**Assignment 20.1**

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

1) Model building  
2) Model testing  
3) Applying the model

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

The standard approach to supervised learning is to split the sample into “training data” and the “test data”. The model is built on the “training data” and then tested on the “test data” to determine its accuracy.

**3. What is Training set and Test set?**

In machine learning, the set of data that is used to discover the predictive relationship is known as ‘Training Set’. Training set is an example set given to the model, while Test set is used to test the accuracy of the hypotheses generated by the model.

**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 is a method in ensemble for improving unstable estimation or classification schemes. Bagging both can reduce errors by reducing the variance term.

Boosting method is 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?**

**Cross-validation**: Use your initial training data to generate multiple mini train-test splits. Use these splits to tune your model.

**Remove features**: Manually improve generalizability by removing irrelevant input features.

**Ensembling**: Ensembles are machine learning methods for combining predictions from multiple separate models.