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

Machine Learning is the field of study that gives the computer's ability to learn without being explicitly programmed.

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

1. Problems which requires lot of conditions or rules to solve.

2. Complex problems where current solutions are using traditional approaches where the following approach is not a good solution to solve the problem.

3. Machine Learning systems can adapt to new data where there are fluctuating environments.

4. Machine Learning systems can get insights from large amount of data and complex problems.

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

In supervised learning technique, the data we feed to the ML algorithm includes the desired output, called **labels.**

The algorithm first learning from the training using the desired solutions and the predict on the test set, the test does not contain the desired solution which means the model predict on unknown data.

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

Regression and Classification are two most important tasks of supervised learning.

The main difference between regression and classification:

In **Regression,** the target or label variable is **continuous**

Example: height, weight, price, age etc…

In **Classification**, the target or label variable is **discrete** in nature

Example: Number of children, Number of bank accounts, gender, etc…

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

1. Clustering – Clustering means grouping the similar data points.

2. Anomaly detection and novelty detection – Anomaly detection is for detecting the unusual things like catching manufacturing defects, fraud transactions etc…

3. Dimensionality reduction – It is used to reduce the number of features without losing to much information.

4. Association rule learning – Main aim in association rule learning is to dig the insights from large amount of data and discovering the interesting relations between features.

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

Reinforcement learning is the model that would be best to make a robot walk through various unfamiliar terrains.

The working of reinforcement learning is

1. Observe

2. Select action using a policy

3. Perform Action

4. Get reward or penalty

5. Update policy

6. Iterate until an optimal policy is found.

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

I will use clustering algorithms for dividing customers into different groups

1. K-means and K-means++

2. DBSCAN

3. Hierarchical clustering

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

I will consider the problem of spam detection to be a supervised learning problem.

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

Online learning is also known as incremental learning, means training the system incrementally by feeding the data sequentially or by mini batches.

It is great for the systems that receive data continuously or need to adapt for rapid changes.

If we have limited computing resource, once a algorithm has learned about data it does not need them anymore we can discard them, this saves a huge amount of space.

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

Out of core learning means when we have huge data that cannot be fitted into main memory we can load a part of data, perform training and repeat the process.

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

Instance based learning algorithms makes predictions using a similarity measure, means comparing the similarity between the data points and then predict accordingly using a similarity measure.

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

Model parameters are those parameters that are internal to the model and estimated from the data automatically.

Hyper parameters are the parameters which are set manually for the optimization 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 the optimal value of parameters in a model that will give best result on unseen data.

The most popular method is to use the cost function and minimize the cost function to achieve the success. After minimizing the cost function then that model is used to make predictions

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

Four most important challenges of Machine learning

1. Irrelevant data

2. Insufficient quantity of data

3 Poor Quality of data

4. Over fitting and under 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?

If the model is performing well on training data but fails to generalize on new situations the probably the model is **over fitted.**

To overcome the over fitting problem

- Provide sufficient amount of quality data
- Reduce the noise in the training data i.e remove outliers, handle missing values, fix data errors
- Simplify the model with few parameters
- Perform hyper parameter tuning

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

If we want to know our model is working or not on new data is to put the model on production and monitor it. If it works then fine, else if it working very bad then users will complain and this is not the best idea.

A better option is to split the data into training and test set.

- Training set is used for training the model
- Testing set is used to test the model, the error rate on test set is generalization error(Out of sample error)

17. What is a validation set's purpose?

If our model is not performing on test set, so we tried different parameters and measure generalization error multiple times on same test set and got out parameters to perform its best on test set. But this means the model is going to perform very poorly on new unseen data.

So the solution for this problem is to hold out some part of training data and evaluate the models on hold out set. The hold out is called **Validation set.**

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

The train-dev kit is also known as validation set. We need when the model is performing well on test set but not on new unseen data.

It is used from the part of the training data.

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

The model will not perform well on new unseen data if we use the test set multiple times to tune hyper parameters.