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

Answer 1: Machine learning is a field of artificial intelligence (AI) that focuses on the development of algorithms and models that enable computers to learn and make predictions or decisions without being explicitly programmed. It involves the use of data to train models, allowing them to improve their performance over time.

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

Answer 2: Four distinct types of issues where machine learning shines are:

* Image Recognition: Identifying and classifying objects or patterns in images.
* Natural Language Processing (NLP): Understanding and processing human language.
* Recommendation Systems: Providing personalized suggestions based on user preferences.
* Medical Diagnosis: Analyzing medical data for disease diagnosis and treatment recommendations.

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

Answer 3: A labeled training set is a dataset used to train a supervised learning algorithm. It consists of input-output pairs, where the input data is accompanied by corresponding labels or target values. The algorithm learns to map input features to the correct output by generalizing patterns from the labeled examples.

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

Answer 4: Two Important Supervised Tasks:

* Regression: Predicting a continuous output variable.
* Classification: Assigning input data to predefined categories or classes.

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

Answer 5: Four Examples of Unsupervised Tasks:

* Clustering: Grouping similar data points together.
* Dimensionality Reduction: Reducing the number of features while preserving information.
* Association Rule Learning: Discovering interesting relationships in data.
* Generative Modeling: Creating new instances like the training data.

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

Answer 6: A reinforcement learning model, such as Deep Q-Learning, would be suitable for training a robot to walk through various unfamiliar terrains.

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

Answer 7: A clustering algorithm, such as K-means or hierarchical clustering, could be used to divide customers into different groups based on their characteristics.

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

Answer 8: Spam detection is typically treated as a supervised learning problem, as the algorithm needs labeled examples (spam and non-spam) to learn from.

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

Answer 9: An online learning system updates the model continuously as new data becomes available, adapting to changing patterns and ensuring real-time adjustments.

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

Answer 10: Out-of-core learning involves training models on datasets that are too large to fit into a computer's memory. It differs from in-core learning, where the entire dataset can be loaded into memory.

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

Answer 11: Instance-based learning algorithms, such as k-Nearest Neighbors (k-NN), make predictions based on similarity measures between instances.

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

Answer 12: Model Parameter: Parameters learned from the training data (e.g., weights in a neural network).

Hyperparameter: Settings that are set before the training process (e.g., learning rate or number of hidden layers).

**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?**

Answer 13:

* Criteria: Accuracy, simplicity, and the ability to generalize.
* Method: The most popular method is to use a cost function to measure the difference between predicted and actual outcomes.
* Prediction Method: Models use learned parameters to make predictions.

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

Answer 14: Four Machine Learning Challenges are Overfitting and Underfitting, Data Quality, Computational Complexity and Interpretability.

**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?**

Answer 15: If a model performs well on training data but fails to generalize:

* Collect more diverse data.
* Simplify the model or use regularization techniques.
* Try a different learning algorithm.

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

Answer 16: 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?**

Answer 17: A validation set is used during the training process to tune hyperparameters and avoid overfitting to the training data.

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

Answer 18: A train-dev kit is a subset of the training data used to detect problems like data mismatch. It is employed when there's a suspicion that the training and development sets have different distributions.

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

Answer 19: Using the test set to tune hyperparameters can lead to overfitting to the test set, compromising the model's ability to generalize to new data. It may result in optimistic performance estimates that do not reflect real-world performance.