn the mapping from inputs to outputs. The algorithm learns by adjusting its parameters to minimize the difference between predicted and actual outputs.

4.What are the two most important tasks that are supervised?

Classification: Assigning input data to predefined categories (e.g., spam or not spam).

Regression: Predicting a continuous numerical output based on input features (e.g., predicting house prices).

5.Can you think of four examples of unsupervised tasks?

Clustering: Grouping similar data points together without predefined categories.

Dimensionality Reduction: Reducing the number of input features while preserving important information.

Association Rule Learning: Discovering interesting relationships between variables in large datasets.

Anomaly Detection: Identifying unusual patterns or outliers in data.

6.State the machine learning model that would be best to make a robot walk through various unfamiliar terrains?

Reinforcement learning models, particularly Deep Q-Learning or Policy Gradient methods, are suitable for training a robot to navigate and walk through unfamiliar terrains.

7.Which algorithm will you use to divide your customers into different groups?

Clustering algorithms, such as K-means or hierarchical clustering, can be used to divide customers into different groups based on their similar characteristics or behaviors.

8.Will you consider the problem of spam detection to be a supervised or unsupervised learning problem?

Spam detection is typically treated as a supervised learning problem. The algorithm learns from labelled examples of spam and non-spam emails to make predictions on new, unseen emails.

9.What is the concept of an online learning system?

An online learning system learns continuously from incoming data, adapting its model over time. It updates its parameters using each new data point, allowing it to stay current and adapt to changing patterns.

10.What is out-of-core learning, and how does it differ from core learning?

Out-of-core learning is a method used when the dataset is too large to fit into the computer's memory. It involves processing data in smaller chunks (batches) that can fit into memory, allowing the model to learn iteratively.

11.What kind of learning algorithm makes predictions using a similarity measure?

Instance-based learning algorithms, such as k-Nearest Neighbors (k-NN), make predictions based on the similarity measure between instances in the training set and the new data point.

12.What's the difference between a model parameter and a hyperparameter in a learning algorithm?

Model Parameter: Parameters are internal variables learned from the training data, influencing the model's predictions.

Hyperparameter: Hyperparameters are configuration settings external to the model that need to be set before training. They control the learning process.

13.What are the criteria that model-based learning algorithms look for? What is the most popular method they use to achieve success? What method do they use to make predictions?

Model-based learning algorithms aim to minimize the cost function, representing the difference between predicted and actual values. They often use techniques like gradient descent for optimization.

14.Can you name four of the most important Machine Learning challenges?

Data Quality: Inaccurate or incomplete data.

Overfitting/Underfitting: Balancing model complexity.

Interpretability: Understanding and explaining model decisions.

Scalability: Handling large datasets and computational demands.

15.What happens if the model performs well on the training data but fails to generalize the results to new situations? Can you think of three different options?

Overfitting: The model memorizes training data but fails on new data.

Underfitting: The model is too simple and doesn't capture the underlying patterns.

Model Complexity Adjustment: Adjusting hyperparameters to find the right balance.

16.What exactly is a test set, and why would you need one?

A test set is a separate dataset used to evaluate the model's performance after training. It helps assess how well the model generalizes to new, unseen data.

17.What is a validation set's purpose?

A validation set is used during the training process to tune hyperparameters and assess the model's performance on data not seen during training.

18.What precisely is the train-dev kit, when will you need it, how do you put it to use?

A train-dev set is a subset of the training data used to detect problems like data mismatch or model overfitting during the development phase.

19.What could go wrong if you use the test set to tune hyperparameters?

Using the test set to tune hyperparameters can lead to overfitting the model to the test set, compromising its ability to generalize to new data. It may result in optimistic performance estimates.