1)R-squared or Residual Sum of Squares (RSS) which one of these two is a better measure of goodness of fit model in regression and why?

R-squared is generally a better measure of the goodness of fit for a regression mode. R-squared measures the extent to which changes in the dependent variable can be predicted by changes in the independent variable(s). Higher R-squared values indicate a better fit of the regression model to the data.

Residual Sum of Squares (RSS) measures the difference between the observed values of the dependent variable and the predicted values by the model. It represents the sum of the squared differences between the actual and predicted values of the dependent variable.
R-squared is a standardized measure and ranges from 0 to 1, making it easy to compare the fit of different models. In contrast, the magnitude of the RSS value depends on the scale of the dependent variable and can't be easily compared across models.

2) What are TSS (Total Sum of Squares), ESS (Explained Sum of Squares) and RSS (Residual Sum of Squares) in regression. Also mention the equation relating these three metrics with each other.

TSS = $\sum(Yi - \bar{Y})^2$, where Yi is the actual value of the response variable for observation i, and $\bar{Y}$ is the mean of the response variable.

RSS = $\sum(Yi - \hat{Y}i)^2$, which is the sum of squared differences between the actual and predicted values of the response variable.

ESS = $\sum(\hat{Y}i - \bar{Y})^2$, where $\hat{Y}i$ is the predicted value of the response variable for observation i.

TSS=ESS+RSS

3)What is the need of regularization in machine learning?

It is a technique used in ML to reduce the overfitting and underfitting in the data models.

A high biased models pays little attention to train data, and hence ends up with errors on both train data and test data.
Whereas high variance models pay close attention to train data, and hence works well with train data, but has high errors in the test data.

There should be a balance between the two where the model works in an optimum way, for which the technique called regularisation is used.

4)What is Gini–impurity index?

The Gini Index is also known as Gini impurity. It is a measure of how mixed or impure a dataset is. The Gini impurity ranges between 0 and 1, where 0 represents a pure dataset and 1 represents a completely impure dataset.

Gini impurity = $1-\sum(p(i)^2$

5)Are unregularized decision-trees prone to overfitting? If yes, why?

Decision trees, by their very nature, are prone to overfitting, especially when they are deep. Overfitting occurs when a model captures noise or fluctuations in the training data that do not represent the underlying data distribution. n the context of decision trees, overfitting can mean creating too many branches based on outliers or anomalies in the training data.

6)What is an ensemble technique in machine learning?

Ensemble learning is a machine learning technique that enhances accuracy and resilience in forecasting by merging predictions from multiple models. It minimizes errors that may exist in the individual models by combining the outputs of various models and create more precise prediction.

7)What is the difference between Bagging and Boosting techniques?

Bagging

- Combines multiple models trained on different subsets of data.
- To reduce variance by averaging out individual model error.
- Less prone to overfitting due to average mechanism.
- Improves accuracy by reducing variance.

Boosting

- Train models sequentially, focusing on the error made by the previous model.
- Reduces both bias and variance by correcting misclassifications of the previous model.
- Generally not prone to overfitting, but it can be if the number of the model or the iteration is high.
- Achieves higher accuracy by reducing both bias and variance.

8)What is out-of-bag error in random forests?

OOB (out-of-bag) score is a performance metric for a machine learning model, specifically for ensemble models such as random forests. It is calculated using the samples that are not used in the training of the model, which is called out-of-bag samples. These samples are used to provide an unbiased estimate of the model's performance, which is known as the OOB score.

9)What is K-fold cross-validation?

It involves splitting the dataset into k subsets or folds, where each fold is used as the validation set in turn while the remaining k-1 folds are used for training. This process is repeated k times, and performance metrics such as accuracy, precision, and recall are computed for each fold. By averaging these metrics, we obtain an estimate of the model's generalization performance. This method is essential for model assessment, selection, and hyperparameter tuning, offering a reliable measure of a model's effectiveness.

10)What is hyper parameter tuning in machine learning and why it is done?

Hyperparameter tuning is the process of selecting the optimal values for a machine learning model's hyperparameters. Hyperparameters are settings that control the learning process of the model, such as the learning rate, the number of neurons in a neural network, or the kernel size in a support vector machine. The goal of hyperparameter tuning is to find the values that lead to the best performance on a given task. In the context of machine learning, hyperparameters are configuration variables that are set before the training process of a model

begins. They control the learning process itself, rather than being learned from the data. Hyperparameters are often used to tune the performance of a model, and they can have a significant impact on the model's accuracy, generalization, and other metrics.

11)What issues can occur if we have a large learning rate in Gradient Descent?

Gradient descent can overfit the training data if the model is too complex or the learning rate is too high.

12)Can we use Logistic Regression for classification of Non-Linear Data? If not, why?
Non-linear problems can't be solved with logistic regression because it has a linear decision surface. Linearly separable data is rarely found in real-world scenarios.

13)Differentiate between Adaboost and Gradient Boosting.

- Model Flexibility: Gradient Boosting is more flexible and can be used with different weak learners like decision trees. AdaBoost is restricted to decision tree models.
- Overfitting Handling: Gradient Boosting has inbuilt regularization to prevent overfitting. AdaBoost can overfit if run for too many iterations.
- Performance: Gradient Boosting typically achieves better accuracy by reducing bias and variance. AdaBoost is faster and simpler to tune.

14)What is bias-variance trade off in machine learning?

A high biased models pays little attention to train data, and hence ends up with errors on both train data and test data.
Whereas high variance models pay close attention to train data, and hence works well with train data, but has high errors in the test data.
Bias represents the error due to overly simplistic assumptions in the learning algorithm. High bias can cause the model to underfit the data, leading to poor performance on both training and unseen data.
Variance, on the other hand, reflects the model's sensitivity to small fluctuations in the training data. High variance can lead to overfitting, where the model captures noise in the training data and performs poorly on new, unseen data.

15) Give short description each of Linear, RBF, Polynomial kernels used in SVM.
The linear, polynomial and RBF or Gaussian kernel are simply different in case of making the hyperplane decision boundary between the classes.
The kernel functions are used to map the original dataset (linear/nonlinear) into a higher dimensional space with view to making it linear dataset.
Usually linear and polynomial kernels are less time consuming and provides less accuracy than the rbf or Gaussian kernels.