Question 1.

What are the three stages to build the hypothesis or model in machine learning?

Ans.

- Model Building
- Model Testing
- Applying the model

Question 2.

What is the standard approach to supervised learning?

Ans.

The standard approach to supervised learning is to split the set of examples into the training set and the test set.

Question 3.

What is training set and test set?

Ans.

- Training 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. Thereby, if the training set is labeled correctly, the model should be able to learn something from these features.
- Test Set –
  The test set is a dataset used to measure how well the model performs at making predictions on that test set. If the prediction scores for the test set are unreasonable, we'll have to make some adjustments to our model and try again.

Question 4.

What is the general principle of an ensemble method and what is bagging and boosting in ensemble method?

Ans.

- General principle of 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 –
  Bootstrap aggregating, also called bagging, is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. It also reduces variance and helps to avoid overfitting. Although it is usually applied to decision tree methods, it can be used with any type of method. Bagging is a special case of the model averaging approach.
- Boosting –
  Boosting is a machine learning ensemble meta-algorithm for primarily reducing bias, and also variance in supervised learning, and a family of machine learning algorithms that convert weak learners to strong ones.

Question 5.

How can you avoid overfitting ?

Ans.

- Cross-Validation
- Pre-pruning
- Post-pruning
- Regularization
- Feature Selection
- Take more data