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

Machine learning is a type of Artificial Intelligence that allows application to predict some output on the basis of input. And it further divided in some category.

Supervised Learning

Unsupervised Learning

Semi Supervised Learning

Reinforcement Learning

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

Machine Learning algorithm are good for some problem such that spam detection, Fraud detection, Diagnosis, Automatic Car driving.

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

A dataset that comes with a label is called labelled dataset.It is used in supervised machine learning.

for eg let us say that we are trying to determine that if a cat is healthy or sick based on its features then that healthy/sick is a label.

or if you are trying to determine if animal is cat or dog based on its features then that cat/dog is label column

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

For supervised machine learning there are two major tasks is Regression and Classification.

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

Four Example where we use unsupervised task is:-

For Grouping the data

For Image compression

For remove noise from visual data

Use to find relationship between the data

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

Reinforcement Learning model that would be best to make robot walk through various unfamiliar

Terrains.

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

KNN (K Nearest Neighbour) is used to divide customer into different 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. This means you have to provide your model to set of spam and harm message.

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

Online Learning is a concept in which data become available in a sequential order and is used to update the best predictor.

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

Out of core learning refers to the Machine Learning algorithms working with data cannot fit into the memory of a single machine but it can be fit into the local storage.

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

The Machine Learning algorithm which makes predictions using similarity measure is called instance-based because it builds the hypotheses from the training instances. It is also known as memory-based learningorlazy-learning**.**The time complexity of this algorithm depends upon the size of training data. The worst-case time complexity of this algorithm isO (n), where n is the number of training instances.

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

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.

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?

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. Model-based system learn by minimizing a cost function that measures how bad the system is at making predictions on new data, plus a penalty for model complexity if the model is regularized. To make a prediction, a new instance's features are fed into a hypothesis function which uses the minimized theta found by repeatedly running the cost function.

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

Not enough data or noise in data

Outliers and missing value in dataset.

Feature Selection and Feature Extraction.

Overfitting and Underfitting of 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?

This is the case of overfitting to remove the overfitting we can do

1. Removing Feature
2. Removing outliers
3. Collect More data

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

When we want to know how well our model generalizes to new cases, we prefer to use a test set instead of actually data. To build the test set we split the training data into a training set and test set. Our model is training with the training set. Then we use the model to run predictions on the test set. Our error rate on the test set is called the generalization error or out-of-sample error. This error tells us how well our model performs with new data.

17.What is a validation set's purpose?

validation data is usually 20%of our bulk dataset it is used in stacking generally to test the model after hyperparamater tuning

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

while building a machine learning model we always build multiple data based on our dataset

we use training dataset to train all those models. In order to rank these models to select the best one,we

test the accuracy of our models based on validation set(or dev set).

And after selecting our best model and in order to be sure of the accuracy part we use test data as a proxy of unseen dataset to check its accuracy

for instance--

while making multiple models for problem statement we have to do two things

1)select best parameter for our model

2)select best model

so in case of train dev split, we train the data on our training dataset and then select the dev dataset is used to select best model

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

Model will not be generalized to new data.