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

Machine Learning means an algorithm which helps the computers to learn the patterns from the data. Contrary to traditional rule based programing, in machine learning the algorithm tries to establish a relationship between features and labels (independent variables and dependent variables) based on the data itself, without being explicitly programmed. The purpose of this learning is to predict an outcome for a similar but unknown data instance.

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

Machine learning shines in the following issues:

1. Prediction of Housing Prices, Prediction for Loan Default for new loan application
2. E Commerce Recommendation Systems
3. OTT Platform recommendation system
4. Computer Vision problems
5. Natural Language Processing problems
6. What is a labelled training set, and how does it work ?

Labelled Training data consists of the training data and correct output for that training data. In other words, when the correct answer is given along with various features in the data, then we call this type of data as labelled data. The correct answer is called as label and the training attributes are called features.

For example, when we try to predict the heart attack chance based on a data set which consist of blood sugar, blood pressure, body weight, cholesterol as feature and whether heart disease exist as result then this type of data is called labelled data as the correct output is given with the dataset.

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

Regression and classifications task are performed under supervised category.

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

A. Detection of fraudulent financial transaction

B. Virtual Private Assistant

C. Image recognition

D. Recommendation problem

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

Re inforcement learning is suitable to handle this task

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

K Means clustering

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

Spam detection is Supervised algorithm

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

Online learning is a combination of different techniques of ML where data arrives in sequential order and the learning algorithm aims to learn and update the best predictor for future data at every step.

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

Out of the core learning- These are learning algorithm which work with the data that can not fit into a single computer’s RAM. In these type of learning such big datasets are kept either external computer network or web repository but still they try to access the remote data in one sequence. Therefore out of core learning tries to minimise the performance issue while accessing the data which are remotely stored.

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

Instance based learning

1. What's the difference between a model parameter and a hyper parameter in a learning algorithm?

A model parameter is a configuration variable that is internal to the model and whose value can be estimated from data.

A hyper parameter is a configuration that is external to the model and whose value cannot be estimated from data.

1. 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 search for the optimal value of parameters in a model that will give the best results for unknown data.*

*Model based learning algorithm produces a hypothesis function which establish the relationship between independent variable and dependent variable, in other words it tries to find our parameters which maps the dependent variable from independent variable. It minimizes the gradually through a cost function by adjusting its parameters.*

When a new instance arrives their features are fed into a hypothesis function which uses the minimized parameters (theta) found by repeatedly running the cost function.

1. Can you name four of the most important Machine Learning challenges?
2. Over complicated model (Over fitting)
3. Simple model (under fitting)
4. Dataset is filled with missing, unresponsive data
5. Multi collinear data
6. 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 phenomenon is called overfitting

Following approach may be adopted:

1. Simplify the model by removing non essential features
2. Collect more relevant data
3. Remove outliers from the data.
4. What exactly is a test set, and why would you need one?

*Test set is data set that you test your model which has trained using training data, to see how it performs. Test set is necessary so that you can determine how good (or bad) your model performs with unseen data.*

1. What is a validation set's purpose?

In real world situation it is better to have a validation data set which is a second holdout set. We train multiple models with various hyper parameters using the training set, we select the model and hype parameters that perform best on the validation set, and when we are happy about our model we run a single final test against the test set to get an estimate of the generalization error.