1. what are different type of machine learning?

Supervised (labelled data, Regression, classification) , Unsupervised (Clustering), Reinforcement (alpha-go, agent and environment)

1. What is over fitting? How can you avoid this?

Model is memorising the data. Will be have good accuracy on training data but very low in validation data

Regularisation (drop out in neural network) : this involves a cost term for the features involved In the objective function

Makes a simple model :- lesser variables and parameters, the variance can be reduced,

Cross validation methods like k-fold

Lasso regularisation can be used to penalise the parameters which cause over fitting

1. What is training and test set in machine learning model? How much data will you allocate for modelling?

Training sets are examples given to model to study and learn and test sets are given to model to evaluate the models performance 70:30 is the ratio

1. How do you handle missing values in a dataset?

Drop those rows or column or replace them completely with some other value

isnull(), Dropna() will help to find the columns / row contains missing values

fillna() will replace wrong values with placeholder value(0)

1. How can you choose the classifier based on training set size

If the training data set is small. Model with high bias and low variance will works better

When training is large models with low bias and high variance like decision tree will work

1. Explain confusion matrix or error matrix:

Confusion Metrix is a specific table used to measure the performance of a classification problem. It is mostly used in supervised learning (In unsupervised learning it is called matching matrix. It has two dimension. Both dimension has identical sets of features.

1. What is false positive and false negative ?

False positive are the cases where the model predicted as true but are actually false.

False negative are the cases where the model is predicted as false but are actually true

Accuracy, precision, f1 score

1. What is deep learning?

Learn like human using artifitual neural network. Subset of ML, deep neural network, Feature engineering done by system.

Machine learning enable machiens to take decision on their on based on past data, Need small amount of training data and low end system. Most features need to be identified in advance and manually coded. Problem is divided into parts and solved individually and combined.

In deep learning enables machines to take decision with the help of neural network. Needs large amount of training data and needs high end system to work. System learns the features from the data and problem is solved in end to end manner.

1. What are the application of supervised machine learning?
2. What is semi supervised learning ?
3. What are unsupervised ml?

Clustering : If most of the characters are similar we can put that into same cluster.

Association: We identify the patterns of the association between different variable.( Suggestion of other item to by in ecommerce website.

1. Inductive machine learning, Deductive machine learning ?

Inductive Machine Learning: It observes instance based on defined principle to draw conclusion (studying things by watching video)

Deductive Machine Learning: It draws conclusion from already defined general principle.

1. What is knn and k-means?

Knn supervised classification and k-means is unsupervised clustering algorithm.

Knn classifies unlabelled observation based on its K( can be any number) surrounding neighbours. K-means: The point in each cluster are similar to each other and each cluster is different from its neighbouring clustering.

1. What is naive bayes classifier?

Probability based classification . It makes assumption that may or may not turn out to be correct. The algorithm assumes the presence of one feature of a class not depend on the presence of other feature.

1. What is random forest?

Random forest is supervised machine learning algorithm that is generally used for classification problem. It consit of multiple decision tree in during training phase. Ensamble learner. Consist of more than one algorithem.