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

Machine Learning is a field of inquiry devoted to understanding and building methods that learn, that is, methods that leverage data to improve performance on some set of tasks. It is a part of AI.

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

Machine learning algorithms have had good results on problems such as spam detection in email, cancer diagnosis, fraudulent credit card transactions, and automatically driving vehicles.

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

The training set is used to train the algorithm, and then you use the trained model on the test set to predict the response variable values that are already known. The final step is to compare the predicted responses against the actual (observed) responses to see how close they are.

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

The two most common supervised tasks are classification and regression.

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

Four common unsupervised tasks included clustering, visualization, dimensionality reduction, and association rule learning.

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

The best Machine Learning algorithm to allow a robot to walk in unknown terrain is Reinforced Learning, where the robot can learn from the response of the terrain to optimize itself.

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

We will use the k-means clustering algorithm to derive the optimum number of clusters and understand the underlying customer segments based on the data provided.

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

Spam detection is a supervised machine learning problem. This means you must provide your machine learning model with a set of examples of spam and ham messages and let it find the relevant patterns that separate the two different categories

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

A learning system based on formalized teaching but with the help of electronic resources is known as E-learning. While teaching can be based in or out of the classrooms, the use of computers and the Internet forms the major component of E-learning.

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

Out-of-core learning refers to a set of algorithms working with data that cannot fit into the memory of a single computer, but that can easily fit into some data storage such as a local hard disk or web repository.

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

What type of learning algorithm relies on a similarity measure to make predictions? A learning algorithm that relies on a similarity measure to make predictions is the instance-based algorithm.

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

Model parameters are estimated based on the data during model training and model hyperparameters are set manually and are used in processes to help estimate model parameters. Model hyperparameters are often referred to as parameters because they are the parts of machine learning that must be set manually and tuned.

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?

The goal of a model-based algorithm is to be able to generalize to new examples. To do this, model-based algorithms search for optimal values for the model's parameters often called `theta`. This searching, or "learning", is what machine learning is all about. Model-based systems learn by minimizing a cost function that measures how bad the system is at making predictions on new data, plus a penalty for model complexity if the model is regularized. To make a prediction, a new instance's features are fed into a hypothesis function which uses the minimized `theta` found by repeatedly running the cost function.

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

- Poor Quality of Data. ...
- Underfitting of Training Data. ...
- Overfitting of Training Data. ...
- Machine Learning is a Complex Process. ...
- Lack of Training Data. ...
- Slow Implementation. ...
- Imperfections in the Algorithm When Data Grows

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?

It is the situation of overfitting the model.

- The problem of model performance mismatch that may occur when evaluating machine learning algorithms.
- The causes of overfitting, under-representative data samples, and stochastic algorithms.
- Ways to harden your test harness to avoid the problem in the first place.

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

Test set—a subset used to put the trained model to the test

Your goal is to develop a model that generalizes well to new data, assuming your test set fits the two constraints mentioned above. Our test set acts as a stand-in for new information. Let's say that the model learned for the training data is really basic.

17. What is a validation set's purpose?

A validation set is a set of data used to train artificial intelligence (AI) with the goal of finding and optimizing the best model to solve a given problem. Validation sets are also known as dev sets. A supervised AI is trained on a corpus of training data.

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

A validation data set is a data set of examples used to tune the hyperparameters (i.e. the architecture) of a classifier. It is sometimes also called the development set or the "dev set". An example of a hyperparameter for artificial neural networks includes the number of hidden units in each layer.

The goal of the dev-set is to rank the models in terms of their accuracy and helps us decide which model to proceed further with. Using the Dev set we rank all our models in terms of their accuracy and pick the best performing model.

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

If we use training data for hyperparameter tuning, we'll be unable to check whether or not our model is performing well on new data - and this would result in overfitting