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

Machine Learning: Machine learning is a subset of artificial intelligence that involves the development of algorithms and models that enable computers to learn patterns from data and make predictions or decisions without being explicitly programmed. It's about training systems to improve their performance on a specific task over time, by learning from experience (data).

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

Image Recognition: Machine learning is used to classify and recognize objects in images, such as facial recognition, object detection, and medical image analysis.

Natural Language Processing (NLP): It's used for sentiment analysis, language translation, chatbots, and understanding human language.

Recommendation Systems: Machine learning can suggest products, movies, music, or content based on a user's preferences and behavior.

Autonomous Vehicles: Machine learning plays a vital role in enabling self-driving cars to perceive and navigate the environment.

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

A labeled training set is a dataset used in supervised machine learning. It consists of input data along with corresponding output labels. The labels represent the correct answers or outcomes for the given inputs. The model learns from this dataset by identifying patterns and relationships between the inputs and labels, and then uses this knowledge to make predictions on new, unseen data.

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

Classification: Assigning a label or category to input data points. Example: Spam detection (classifying emails as spam or not).

Regression: Predicting a continuous numerical value based on input features. Example: Predicting house prices based on features like size, location, etc.

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

Clustering: Grouping similar data points together based on some similarity measure, without any predefined labels.

Dimensionality Reduction: Reducing the number of features while retaining the essential information.

Anomaly Detection: Identifying unusual or rare data points that deviate significantly from the norm.

Topic Modeling: Discovering hidden thematic structures within a collection of documents.

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

The best machine learning model to make a robot walk through various unfamiliar terrains could be a Reinforcement Learning algorithm. Specifically, a Deep Reinforcement Learning model, such as Deep Q-Networks (DQN) or Proximal Policy Optimization (PPO), could be used. These algorithms learn how to take actions in an environment to maximize a cumulative reward signal.

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

The algorithm to divide customers into different groups is typically a clustering algorithm, and one commonly used algorithm for this purpose is K-Means clustering.

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

The problem of spam detection is typically approached as a supervised learning problem. The model is trained on a labeled dataset of emails where each email is labeled as spam or not-spam.

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

An online learning system is a machine learning system that can adapt to new data on the fly, updating its model and making predictions in real-time as new data arrives.

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

Out-of-core learning refers to training a machine learning model on a dataset that is too large to fit entirely in memory. The data is processed in chunks or batches, allowing the model to learn from the data efficiently despite memory limitations. Core learning, as mentioned, doesn't imply a common term in machine learning.

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

A kind of learning algorithm that makes predictions using a similarity measure is an instance-based learning algorithm. These algorithms make predictions by comparing the new data point with instances from the training dataset.

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

In a learning algorithm, a model parameter is a setting that the algorithm learns from the training data (weights in a neural network, coefficients in a linear regression). A hyperparameter is a configuration setting that is set before the learning process begins, like the learning rate, number of hidden layers, etc.

**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 look for good generalization to new, unseen data. The most popular method they use is regularization, which adds constraints to the model to prevent overfitting. They use the learned model to make predictions by applying it to new input data.

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

Overfitting: When a model performs well on the training data but fails to generalize to new data.

Data Scarcity: Insufficient data can hinder the model's ability to learn.

Bias and Fairness: Models can inherit biases present in the training data, leading to unfair predictions.

Interpretability: Complex models can be difficult to interpret, making it hard to understand their decision-making.

**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?**

If a model performs well on training data but fails to generalize:

Simplify the Model: Use a simpler model with fewer parameters to reduce overfitting.

Regularization: Apply techniques like L1 or L2 regularization to control model complexity.

Collect More Diverse Data: Gather more diverse and representative data for training.

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

A test set is a separate portion of the dataset that is not used during model training. It's used to evaluate the model's performance on new, unseen data, providing an estimate of how well the model generalizes.

**17.What is a validation set's purpose?**

The purpose of a validation set is to tune hyperparameters and assess model performance during training. It's a subset of the training data used to make decisions about model architecture and hyperparameter settings.

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

The train-dev kit (training-development kit) refers to a portion of the dataset that's used for both model training and hyperparameter tuning. It's particularly useful when dealing with small datasets to ensure a balance between training and evaluation.

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

If you use the test set to tune hyperparameters, you risk overfitting to the test set. This can lead to the model performing well on the test set by chance but failing to generalize to new data. It's essential to use a separate validation set for hyperparameter tuning and keep the test set untouched for final evaluation.