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

Model Building – Fit the model using fit function on training data

Apply the model – Apply it on test data using predict function

Test the model – You have to use score() to test the model and its accuracy

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

Supervised learning is a labelled approach to solve machine learning problems. Here we divide the data into training and test data 70-30 ratio. We fit the model on training data and test it on test data. We should be careful so that the model wont overfit. Some of the tricks are using cross-validation, stratify the labels etc.

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

Training set is the set to train the model for a machine learning algorithm. We can use feature selection to reduce the number of features involved in training the algorithm. If the training set is consistent, the model should learn the behaviour of the data and output a standard function or assumption about the given data

Test set is used to validate our machine learning algorithm. It can also be used to check the overfitting and underfitting problem. We have specific examples to deal with bias and variance tradeoff. Ideally, there should be a sweet spot between overfitting and underfitting the data

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

The general principle of ensemble methods is to use a wide variety of algorithms on the given data for maximum performance or accuracy.

Bagging is also called as Bootstrapped aggregation which is to build a numerous decision trees and improving the model performance which also reduces variance and avoid overfitting. It increases stability of the model.

Boosting is a machine learning approach to reduce bias and variance in supervised learning. It converts weak learners to strong learners by using utilizing ensemble algorithms iteratively until all weak learners are converted to strong learners.

1. **How can you avoid overfitting ?**

You can avoid overfitting by Cross validation, regularization, feature selection. Overfitting can happen when there is dearth of data and you need to generate the data by permutations of the sample. You can also avoid overfitting by choosing more generalized models such as Random Forests by using bagging and boosting ensemble models.