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?

Ans: R-squared is often used when you want to understand the proportion of variance explained by the model, especially in the context of comparing different models. It can provide insights into how well the predictors collectively contribute to explaining the variation in the dependent variable.

- RSS is useful when you want to evaluate the absolute goodness of fit, focusing on the magnitude of the prediction errors. If minimizing prediction errors is a primary concern, RSS is a more appropriate choice.

In practice, both measures can be valuable. R-squared provides a high-level summary of the explanatory power of your model, while RSS gives you a detailed view of the model's prediction accuracy. The choice between them should align with your specific goals and the questions you want to answer about your regression model.

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

Ans: Formula:

Total Sum of Squares = Explained Sum of Squares + Residual Sum of Squares

The Explained sum of squares tells you how much of the variation in the dependent variable your model explained.
Explained SS = $\Sigma(Y\text{-Hat} - \text{mean of } Y)^2$

The Residual sum of squares tells you how much of the dependent variable's variation your model did not explain. It is the sum of the squared differences between the actual Y and the predicted Y:
Residual Sum of Squares = $\Sigma e2$

The Total sum of squares tells you how much variation there is in the dependent variable.
Total SS = $\Sigma(Yi - \text{mean of } Y)^2$

3] What is the need of regularization in machine learning?
Ans:
While training a machine learning model, the model can easily be overfitted or under fitted. To avoid this, we use regularization in machine learning to properly fit a model onto our test set. Regularization techniques help reduce the chance of overfitting and help us get an optimal model.

4] What is Gini–impurity index?
Ans: Gini impurity measures how often a randomly chosen element of a set would be incorrectly labeled if it were labeled randomly and independently according to the distribution of labels in the set. It reaches its minimum (zero) when all cases in the node fall into a single target category.

5] Are unregularized decision-trees prone to overfitting? If yes, why?
Ans:
decision trees can be prone to overfitting, which occurs when the tree is too complex and fits the training data too well, but performs poorly on new data. To overcome overfitting, techniques such as pruning and ensemble methods can be used.

6] What is an ensemble technique in machine learning?
Ans:
Ensemble learning is a machine learning technique that enhances accuracy and resilience in forecasting by merging predictions from multiple models. It aims to mitigate errors or biases that may exist in individual models by leveraging the collective intelligence of the ensemble

7] What is the difference between Bagging and Boosting techniques?
Ans:
Bagging is the simplest way of combining predictions that belong to the same type while Boosting is a way of combining predictions that belong to the different types. Bagging aims to decrease variance, not bias while Boosting aims to decrease bias, not variance.

8**] What is out-of-bag error in random forests?**
Ans:
Out-of-bag (OOB) error, also called out-of-bag estimate, is a method of measuring the prediction error of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating (bagging).

9] What is K-fold cross-validation?
Ans:
K-fold cross-validation is a technique for evaluating predictive models. The dataset is divided into k subsets or folds. The model is trained and evaluated k times, using a different fold as the validation set each time. Performance metrics from each fold are averaged to estimate the model's generalization performance.

10] What is hyper parameter tuning in machine learning and why it is done?
Ans:
When you're training machine learning models, each dataset and model needs a different set of hyperparameters, which are a kind of variable. The only way to determine these is through multiple experiments, where you pick a set of hyperparameters and run them through your model. This is called hyperparameter tuning.

11] What issues can occur if we have a large learning rate in Gradient Descent?
Ans:
Too large learning rate can lead to exploding or oscillating performance over the training epochs and to a lower final performance, the algorithm may overshoot the minimum, and if it is too low, the algorithm may take too long to converge.

12] Can we use Logistic Regression for classification of Non-Linear Data? If not, why?
Ans:
Logistic regression has traditionally been used to come up with a hyperplane that separates the feature space into classes. But if we suspect that the decision boundary is nonlinear we may get better results by attempting some nonlinear functional forms for the logit function.

13] Differentiate between Adaboost and Gradient Boosting.
Ans:
AdaBoost is the first designed boosting algorithm with a particular loss function. On the other hand, Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem. This makes Gradient Boosting more flexible than AdaBoost.
Adaptive Boosting or AdaBoost, it minimises the exponential loss function that can make the algorithm sensitive to the outliers. With Gradient Boosting, any differentiable loss function can be utilised. Gradient Boosting algorithm is more robust to outliers than AdaBoost.

AdaBoost minimises loss function related to any classification error and is best used with weak learners. The method was mainly designed for binary classification problems and can be utilised to boost the performance of decision trees. Gradient Boosting is used to solve the differentiable loss function problem. The technique can be used for both classification and regression problems.

In the case of Gradient Boosting, the shortcomings of the existing weak learners can be identified by gradients and with AdaBoost, it can be identified by high-weight data points.

14] What is bias-variance trade off in machine learning?
Ans:
In machine learning, as you try to minimize one component of the error (e.g., bias), the other component (e.g., variance) tends to increase, and vice versa. Finding the right balance of bias and variance is key to creating an effective and accurate model. This is called the bias-variance tradeoff.

15] Give short description each of Linear, RBF, Polynomial kernels used in SVM.

Ans:

Linear Kernel is the simplest and most commonly used kernel function, and it defines the dot product between the input vectors in the original feature space.

Radial basis function (RBF) kernel is a nonlinear kernel function that maps the input data into a higher-dimensional feature space using a Gaussian function.

Polynomial Kernel is a nonlinear kernel function that employs polynomial functions to transfer the input data into a higher-dimensional feature space.