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

Machine learning is the science of programming computers so that they can learn from the data

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

In health care industry to predict the probability of getting a particular disease

In Supply chain to predict the occurrence of issue for any product

In ecom to provide the personalized recommendations for users

In IT industry to detect malicious users or prevent from hacking attacks

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

Labelled training set means dataset we feed to an algorithm which has the desired solutions also included along with the features

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

Regression and classification

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

Clustering, visualization, dimensionality reduction and association learning rule

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

Reinforcement learning as it can learn from the responses of the terrains to optimize itself

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

Clustering algorithm

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?

In Online learning model learns incrementally by feeding the model data instances sequentially either individually or by small groups called mini batches

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

The algorithm loads part of data, runs a training step on that data and repeats the process until it has run on all the data

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

Instance based learning

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

Hyper parameter is a parameter of learning algorithm not of the model and model has one or more model parameters that determine what it will predict given a new instance. A learning algorithm tries to find optimal values for these parameters such that the model generalizes well to new instances. A hyperparameter is a parameter of the learning algorithm itself, not of the 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 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 bad 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 predictions function using the parameter values found by the learning algorithm.

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

1) Insufficient Quantity of Training Data

2) Nonrepresentative Training Data

3) Poor-Quality Data

4) Irrelevant Features

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?

The model is said to be overfit in this case.

This can be overcome by below:

Gathering more training data

Reduce noise in the data by removing outliers and data pre-processing

Removing the unwanted features

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

A test set is subset of data set that will be used to test the model. This 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?

Validation set is sub set of train data set which is held to validate the selected model and hyperparameters. It is used to compare models and select the best model by tuning hyper parameters

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

It is used to rank the models in term of their accuracy and helps to decide which model to proceed further with. Train set is used to give parameters of each of the model choice but can’t decide which model is good one based on train set so Dev set is used to rank models in terms of accuracy

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

The model will be tuned for that test set but is unlikely to perform well on new sets which can cause over fitting