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

**Answer : Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.**

**Machine learning (ML) is a type of artificial intelligence ([AI](https://searchenterpriseai.techtarget.com/definition/AI-Artificial-Intelligence)) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning [algorithms](https://whatis.techtarget.com/definition/algorithm) use historical data as input to predict new output values.**

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

**1. Prediction**

**2. Spam and malware detection**

**3. Recommendation**

**4. Fraud Detection**

**5. Image recognition**

**6. voice recognition**

**7. Social Media**

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

**Answer : In dataset we will use some data to train model and some data to test model accuracy. The data use for training model is called labeled training data set.**

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

**Answer :  The two most common supervised learning tasks are regression and classification**

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

**Answer : Four common unsupervised tasks inclused 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?

**Answer : The best Machine Learning algorithm to allow a robot to walk in unknown terrain is Reinforced Learning**

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

**Answer : Clustering algorithm are use for customer segmentation.**

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

**Answer : Supervised Learning.**

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

**Answer : 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?

**Answer : 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?

**Answer : 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?

**Answer :**

**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 the machine learning that must be set manually and tuned.**

**Basically, parameters are the ones that the “model” uses to make predictions etc. For example, the weight coefficients in a linear regression model. Hyperparameters are the ones that help with the learning process. For example, number of clusters in K-Means, shrinkage factor in Ridge Regression. They won’t appear in the final prediction piece, but they have a large influence on how the parameters would look like after the learning step.**

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?

**Answer :**

**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?

**Answer :**

1. **Data Collection**
2. **Less Amount of Training data**
3. **Non-Representative Training data**
4. **Poor Quality of data**
5. **Irrelavant / Unwanted feature**
6. **Overfitting of Training data**
7. **Underfitting of training data**
8. 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?

**Answer :**

**1. Overfitting**

**2. Underfitting.**

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

**Answer : The “training” data set is the general term for the samples used to create the model, while the “test” or “validation” data set is used to qualify performance. Perhaps traditionally the dataset used to evaluate the final model performance is called the “test set”.**

17.What is a validation set's purpose?

**Answer : 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, how do you put it to use?

**Answer :**

**the dev set is a small set. So we try something on the dev set and come to a conclusion and then go to the train set to train it properly and check.**

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

Answer :

**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).**