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

Ans : Build model, Test model and Deploy model.

2. What is the standard approach to supervised learning?

Ans : Splitting the dataset into training dataset and test dataset.

3. What is Training set and Test set?

Ans : Training set is the example from which machine learns and Test set is the exam where machine predicts the result.

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

Ans : The ensemble method is to combine the predictions of different models built with a given learning algorithm in order to improve accuracy over a single model. In bagging, the training stage is parallel and each model is built independently. In boosting, the training stage is sequential and each classifier is trained on data, taking into account the previous classifiers’ success.

5. How can you avoid overfitting?

Ans : We know that overfitting leads to great performance of the model while working on training dataset as it learns the data in details including any noise around the data but the performance on test dataset reduces significantly because the model fails to generalize .

The moment we see error on test dataset increases but the error for training dataset decreases, we conclude that training data is overfitted and hence we can start reducing features iteratively and keep evaluating performance of the model on both training and test dataset unless we reach a sweet spot between overfitting and underfitting.