Assignment ML-1

--------------------------------

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

Ans: a)      Model building

b)      Model testing

c)       Applying the model

2. What is the standard approach to supervised learning?

Ans:**Supervised learning** is the machine **learning** task of inferring a function from **supervised training**data. The **training** data consist of a set of **training** examples.

3. What is Training set and Test set?

Ans: 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.

A **Test Set** is a group of **tests** that belong to specific tasks or feature, or have some other reason to be run together. Good candidates for **Test Sets** can be, for example:**Tests** that focus on a certain part of the system, such as GUI or Database. **Tests**that belong to a certain task, such as Regression or Sanity.

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

boosting in ensemble method?

Ans: **Ensemble methods** are learning algorithms that construct a. set of classifiers and then classify new data points by taking a (weighted) vote of their predictions. The original **ensemble method** is Bayesian aver- aging, but more recent algorithms include error-correcting output coding, Bagging, and boosting.

Bootstrap aggregating, also called **bagging**, is a **machine learning** ensemble meta-algorithm designed to improve the stability and accuracy of **machine learning**algorithms used in statistical classification and regression. It also reduces variance and helps to avoid overfitting.

**Boosting** is a **machine learning** ensemble meta-algorithm for primarily reducing bias, and also variance in supervised **learning**, and a family of **machine learning**algorithms which convert weak learners to strong ones. ... Algorithms that achieve hypothesis **boosting** quickly became simply known as "**boosting**".

5. How can you avoid overfitting ?

Ans:

* Add more data.
* Use data augmentation.
* Use architectures that generalize well.
* Add regularization (mostly dropout, L1/L2 regularization are also possible)
* Reduce architecture complexity.