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

Machine Learning is a field of study which uses statistical tool for analysing, interpreting and for doing descriptive analysis of the given dataset.

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

Sentiment Analysis

Customer Segmentation

Speech Recognition

Face Detection

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

A labeled training set is a dataset in which for some given input features, we are provided with actual output. The model is trained based on this.

After this the test data is passed through the model and the output is recorded. Now the output is compared with the predicted output of the model and the accuracy is measured.

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

Classification and Regression

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

Clustering

Visualisation

Dimensionality Reduction

Association Rule Learning

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

The best model in such a case is Reinforced ML, where the model can learn from Incoming Data/Experience.

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

If the group has labels, then supervised ML

If not, then unsupervised(Clustering) ML.

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

Spam detection is a kind of supervised machine learning technique as it gives the output as “Yes” or “No”.

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

Online learning system is a learning system in which the model learns as data is given in small stems or on the fly.

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

It is a way of training our model on a dataset which can’t fit in our main memory.

In this the data is read by breaking the data into chunks. This is done using data reader. After this the process of feature extraction takes place and at last model is trained based on this. It might be possible that even after feature selection, Our data is large as compared to our main memory. So, a process of incremental learning is used where our algorithm is trained by simultaneously giving mini batches of data. This method can only be used in those algorithm in which partial\_fit method is available like Naïve Bias, XGD Regressor etc.

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

Instance based Algorithm – This type of algorithm learns the training dataset by heart and then make predictions using some similarity index.

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

Model parameters are those that must be determined using training dataset abd is used to predict new instances.

Hyperparameter is for learning algorithm and is not used for a 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 algorithm looks for the best values of a parameter that will give the best results for new instaces. The method used is reducing the cost/loss function associated with them so that maximum accuracy can be achieved. The model makes prediction using newer data and the parameter in its function.

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

Overfitting

Underfitting

Not enough Training Data

Non representative Training 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 kind of situation leads to model overfitting.

For avoiding overfitting condition we uses the regularisation parameters for generalizing the results to new situations.

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

Test set is a percentage of dataset kept out of original dataset through which the accuracy of our model is calculated. The input from these dataset are passed on to the model and predictions are recorded. These predicted values are then compared with our actual ones and accuracy is calculated.

17.What is a validation set's purpose?

This dataset is used to compare the performance of various prediction algorithms that were created based on the training set. The best performing algorithm is selected.

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

This basically helps us to rank the models in terms of their accuracy and helps us to decide with which model to proceed further.

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

Our model will not be generalised for new examples.