

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

Machine learning (ML) is **a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so**. Machine learning algorithms use historical data as input to predict new output values

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

There are four types of machine learning algorithms: **supervised, semi-supervised, unsupervised and reinforcement**.

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 **regression and classification**

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

Common unsupervised tasks include **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?

Reinforcement learning

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

One very common machine learning algorithm that's suitable for customer segmentation problems is the **k-means clustering algorithm**

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

It is a supervised learning

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

online machine learning is **a method of machine learning in which data becomes available in a sequential order and is used to update the best predictor for future data at each step**

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

Out-of-core (or "external memory") learning is **a technique used to learn from data that cannot fit in a computer's main memory (RAM)**.

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

Learning algorithm that relies on a similarity measure to make predictions is **instance-based algorithm**.

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

Model parameters are estimated from data automatically and model hyperparameters are set manually and are used in processes to help estimate model parameters

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 for 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.

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

Four main challenges in Machine Learning include **overfitting the data (using a model too complicated), underfitting the data (using a simple model), lacking in data and nonrepresentative data.**

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?

Underfitting refers to a model that can neither model the training data nor generalize to new data

Options : Increase model complexity. Increase the number of features, performing feature engineering. Remove noise from the data

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.**

It is need for testing model accuracy

17.What is a validation set's purpose?

Validation sets are used **to select and tune the final AI model**. Training sets make up the majority of the total data, averaging 60 percent. In testing, the models are fit to parameters in a process that is known as adjusting weights. The validation set makes up about 20 percent of the bulk of data used.

18.What precisely is the train-dev kit, when will you need it, 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"

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

If you use this data to choose hyperparameters, you actually give the model a chance to "see" the test data and to develop a bias towards this test data. Therefore, **you actually lose the possibility to find out how good your model would actually be on unseen data** (because it has already seen the test data)