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

Machine learning is a subset of artificial intelligence that involves the use of algorithms and statistical models to enable computers to learn from data and make predictions or decisions without being explicitly programmed.

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

Heath sectors, Image and speech recogonition, spam email detections, predictive analysis, fraud detection and NLP etc

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

In machine learning, a labeled training set is a dataset that has been labeled or tagged with predefined outcomes or target variables. These labels or tags serve as a guide for the machine learning algorithm to learn from and make predictions on new, unseen data.

With supervised learning, on the other hand, humans must tag, label, or annotate the data to their criteria, in order to train the model to reach the desired conclusion (output).

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

Regression and classification.

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

Common unsupervised tasks include clustering, visualization,Association rule mining and Anomaly detection.

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

A reinforcement learning model. Reinforcement learning is a type of machine learning that involves an agent (in this case, the robot) learning to make decisions in an environment to maximize a reward signal (in this case, successfully walking through unfamiliar terrains).

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

Clustering more specifically k-,means clutering.

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

The problem of spam detection is typically considered as a supervised learning problem.

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

machine learning algorithm that is designed to learn and adapt from streaming data in real-time. In contrast to traditional batch learning, where the algorithm is trained on a fixed dataset before deployment, online learning algorithms continuously update their model parameters as new data becomes available.

Some popular examples of online learning algorithms include online linear regression, perceptron, and stochastic gradient descent. Online learning systems have been used in a wide range of applications, including fraud detection, recommendation systems, and anomaly detection.

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

In core learning, the entire dataset is loaded into memory and processed at once. This is possible when the dataset is relatively small and can fit into the memory of the machine used for training. However, when dealing with large datasets, this approach can be impractical or even impossible.

Out-of-core learning algorithms are designed to handle large datasets that cannot be loaded entirely into memory. Instead, they process the data in smaller chunks, one at a time, and update the model's parameters after each batch.

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

instance-based learning algorithm or a lazy learning algorithm.

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

| PARAMETERS | HYPERPARAMETER |
| --- | --- |
| They are required for making predictions | They are required for estimating the model parameters |
| They are estimated by optimization algorithms(Gradient Descent, Adam, Adagrad) | They are estimated by hyperparameter tuning |
| They are not set manually | They are set manually |
| The final parameters found after training will decide how the model will perform on unseen data | The choice of hyperparameters decide how efficient the training is. In gradient descent the learning rate decide how efficient and accurate the optimization process is in estimating the parameters |

A model parameter is a variable that is internal to the model and is learned during the training process. The model parameter values are adjusted by the learning algorithm to fit the training data and optimize the model's performance. Examples of model parameters include the weights and biases of a neural network or the coefficients of a linear regression model.

On the other hand, a hyperparameter is a setting that is external to the model and is specified before the training process begins. Hyperparameters are not learned from the data but are chosen by the user based on prior knowledge or trial and error. Examples of hyperparameters include the learning rate of an optimization algorithm, the number of hidden layers in a neural network, or the regularization strength of a model.

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

A test set in machine learning is a secondary (or tertiary) data set that is used to test a machine learning program after it has been trained on an initial training data set.

a test set is a separate dataset used to evaluate the performance of a trained machine learning model on new, unseen data. It helps to ensure that the model is not overfitting and can make accurate predictions on new data.

17.What is a validation set's purpose?

he validation helps to deal with issues like overfitting, where the program may not be calibrated well to handle future data

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

In machine learning, a train-dev kit (also known as a train-dev-test split) is a technique used to split a dataset into three subsets: a training set, a development set (also called a validation set), and a test set.

The purpose of the train-dev kit is to evaluate the performance of a machine learning model during development and testing stages, as well as to prevent overfitting, which occurs when a model is trained to fit the training data too closely and fails to generalize to new, unseen data.

To put a train-dev kit to use in machine learning, you would typically follow these steps:

1. Split the data

2. Train the model

3. Evaluate the model

4. Test the final model

5. Iterate and improve

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

This will produce an overestimate. Test set should be used only for testing not for parameter tuning.