# Assignment 8.2

## Problem Statement

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

2. What is the standard approach to supervised learning?
   **Answer:** 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?
   **Answer:**
   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.

4. What is the general principle of an ensemble method and what is bagging and boosting in ensemble method?
   **Answer:** The general principle of an ensemble method is to combine the predictions of several models built with a given learning algorithm to improve robustness over a single model.

   Bagging is a method in ensemble for improving unstable estimation or classification schemes. Bagging 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?
   **Answer:** By using a lot of data overfitting can be avoided, overfitting happens relatively as you have a small dataset, and you try to learn from it. But if you have a small database and you are forced to come with a model based on that. In such situation, you can use a technique known as cross validation. In this method the dataset splits into two sections, 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.