**Assignment 20.1**

**Question1: What are the three stages to build the hypotheses or model in machine learning?**

Three stages to build hypothesis or model in model in machine learning are:

i.     Model building: In this stage the model is built. If it is supervised machine learning, then a appropriate training algorithm is chosen for the dataset and training is performed to iteratively build the model

ii.      Model testing: Data in machine learning is split into training set and test set. In this stage the model is tested against training dataset and various metrics (like Accuracy, Recall, Precision) appropriate for the model are calculated

iii.       Applying the model: The model is deployed in production

**Question2: What is the standard approach to supervised learning?**

Standard approach to supervised learning is first split the data into the training data set and the test data set. Chosen machine learning model is trained against the training set and the model is built. Next, model is tested against test data set and various metrics (like accuracy, recall, precision etc) are calculated and compared. In next step model is deployed

**Question3: What is training and test set**

In supervised machine learning, The dataset is split into training and test set. Training set is an examples given machine learning algorithm to build the model. Test set is used to test the accuracy of the hypotheses generated by the machine learning algorithm Training set are distinct from Test set.

**Question 4. What is the general principle of an ensemble method and what is bagging and**

**boosting in ensemble method?**

To solve a specific computational program, numerous models such as classifiers are strategically made and combined. This process is known as ensemble learning.

The general principle of 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 method for improving unstable estimation or classification schemes

Boosting is an ensemble technique that attempts to create a strong classifier from number of weak classifiers by primarily reducing bias also variance in supervised learning

**Question 5. How can you avoid overfitting ?**

Overfitting can be avoided using techniques below:

1. Large Data set: Train and test against large dataset
2. Cross validation: In this method the dataset splits into two section, testing and training datasets, the testing dataset will only test the model while, in training dataset, the datapoints will come up with the model.
3. Early Stopping: Its rules provide us the guidance as to how many iterations can be run before learner begins to over-fit.
4. Pruning: Pruning is extensively used while building related models. It simply removes the nodes which add little predictive power for the problem in hand.
5. Regularization: It introduces a cost term for bringing in more features with the objective function. Hence it tries to push the coefficients for many variables to zero and hence reduce cost term.