1. **Explain the difference between supervised and unsupervised machine learning?**

   In supervised machine learning algorithms, we have to provide labelled data, for example, prediction of stock market prices, whereas in unsupervised we need not have labelled data, for example, classification of emails into spam and non-spam.

2. **Explain the difference between KNN and k.means clustering?**

   K-Nearest Neighbours is a supervised machine learning algorithm where we need to provide the labelled data to the model it then classifies the points based on the distance of the point from the nearest points.

   Whereas, on the other hand, K-Means clustering is an unsupervised machine learning algorithm thus we need to provide the model with unlabelled data and this algorithm classifies points into clusters based on the mean of the distances between different points

3. **What is the difference between classification and regression?**

   Classification is used to produce discrete results, classification is used to classify data into some specific categories .for example classifying e-mails into spam and non-spam categories.

   Whereas, We use regression analysis when we are dealing with continuous data, for example predicting stock prices at a certain point of time.

4. **How to ensure that your model is not overfitting?**

   Keep the design of the model simple. Try to reduce the noise in the model by considering fewer variables and parameters.

   Cross-validation techniques such as K-folds cross validation help us keep overfitting under control.

   Regularization techniques such as LASSO help in avoiding overfitting by penalizing certain parameters if they are likely to cause overfitting.

5. **What is meant by 'Training set' and 'Test Set'?**

   We split the given data set into two different sections namely,'Training set' and 'Test Set'.

'Training set' is the portion of the dataset used to train the model.

'Testing set' is the portion of the dataset used to test the trained model.

6. **List the main advantage of Navie Bayes?**

   A Naive Bayes classifier converges very quickly as compared to other models like logistic regression. As a result, we need less training data in case of naive Bayes classifier .

7. **Explain Ensemble learning.**

   In ensemble learning, many base models like classifiers and regressors are generated and combined together so that they give better results. It is used when we build component classifiers that are accurate and independent. There are sequential as well as parallel ensemble methods.

8. **Explain dimension reduction in machine learning.**

   Dimension Reduction is the process of reducing the size of the feature matrix. We try to reduce the number of columns so that we get a better feature set either by combining columns or by removing extra variables.

9. **What should you do when your model is suffering from low bias and high variance?**

   When the model's predicted value is very close to the actual value the condition is known as low bias. In this condition, we can use bagging algorithms like random forest regressor.

10. **Explain differences between random forest and gradient boosting algorithm.**

    Random forest uses bagging techniques whereas GBM uses boosting techniques.

    Random forests mainly try to reduce variance and GBM reduces both bias and variance of a model