**Assignment 18 - Machine Learning -1**

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

Following are the stages to build a model:

* 1. Decide an algorithm and data preparation
  2. Prepare testing and training data
  3. Train and test the algorithm
  4. Deployment and monitor

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

The standard approach for supervised learning is to divide the data into Training and Test set.

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

A training set is a dataset used to train a model. In training the model, specific features are picked out and are then incorporated into the model. Training set is the data given to the learner. Training set are distinct from test set.

The test set is a dataset used to test the accuracy of the hypotheses generated by the learner. It is used to measure how well the model performs at making predictions on that test set.

1. **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.  While boosting method are used sequentially to reduce the bias of the combined model.  Boosting and Bagging both can reduce errors by reducing the variance term.

1. How can you avoid overfitting ?

With the large dataset overfitting can be avoided. Overfitting happens usually when we have a small dataset and we try to build a model from it. If our model does much better on the training set than on the test set, then we’re likely overfitting. Cross-validation is a one of the powerful preventative measure against overfitting.