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

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.

It 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?**

1. Image Recognition
2. Speech Recognition
3. Traffic prediction
4. Product recommendations
5. Email Spam and Malware Filtering

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

A labeled training set is a training set that contains the desired solution ( label) for each instance.

The training set is used to train the algorithm, and then it is used for the trained model on the test set to predict the response variable values that are already known

Labeled data is a designation for pieces of data that have been tagged with one or more labels identifying certain properties or characteristics, or classifications or contained objects. Labels make that data specifically useful in certain types of machine learning known as supervised machine learning setups.

**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?**

Everyday 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 is likely to perform best if we want a robot to learn to walk in various unknown terrains since this is typically the type of problem that Reinforcement Learning tackles. It might be possible to express the issue as a supervised or semi-supervised learning problem, but it would be less natural.

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

If it is not known that how to define the groups, then a clustering algorithm ( unsupervised learning ) is used to segment customers into clusters of similar customers. However, if it is known that what groups are there, then it can feed many examples of each group to a classification algorithm ( Supervised learning ), and it will classify all customers into these groups.

**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?**

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 at once.

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

Out-of-core algorithms can handle vast quantities of data that cannot fit in a computer’s main memory. An out-of-core learning algorithm chops the data into mini-batches and uses online learning techniques to learn from these mini-batches.

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

An instance-based learning system learns the training data by heart; then, when given a new instance, it uses a similarity measure to find the most similar learned cases and uses them to make predictions.

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

A model has one or more model parameters that determine what it will predict given a new instance ( e.g., the slope of a linear model).

A learning algorithm tries to find optimal values for these parameters such that the model generalizes well to the new instance.

A hyperparameter is a parameter of the learning algorithm itself, not of the model (e.g., the amount of regularization to apply).

**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 algorithms search for an optimal value for the model parameters such that the model will generalize well to new instances.

We usually train such systems by minimizing a cost function that measures how corrupt the system is at making predictions on the training data, plus a penalty for model complexity if the model is regularized.

To make predictions, We feed the new instance’s features into the model’s prediction function, using the parameter values found by the learning algorithm.

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

Some of the main challenges in machine learning are the lack of data, poor data quality, nonrepresentative data, uninformative features, excessively simple models that underfit the training data, and overly complicated models that overfit the 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 a model performs excellently on the training data but generalizes poorly to new instances, the model is likely overfitting the training data ( or we got fortunate on the training data).

Possible solutions to overfitting are getting more data, simplifying the model ( selecting a simpler algorithm, reducing the number of parameters or features used, or regularizing the model ), or reducing the noise in the training data.

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

A test set is used to estimate the generalization error that a model will make on new instances, before the model is launched in production.

**17.What is a validation set's purpose?**

A validation set is used to compare models. It makes it possible to select the best model and tune the hyperparameters.

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

The train-dev set is used when there is a risk of mismatch between the training data and the data used in the validation and the test datasets ( which should always be as close as possible to the data used once the model is in production).

The train-dev set is a part of the training set that’s held out ( the model is not trained on it ). The model is trained on the rest of the training set and evaluated on both the train-dev set and the validation set.

If the model performs well on the training set but not on the train -dev set, then the model is likely overfitting the training set.

If it shows well on both the training set and the train-dev set, but not the validation set, then there is probably a significant data mismatch between the training data and the validation + test data, and you should try to improve the training data to make it look more like the validation + test data.

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

If we tune hyperparameters using the test set, we risk overfitting the test set, and the generalization error we measure will be optimistic ( you may launch a machine learning model that performs worse than you expect).

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