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

Machine learning refers to understanding and developing methods that allow a machine to “Learn”. ML models use algorithms to identify certain patterns in a dataset and predict a suitable outcome of a similar situation. These models are useful in situations where traditional explicit programming is not able to deliver the desired results due to constantly changing parameters and outcomes, Machine learning allows us to perform many tasks like predicting outcomes or analysing large amounts of data in a relatively small amount of time that can provide insights and allow for precise decisions that would otherwise be impossible or take an impractical amount of time, human effort and error.

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

Machine Learning can be used in areas where constant monitoring of data is required or a large amount of data needs processing. For example:

1. Email Spam Detection
2. Safety monitoring of chemical plants for Early detection of any Faults
3. Statistical Modelling of sales data of a company
4. Analysing Weather patterns and predicting upcoming conditions

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

Labelled training datasets contain labels for the provided data features that give context about the data being processes that can allow for better understanding of data and deciding which features are actually essential for building of the model

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

Classification and Regression

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

Dimensionality reduction and Clustering

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

Reinforcement learning is best suited for this task as it would allow the robot to optimize its decisions based on the response of the terrain

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

Clustering

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

Supervised. It needs to have a set of example mails to find relevant patterns and determine if the messages are spam or not

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

Online learning system is a different approach to ML as the data here is constantly fed to the system and the model is constantly changing based on the new information gathered from the new data. It employs a combination of various ML techniques where data comes in a sequential manner and the algorithm updates the best predictor method constantly for best results.

Traditional ML models often use batch training of data where the model is trained “manually” to update it at regular intervals to keep its predictions relevant. This creates a problem in scenarios where the behaviour of data is constantly changing and the predictions of the model may not reflect the absolute latest trends in the data since it is not being updated fast enough.

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

Out-of-core ML is used in scenarios where the dataset is too large to fit into the Memory or RAM of the machine and has to be accessed from the storage such as a hard drive or a webserver.

Unlike Core learning where whole of the dataset is ready for random access by the model as it is loaded into the memory, out-of-core approach requires handling the data differently. This involved “Streaming” the data, which essentially means feeding small chunks of data to the model in batches instead of the whole dataset directly

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

Instance based learning methods like KNN use similarity measure for classification and regression problems

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

For a learning algorithm, model parameters are determined by the training data like the dimensions of the model, number of datapoints etc.

Hyperparameters on the other hand are parameters that change the behaviour of the model and must be tuned accordingly to get the optimal performance. These are different for various algorithms as each used differing parameters like the values of constants and learning methods

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 look for the optimal parameters that give the best results for the new instances. These often use a cost function and determine parameters that can minimize the function. The model then makes a prediction using the values from the new instance and the parameters that delivered the optimal results

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

Building an ML model can be tricky as one might face numerous challenges in the process.

Overfitting is a challenge quite common while building a learning model born as a result of achieving high accuracy while training which can lead to poor performance in real world usage.

Underfitting is another challenge as a result of avoiding overfitting, this can also result in a poor model performance of the model in testing and deployment

Quality of data can be a big issue while trying to build a ML model. If the data received for training contains null, missing or wrong values in essential features, the model produced will not be very effective due to lack of information

Keeping a model updated with ever growing data and maintaining the effectiveness is also a big issue. With new data the current implementation of the model may not continue to give the optimal results and hence will need constant tuning or just complete replacement as it may not be able to remain effective for the intended purpose. Hence regular monitoring and maintenance is required.

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 is the case of and overfitted model that does not preform well in prediction using the new data.

To resolve this, we can use data augmentation, removal of outliers or selecting a different algorithm for the model

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

A test set is a dataset with known outcomes which is used to determine the accuracy of a model outside of a training dataset. A test set is needed to determine how well a model performs outside of a training environment and gauging its feasibility in a real world scenario.

17.What is a validation set's purpose?

A validation dataset is used to evaluate the performance of a model during data preparation and hyperparameter tuning. This term is however often used interchangeably with “test” as well as they serve same purpose essentially.

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

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

Using test dataset to tune hyperparameters can lead to the model developing a bias towards the test data as the model can “see” the data. This creates a problem of data leak where the model does not actually know the test data but can develop a bias towards its nature reducing the ability of the test data to properly evaluate the performance of the model.