MACHINE LEARNING

Q1 to Q15 are subjective answer type questions, Answer them briefly.

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 generally a better measure of the goodness of fit for a regression model than the residual sum of squares (RSS). It is dimensionless and ranges from 0 to 1, where a value closer to 1 indicates a better fit. The RSS, on the other hand, is the sum of the squared differences between the observed actual outcomes and the outcomes predicted by the regression model.
The reason why R^2 is often preferred over RSS as a measure of goodness of fit is due to its standardized nature.

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: The sum of squares total (SST) or the total sum of squares (TSS) is the sum of squared differences between the observed dependent variables and the overall mean.
The sum of squares due to regression (SSR) or explained sum of squares (ESS) is the sum of the differences between the predicted value and the mean of the dependent variable.
The sum of squares error (SSE) or residual sum of squares (RSS, where residual means remaining or unexplained) is the difference between the observed and predicted values.
the Relationship between SSR, SSE, and SST
Mathematically, **SST** = **SSR** + **SSE,**

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

Answer: The primary goal of regularization is to reduce the model's complexity to make it more generalizable to new data, thus improving its performance on unseen datasets.
Regularization is techniques used to calibrate **machine learning models to avoid overfitting or underfitting**.
4. What is Gini–impurity index?

Answer: Gini Impurity is a measurement used to build Decision Trees to determine how

the features of a dataset should split nodes to form the tree.

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

Answer: Decision trees are prone to overfitting, especially when a tree is particularly deep. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions. A decision tree will always overfit the training data if we allow it to grow to its max depth.

6. What is an ensemble technique in machine learning?

Answer: Ensemble learning refers to a machine learning approach where several models are trained to address a common problem, and their predictions are combined to enhance the overall performance.

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

Answer: The difference between Bagging and Boosting techniques is Bagging (Bootstrap Aggregating) reduces variance by averaging multiple models, while boosting reduces bias by combining weak learners sequentially to form a strong learner.

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 (bagging).

9. What is K-fold cross-validation?

Answer: In K Fold cross validation, the data is divided into k subsets. Now the holdout method is repeated k times, such that each time, one of the k subsets is used as the test set/ validation set and the other k-1 subsets are put together to form a training set.

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

Answer: Hyperparameters directly control model structure, function, and performance. Hyperparameter tuning allows data scientists to tweak model performance for optimal results. This process is an essential part of machine learning, and choosing appropriate

hyperparameter values is crucial for success.

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

Answer: When the learning rate is too large, gradient descent can suffer from divergence. This means that weights increase exponentially, resulting in exploding gradients which can cause problems such as instabilities and overly high loss values.

12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?

Answer: Logistic regression is simple and easy to implement, but it also has some drawbacks. One of them is that it assumes a linear relationship between the input features and the output. This means that it cannot capture the complexity and non-linearity of the data.

13. Differentiate between Adaboost and Gradient Boosting.

Answer: The most significant difference is that gradient boosting minimizes a loss function like MSE or log loss while AdaBoost focuses on instances with high error by adjusting their sample weights adaptively.

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

Answer: The bias-variance trade off in machine learning is known as the difference between the prediction of the values by the Machine Learning model and the correct value.

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

Answer: 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.

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.

The linear kernel produces a decision boundary that is a hyperplane in the feature

space. This hyperplane separates data points from different classes in a linear fashion.