Q1) What are the three stages to build the hypotheses or model in machine learning?

The three stages to build model in ML are:

a) Model Building
b) Model Testing
c) Applying the model

Q2) What is the standard approach to supervised learning?

Supervised learning is the machine learning task of learning a function that maps an input to an output

based on example input-output pairs. It infers a function from labeled training data consisting of a set of training examples.

A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples.

So the standard approach to supervised learning is to split the set of example into the training set and the test.

Q3) What is Training set and Test set?

Set of data is used to discover the potentially predictive relationship known as 'Training Set'.

Training set is an examples given to the model, while Test set is used to test the accuracy of the hypotheses generated by the model.

So in general Training Set is a dataset which is used to train the model based on a specific features.

While Test set is used to predict accuracy of model on test data.

Q4) What is the general principle of an ensemble method and what is bagging and

boosting in ensemble method?

Ensemble method is using multiple learning algorithm together for the same task to obtain better result than individual learning.

so The general principle of an ensemble method is to combine the predictions of several models built with a given learning algorithm in order to improve robustness over a single model.

So the output/accuracy of ensemble method should be higher than any of individual model else there is no more use of putting extra

computation power.

1)Bagging (bootstrap Aggregation): bagging is multiple model of same learning Algorithm trained with subsets of dataset randomly picked from training dataset.

2) Boosting: We give more emphasis on dataset which gives wrong prediction and again create the bag with old wrong prediction.boosting method are used sequentially to reduce the bias of the combined model.

Q5) How can you avoid overfitting?

A hypothesis h is said to overfit training data if there is another hypotheses h1 such that h1 has more error than h on training data but less error than h on test data.

overfitting happens relatively as on small dataset, and model try to learn from it. So one approach is to add more data so to avoid overfit.

But if we have a small data and are forced to come with a model based on that. In such situation, we can use a technique known as cross validation.

In this method the dataset splits into two section, testing and training datasets, the testing dataset will only test the model while, in training dataset, the data points will come up with the model.

In this technique, a model is usually given a dataset of a known data on which training (training data set) is run and a dataset of unknown data against which the model is tested. The idea of cross validation is to define a dataset to “test” the model in the training phase.