**Answer 1:**

The three stages to build the hypotheses or model in machine learning are:

a) Model building

b) Model testing

c) Applying the model

**Answer 2:**

The Standard approach to supervised to learning is to divide the data into Training set and Test set, Here we know the label and we can use test set to test the model by inputting it to the training set model.

**Answer 3:**

The Training set is nothing but the actual data or examples given to the learner, while the Test set is something which is divided from the training set and used to test the accuracy of the hypotheses generated by the learner and also used to test the model for accuracy.

**Answer 4:**

The general principle of the ensemble method is that it create multiple models and then combine them to produce improved results. Ensemble methods usually produces more accurate solutions than a single model would produce.

**Bagging** is an ensemble method. First, we create random samples of the training data set Then, we build a classifier for each sample. Finally, results of these multiple classifiers are combined using average or majority voting. Bagging helps to reduce the variance error.

**Boosting**provides sequential learning of the predictors. The first predictor is learned on the whole data set, while the next are learnt on the training set based on the performance of the previous set**.**Itstarts by classifying original data set and giving equal weights to each observation. If classes are predicted incorrectly using the first learner, then it gives higher weight to the missed classified observation. Being an iterative process, it continues to add classifier learner until a limit is reached in the number of models or accuracy.

**Answer 5:**

We can reduce the overfitting using following methods

**Cross-validation:** Cross-validation is a powerful preventative measure against overfitting, Cross validation allows you to tune hyper parameters with only your original training set. This allows you to keep your test set as a truly unseen dataset for selecting your final model.

**Train with more data:** Training with more data can help algorithms detect the signal better, you should always ensure your data is clean and relevant