1.What are the three stages to build the hypotheses or model in machine learning?

Model Building

Applying the Model

Model Testing.

2. What is the standard approach to supervised learning?

Split the set in to Training set and Test set is the standard approach.

3. what is training set and test set?

In machine learning, a training set is a dataset used to train a model.  In training the model, specific features are picked out from the training set.  These features are then incorporated into the model.

The test set is a dataset used to measure how well the model performs at making predictions on that test set

4. What is the general principle of an ensemble method and what is bagging and

boosting in ensemble method?

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.

Bagging:

Bagging stands for bootstrap aggregation. One way to reduce the variance of an estimate is to average together multiple estimates.

Boosting:

It is a general ensemble method that creates a strong classifier from a number of weak classifiers.

This is done by building a model from the training data, then creating a second model that attempts to correct the errors from the first model. Models are added until the training set is predicted perfectly or a maximum number of models are added.

5.How can you avoid overfitting?

Overfitting is traditionally defined as training some flexible representation so that it memorizes the data but fails to predict well in the future.

The commonly used methodologies are:

- Cross- Validation: A standard way to find out-of-sample prediction error is to use 5-fold cross validation.
- Early Stopping: Its rules provide us the guidance as to how many iterations can be run before learner begins to over-fit.
- Pruning: Pruning is extensively used while building related models. It simply removes the nodes which add little predictive power for the problem in hand.
- Regularization: It introduces a cost term for bringing in more features with the objective function. Hence it tries to push the coefficients for many variables to zero and hence reduce cost term