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

ANS Machine learning is a field of artificial intelligence that uses statistical techniques to enable computer systems to improve their performance on a specific task through experience, without being explicitly programmed.

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

ANS Machine learning shines in image recognition, natural language processing, anomaly detection, and recommendation systems.

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

ANS A labeled training set is a dataset in which the desired output is known for each example in the dataset. The machine learning algorithm uses this dataset to learn how to map inputs to outputs by adjusting its parameters to minimize the difference between its predictions and the known output.

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

ANS The two most important tasks that are supervised are classification, where the goal is to predict a categorical output, and regression, where the goal is to predict a continuous output.

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

ANS Four examples of unsupervised tasks are clustering, dimensionality reduction, density estimation, and anomaly detection.

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

ANS A reinforcement learning model would be best to make a robot walk through various unfamiliar terrains as it enables an agent to learn from its environment through trial-and-error.

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

ANS A clustering algorithm would be used to divide customers into different groups.

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

ANS The problem of spam detection is a supervised learning problem as the desired output is known for each example in the dataset.

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

ANS An online learning system is a machine learning system that can update its model based on new data that is fed to it continuously.

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

ANS Out-of-core learning is a technique for training machine learning models on large datasets that do not fit in memory. It differs from core learning in that it processes data in smaller batches, reading and writing data to and from disk.

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

ANS A nearest neighbor algorithm makes predictions using a similarity measure.

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

ANS A model parameter is a configuration variable that is internal to the model and is learned from the training data. A hyperparameter is a configuration variable that is external to the model and is set by the data scientist before training the model.

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?

ANS Model-based learning algorithms look for a model that best fits the training data while also minimizing overfitting. The most popular method they use to achieve success is to optimize a loss function. They use the learned model to make predictions on new data.

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

ANS Four of the most important Machine Learning challenges are overfitting, data preprocessing, bias and fairness, 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?

ANS If the model performs well on the training data but fails to generalize the results to new situations, three different options are to collect more data, simplify the model, or regularize the model.

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

ANS A test set is a dataset that is used to evaluate the performance of a trained machine learning model on new, unseen data. It is important to have a test set to avoid overfitting the model to the training data and to get an estimate of how well the model will perform on new data.

17.What is a validation set's purpose?

ANS A validation set is a dataset that is used to tune the hyperparameters of a machine learning model. It is used to evaluate the model's performance on new data during training and to prevent 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?

ANS The train-dev kit is a dataset that is used to assess the performance of a machine learning model during development. It is used to detect and diagnose problems early in the development process, such as overfitting or underfitting. It can be used by splitting a portion of the training data as a train-dev set and using it to evaluate the model's performance during development.

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

ANS.If the test set is used to tune hyperparameters, the model may overfit the test set, resulting in poor generalization performance on new data. This defeats the purpose of having a test set to evaluate the model's performance on new data. Instead, a separate validation set should be used to tune the hyperparameters, and the test set should only be used to evaluate the final performance of the trained model.