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

Machine Learning is the subset of artificial intelligence which gives us the facility to feed the machine with data and the machine has the ability to learn from that data and extract meaningful information.

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

Customer segmentation, sales forecasting, fraud detection, product recommendation

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

Labelled training set means we have specified label against every data point. These labels help in supervising the model prediction.

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

Regression and classification

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

Customer segmentation, anomaly detection, products association, dimensionality reduction

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

Reinforcement learning

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

Clustering algorithms like k-means and hierarchical clustering

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

A supervised classification problem.

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

In online learning system, we can work on live data. This type of data changes frequently. Online learning also facilitates us working with out of core data.

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

Sometimes we have very huge sized data. We cannot feed all the data at one time. It will crush the processor or make it very slow. Out-of-core learning helps in dealing with that. It accepts small chunk of data in a sequential manner and process it.

Core learning uses all the available processor of the system. It usually supports small to medium sized dataset.

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

K nearest neighbours or KNN

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

Model parameter are mathematical parameters which are trained during model learning.

Hyperparameters facilitates in training and optimizing the model parameters.

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 tries to identify the hidden pattern in the dataset. It uses cost function optimization. It takes a number of iteration to achieve that. Finally based on minimum cost function, it makes the prediction.

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

* Unavailability of significant amount of data
* Data cleaning complexity
* Lack of representative data
* Under fitting, over fitting

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. We can use following techniques to deal with that…

* We can use some simpler models.
* Use higher number of constraints.
* Data cleaning.
* Provide some more significant variables .

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

Test set is used to test the performance of model. It challenges our model on completely unseen type of data.

17.What is a validation set's purpose?

Validation set is a significant sized subset of training set. It is used to optimize the model performance.

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

A validation set is called the train-dev set. It is used to optimize the model performance. We hold out the few data point from the training set or we can use k-fold cross validation.

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

In that case the model will perform greatly even on the test set because the model already knows the test data points. But as we feed completely unseen data points, the performance will go down drastically.