1. **What does one mean by the term "machine learning"?**
   * Machine learning is a branch of artificial intelligence where systems learn from data and improve their performance over time without explicit programming. It focuses on developing algorithms that can learn from and make predictions or decisions based on data.
2. **Can you think of 4 distinct types of issues where it shines?**
   * Machine learning shines in:
     1. Image and speech recognition
     2. Natural language processing
     3. Recommendation systems
     4. Predictive analytics in finance and marketing
3. **What is a labeled training set, and how does it work?**
   * A labeled training set consists of input data paired with correct output labels. It serves as a basis for the machine learning algorithm to learn patterns and relationships between inputs and outputs. By training on labeled data, the algorithm adjusts its parameters to minimize prediction errors.
4. **What are the two most important tasks that are supervised?**
   * The two most important supervised learning tasks are:
     1. Classification: Predicting a categorical label (e.g., spam or not spam).
     2. Regression: Predicting a continuous-valued output (e.g., predicting house prices).
5. **Can you think of four examples of unsupervised tasks?**
   * Examples of unsupervised learning tasks include:
     1. Clustering: Grouping similar data points together (e.g., customer segmentation).
     2. Dimensionality reduction: Reducing the number of input variables (e.g., principal component analysis).
     3. Anomaly detection: Identifying unusual data points (e.g., fraud detection).
     4. Association rule learning: Discovering interesting relations between variables (e.g., market basket analysis).
6. **State the machine learning model that would be best to make a robot walk through various unfamiliar terrains?**
   * Reinforcement learning models, such as Deep Q-Networks (DQN), would be suitable for training a robot to navigate unfamiliar terrains by rewarding it for making correct movements and penalizing incorrect ones.
7. **Which algorithm will you use to divide your customers into different groups?**
   * For dividing customers into different groups (clustering), algorithms like K-means clustering or hierarchical clustering are commonly used.
8. **Will you consider the problem of spam detection to be a supervised or unsupervised learning problem?**
   * Spam detection is typically considered a supervised learning problem because it involves classifying emails (input) into spam or non-spam categories (output), based on labeled training data.
9. **What is the concept of an online learning system?**
   * An online learning system updates the model continuously as new data becomes available, rather than in batch mode. It adapts to changing conditions and can incorporate new observations dynamically.
10. **What is out-of-core learning, and how does it differ from core learning?**
    * Out-of-core learning refers to techniques used when data cannot fit into a computer's main memory (RAM). It involves processing large datasets that are stored on disk sequentially, rather than all at once in memory (core learning).
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 of new instances to instances in the training data.
12. **What's the difference between a model parameter and a hyperparameter in a learning algorithm?**
    * Model parameters are learned from the training data and define the structure of the model (e.g., weights in neural networks). Hyperparameters are set before training and control the learning process (e.g., learning rate, 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?**
    * Model-based learning algorithms aim to minimize prediction errors on new, unseen data. They typically achieve this by fitting a model to training data (e.g., using regression or classification techniques) and use this model to make predictions based on new inputs.
14. **Can you name four of the most important Machine Learning challenges?**
    * Four important challenges in machine learning include:
      1. Overfitting/underfitting
      2. Data scarcity or quality issues
      3. Interpretability of models
      4. Ethical considerations (bias, fairness, privacy)
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 overfits (performs well on training data but not on new data), options include:
      1. Regularization techniques to penalize complex models.
      2. Using more diverse training data.
      3. Simplifying the model architecture.
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 held out from training and used to evaluate the model's performance on unseen data. It helps assess how well the model generalizes to new observations.
17. **What is a validation set's purpose?**
    * A validation set is used to tune hyperparameters and evaluate different model architectures during the training process. It helps prevent overfitting by providing an independent dataset for model selection.
18. **What precisely is the train-dev kit, when will you need it, how do you put it to use?**
    * The train-dev set (train-dev kit) is used to diagnose data mismatch issues between the training and validation sets. It helps identify problems such as differences in data distributions that affect model performance.
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, which can lead to an overly optimistic evaluation of the model's performance on new data. This defeats the purpose of having an independent test set for unbiased evaluation.