**1. What does one mean by the term “machine learning”?**

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?**

[Recommendation engines](https://whatis.techtarget.com/definition/recommendation-engine)

fraud detection,

spam filtering,

malware threat detection,

[business process automation](https://searchcio.techtarget.com/definition/business-process-automation) (BPA)

[predictive maintenance](https://whatis.techtarget.com/definition/predictive-maintenance-PdM).

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

In machine learning, data labeling is the process of identifying raw data (images, text files, videos, etc.) and adding one or more meaningful and informative labels to provide context so that a machine learning model can learn from it.

In machine learning, a properly labeled dataset that you use as the objective standard to train and assess a given model is often called “ground truth.” The accuracy of your trained model will depend on the accuracy of your ground truth, so spending the time and resources to ensure highly accurate data labeling is essential.

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

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

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

* Customer segmentation or understanding different customer groups around which to build marketing or other business strategies.
* Genetics, for example clustering DNA patterns to analyze evolutionary biology.
* Recommender systems, which involve grouping together users with similar viewing patterns to recommend similar content.
* [Anomaly detection](https://pages.dataiku.com/anomaly-detection-at-scale-guidebook), including fraud detection or detecting defective mechanical parts (i.e., predictive maintenance)

**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?**

**a clustering algorithm (unsupervised learning)** to segment your customers into clusters of similar customers eg. **The k-means clustering algorithm**

**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**

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

*Online learning system is a learning system in which the machine learns as data is given in small streams continuously.*

**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 is the difference between a model parameter and a hyperparameter in a learning**

**algorithm?**

Model parameter determines how a model will predict given a new instance; model usually has more than one parameter (i.e., slope of a linear model). Hyperparameter is a parameter for the learning algorithm, not of a model.

**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 & its purpose?**

validation set is this subset of the training dataset, called the validation set, that can be used to get an early estimate of the skill of the model.

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

**Training Data** Learning algorithm like gradient descent use training data iteratively to learn the parameters of the model.

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

**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.