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

Machine learning is a branch of AI focussed on building system that learn from data. At its core, machine learning allows computers to identify patterns and make decisions with minimal human intervention.

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

Fraud Detection.

Recommendation Systems.

Healthcare and Medical Diagnosis.

Predictive Maintenance in Manufacturing.

Natural Language Processing

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

A labeled training set is a collection of data used in machine learning where each piece of data is paired with a correct answer or outcome, known as a label.

how does it work?

**Collection of Data:** First we gather the data which the system what to learn.

**Labeling:** For each sample data, we specify the correct answer.

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

In supervised learning, the two most important tasks are classification and regression:

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

In Unsupervised learning, the four examples are Clustering, Dimensionality Reduction, Anomaly Detection, Association Rule Learning.

**6.State the machine learning model that would be best to make a robot walk through various**

**unfamiliar terrains?**

Reinforcement Learning (RL) is the best fit for this task.

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

Logistic regression, SVM, K means clustering

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

**problem?**

Supervised learning problem

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

The concept of an online learning system in machine learning refers to a model that learns incrementally, as data becomes available, rather than being trained on a complete dataset all at once.

Example:

Recommender Systems, Fraud Detection.

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

Out-of-core learning refers to a set of techniques used to train machine learning models on data sets that are too large to fit into the memory of a single computer. Out-of-core learning is particularly useful for handling big data scenarios where the volume of data exceeds the system's memory capacity.

**how does it differ from core learning?**

Memory Requirement:

Out-of-Core Learning: It is designed to handle datasets that are too large to fit into the system's RAM all at once. It processes the data in small batches, reading them from external storage (like disk), which allows it to work with very large datasets by not requiring all the data to be loaded into memory simultaneously.

In-Memory (Core) Learning: Requires that the entire dataset be loaded into the system's RAM for processing. This approach is limited by the amount of available memory, restricting it to datasets that can comfortably fit within the system's RAM.

Performance and Scalability:

Out-of-Core Learning: While it enables processing of larger datasets beyond memory limits, it can be slower than in-memory processing due to the overhead of reading data from slower external storage in batches.

In-Memory (Core) Learning: Tends to be faster for datasets that fit into RAM due to quicker data access speeds, but it lacks the scalability to handle datasets larger than the available memory.

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

k-Nearest Neighbours (k-NN) algorithm or Supervised learning algorithms.

**12.Whats the difference between a model parameter and a hyperparameter in a learning**

**algorithm?**

**Model parameter:** Model parameters are the configurations that are learned from the data during the training process. They are internal to the model and directly determine the model's predictions. Essentially, these are the aspects of the model that are optimized to fit the model to the training data.

**Hyperparameter:** Hyperparameters, on the other hand, are the configurations external to the model and cannot be learned from the data. They are set prior to the training process and govern the overall behaviour of the learning algorithm. Hyperparameters are used to tune and control the training process.

**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 aim to construct a model that can make predictions or decisions based on input data. These algorithms look for several key criteria during the learning process to achieve success:

Accuracy, Efficiency, Interpretability.

Most Popular Method to Achieve Success is Gradient Descent.

Method Used to Make Predictions is Forward Propagation

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

Overfitting, Underfitting, Model Interpretability, Explainability, Data Quality and Availability, Dealing with Non-stationary Environments.

**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?**

This situatios is called overfitting.

To avoid this we can Use Regularization Techniques like Lasso, Ridge, ElasricNet.

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

A test set is a collection of data used to evaluate the performance of a machine learning model after it has been trained.

Purpose of a Test Set:

Model Evaluation, Detecting Overfitting.

The test set is crucial for understanding how a model will perform in the real world, where it will encounter data it has not been explicitly trained on. It provides a final, unbiased performance metric that can be used to make informed decisions about deploying the model in a production environment or for further development.

**17.What is a validation sets purpose?**

A validation set serves a critical purpose in the machine learning model development process. It is a subset of the dataset that is used for tuning the model's parameters and for making decisions about the model without using the test set.

The validation set is essential for making informed decisions about the model during the development process, allowing for adjustments to be made based on its performance on unseen data. This iterative process of training, validating, and adjusting helps in developing a model that not only fits the training data well but also generalizes effectively to new, unseen data. The use of a validation set thus contributes significantly to the development of robust, high-performing machine learning models.

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

The "train-dev set" plays a crucial role in the development of machine learning models, especially in scenarios where models are being developed and evaluated on large datasets or where there's a risk of distribution mismatch between the training data and the real-world data the model will encounter. It acts as an intermediary step between training and final evaluation, helping to fine-tune models and select the best candidate before the final assessment on the test set.

When You Need It:

Complex Model Development, Large Datasets, Distribution Mismatch Concern.

How to Use It:

Split Your Data.

Model Training: Train your model on the training set as usual.

Model Evaluation: After training, evaluate your model on the train-dev set. This step is crucial for understanding how your model performs on unseen data but still within the training distribution.

Adjustment and Selection: Use the performance on the train-dev set to make adjustments to your model or to select the best model among several candidates. This can involve tuning hyperparameters, adjusting model complexity, or exploring different model architectures.

Final Evaluation: Once you've selected the best model and made necessary adjustments, you can proceed to evaluate its performance on the test set, which should be kept separate and only used for this final evaluation to ensure an unbiased assess

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

The key problems include:

Overfitting to the Test Set

Biased Performance Estimates

Loss of Model Validation Integrity

Decreased Model Robustness