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

Machine learning is a branch of [artificial intelligence (AI)](https://www.ibm.com/cloud/learn/what-is-artificial-intelligence) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

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

Image Recognition: Image recognition is one of the most common applications of machine learning. ...

Speech Recognition. ...

Traffic prediction: ...

Product recommendations: ...

Self-driving cars: ...

Email Spam and Malware Filtering: ...

Virtual Personal Assistant: ...

Online Fraud Detection:

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**.

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?

e-commerce websites like Amazon use clustering algorithms to implement a user-specific recommendation system.

Dimensionality reduction is a commonly used unsupervised learning technique where the goal is to reduce the number of random variables under consideration.

Products Segmentation

Customer Segmentation

Similarity Detection

Recommendation Systems

Labelling unlabelled datasets

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 response of the terrain to optimize itself.

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

**K-Means clustering** is an unsupervised machine learning algorithm that divides the given data into the given number of clusters

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

Spam detection is a supervised learning problem because the labels are known (spam or no spam).

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

In computer science, 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**, as opposed to batch learning techniques which generate the best predictor by learning on the entire training data set

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?

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?

In a machine learning model, there are 2 types of parameters:

Model Parameters: These are the parameters in the model that must be determined using the training data set. These are the fitted parameters.

Hyperparameters: These are adjustable parameters that must be tuned in order to obtain a model with optimal performance.

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 algorithm search for the optimal value of parameters in a model that will give the best results for the new instances. We often use a cost function or similar to determine what the parameter value has to be in order to minimize the function. The model makes prediction by using the value of the new instance and the parameters in its function.

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?

If the model performs poorly to new instances, then it has overfit on the training data. To solve this, we can do any of the following three: get more data, implement a simpler model, or eliminate outliers or noise from the existing data set.

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

Test set is a set that you test your model (fit using training data) to see how it performs. Test set is necessary so that you can determine how good (or bad) your model performs.

17.What is a validation set's purpose?

Validation set is a set used to compare between different training models.

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

The goal of dev-set is to rank the models in term of their accuracy and helps us decide which model to proceed further with. Using Dev set we rank all our models in terms of their accuracy and pick the best performing model. i.e. dev set ranks models similar to a search engine like google rank pages and then pick the top model and hence act as a filter to remove bad models.

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

If you tune hyperparameters using the test sets, then it may not perform well on the out-of-sample data because the model is tuned just for that specific set.