**1. What are the three stages to build the hypotheses or model in machine learning?**
A: a) Model building b) Model testing c) Applying the model

**2: What is the standard approach to supervised learning?**
A: The standard approach to supervised learning is to split the set of examples into the training set and the test.

**3: What is Training set and Test set?**
A: In various areas of information science like machine learning, a set of data is used to discover the potentially predictive relationship known as 'Training Set'. Training set is an examples given to the learner, while Test set is used to test the accuracy of the hypotheses generated by the learner, and it is the set of example held back from the learner. Training set are distinct from Test set.

**4: What is the general principle of an ensemble method and what is bagging and boosting in ensemble method?**
A: 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 is a method in ensemble for improving unstable estimation or classification schemes. Bagging both can reduce errors by reducing the variance term.
boosting method are used sequentially to reduce the bias of the combined model. Boosting can reduce errors by reducing the variance term.

**5: How can you avoid overfitting?**

A: 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.