**Question 1: What are the three stages to build the hypothesis or model in machine learning?**

Machine Learning model contains following standard steps:-

* Data processing /Cleaning
* Dividing Dataset into training and test set
* Selecting required model /algorithm
* Checking Result and Modification

**Question 2: What is the standard approach to supervised learning?**

Supervised algorithms can further divided into following:

* **Classification:** When the data are being used to predict a category, supervised learning is also called classification.
* **Regression:** When a value is being predicted, as with stock prices, supervised learning is called

Regression.

* **Anomaly detection**. Sometimes the goal is to identify data points that are simply unusual

**Question 3: What is a training set and Test set?**

**What is a Training Set?**

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. Thereby, if the training set is labeled correctly, the model should be able to learn something from these features.

**What is a Test Set?**

The test set is a dataset used to measure how well the model performs at making predictions on that test set. If the prediction scores for the test set are unreasonable, we’ll need to make some adjustments to our model and try again.

**Question 4: What is the general principle of an ensemble method and what is bagging and boosting in ensemble method?**

* **Ensemble:**

Ensemble learning is used when you build component classifiers that are more accurate and independent from each other.

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 in ensemble**

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

* **Boosting in ensemble**

Boosting methods are used sequentially to reduce the bias of the combined model. Boosting can reduce errors by reducing the variance term.

**Question 5: How can you avoid over fitting?**

1. Use regularization (for optimization-based classifiers). Here, you should take some time to set an appropriated value for the regularization parameter;

2. Increase the amount of samples in the training set;

3. Reduce the amount of features.

4. Use ensemble.