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?

Answer - R-Squared is better measure of goodness of fit model in regression because it provides an overall measure of the proportion of variance in the dependent variable that is explained by the model, whereas RSS only measures the magnitude of the residuals.

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.

Answer - TSS is the sum of square of difference of each data point from the mean value of all the values of target variable.

ESS is the sum of the differences between the predicted value and the mean of the dependent variable. In other words, it describes how well our line fits the data.

RSS measures the level of variance in the error term, or residuals, of a regression model. The smaller the residual sum of squares, the better your model fits your data; the greater the residual sum of squares, the poorer your model fits your data.

(TSS)= (ESS) + (RSS).

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

Answer - 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. It helps to reduce the over-complexity of the machine learning model.

4. What is Gini–impurity index?

Answer- Gini Impurity is a measure used in decision tree algorithims to quantify a dataset's impurity level or disorder. It determines how well a decision tree was split.

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

Answer - Yes, unregularized decision trees prone to overfitting because they can handle both numerical and categorical data and can easily interpret the results. Decision trees can also suffer from overfitting, which means that they learn too much from the training data and fail to generalize well to new data.

6. What is an ensemble technique in machine learning?

Answer - 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?

Answer - In bagging, models are trained independently in parallel on different random subsets of the data. Whereas in boosting, models are trained sequentially, with each model learning from the

errors of the previous one. 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?

Answer - 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.

9. What is K-fold cross-validation?

Answer - 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.

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

Answer - Hyperparameter tuning is an essential part of controlling the behavior of a machine learning model. If we don't correctly tune our hyperparameters, our estimated model parameters produce suboptimal results, as they don't minimize the loss function. This means our model makes more errors.

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

Answer - The choice of learning rate can significantly impact the performance of gradient descent. If the learning rate is too high, 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?

Answer – 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.

Answer:-
AdaBoost

In AdaBoost, shift is done by up-weighting observations that were misclassified before.
In AdaBoost "shortcomings" are identified by high-weight data point.
AdaBoost is considered as a special case of Gradient boost in terms of loss function, in which exponential losses.

GradientBoost

Gradient boost identifies difficult observations by large residuals computed in the previous iterations.
In Gradientboost "shortcomings" are identified by gradients.
Gradient boost further dissect error components to bring in more explanation.
Concepts of gradients are more general in nature.

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

Answer - Bias-Variance Tradeoff is crucial in machine learning because it directly impacts a model's predictive performance. A model with high bias will consistently produce predictions that are far from the actual values, while a model with high variance will produce widely varying predictions for different training datasets.

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

Answer –

Linear Kernels - A linear kernel is a type of kernel function used in machine learning, including in SVMs (Support Vector Machines). It is the simplest and most commonly used kernel function, and it defines the dot product between the input vectors in the original feature space.

RBF Kernels - In machine learning, the radial basis function kernel, or RBF kernel, is a popular kernel function used in various kernelized learning algorithms. In particular, it is commonly used in support vector machine classification.

Polynomial Kernels - In machine learning, the polynomial kernel is a kernel function commonly used with support vector machines (SVMs) and other kernelized models, that represents the similarity of vectors (training samples) in a feature space over polynomials of the original variables, allowing learning of non-linear models.