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

Machine learning is subset of Artificial intelligence wherein we try to train machines to mimic human intelligence using different algorithms.

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

Recommendation engine, Fraudulent transactions, Sentiment Analysis, Image Recognition

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

Labelled training set contains both input and output. First it is split into train and test. Then train the model using training data and test the performance using test data and fine tune the model. Then when a new data comes in, the model can predict the outcome based on the learning happened from training data.

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

Classification, Regression

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

K means Clustering, Hidden Markov model, PCA, SVD

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

Reinforcement Learning – Learns based on actions and rewards

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

K means clustering

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

Supervised learning where it learns from already labelled data.

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

In online learning system, data will be fed in a sequential manner rather than batch. The model gets trained as and when the data arrives. This is very much useful when the system does not have the capacity to store the whole data.

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

In core learning, data will be stored in memory of a single machine. But in case of out-of-core learning, data will be stored in external memory.

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

Instance based learning algorithms – learns examples by heart and then uses similarity measures to generalize it.

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

Model parameter is internal to the model and it is learned from the training data. It is used to predict the outcome whereas the hyper parameter is external to the model and is involved in the learning.

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?

It creates a model which is optimized by training with the train data and then used for predicting outcomes when a new data comes in. It optimizes the model using cost function. Learning is continued till the cost function reaches global minima. Corresponding model parameters are estimated and used for prediction.

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

Need a lot of data, overfitting, under fitting, noise data, need machines with high computational power.

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 called overfitting.

1. Reduce the complexity of algorithm
2. Train with more data
3. Reduce irrelevant data

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

We usually split the data for training into two – train data and test data with the ratio 80:20. Test data is used for analysing how well a model predicts based on the training done on train data. Test error is used for telling the model accuracy.

17. What is a validation set's purpose?

Validation set is used to tune the hyper parameters used to train the model. Validation set is created along with train and test set. Then model is fit using training set and optimization done using validation set. Before the model deployment, model accuracy is estimated using test set, where it remains as a data which the model has not seen yet.

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

Train-dev kit contains the data for training. It is needed to build a model. We can download it from the publicly available sites like Kaggle. Based on the domain we are working on, we can download corresponding dataset and use to build our model.

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

If we calculate error based on test set after each iteration, the model gets biased to test set. We will not have any data for testing purpose which the model has not seen before. This will reduce the model’s ability to predict outcomes on new data.