MACHINE LEARNING

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 better measure of goodness of fit model because it is a statistical measure in a regression model that determine the proportion of variance in the dependent variable that can be explained by the independent variable.

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.   Total sum of squares is a quantity that appears as part of standard way of presenting results of such analyses.
Explained sum of squares is the sum of squares measures the deviation of data points away from the mean value.
Residual sum of squares is the sum of squares to measure of the discrepancy between the data and an estimation model such as linear regression.

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

Ans-   Regularization refers to technique that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent over-fitting or underfitting.

4. What is Gini–impurity index?

Ans   The Gini index or Gini impurity is calculated by subtracting the sum of the squared probabilities of each class from one.

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

Ans   Decision trees are non-parametric supervised machine learning approach for classification and regression tasks.

6. What is an ensemble technique in machine learning?

Ans.   Ensemble technique use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone.

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

Ans   Bagging is a method od merging the same type of predictions while Boosting is a method of merging different types of predictions.

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

Ans   The out of bag error is the average error for each calculated using predictions from the trees that do not contain in their respective bootstarp sample.

9. What is K-fold cross-validation?

Ans   K- fold cross validation is when the dataset is split into a K no. of folds and is used to evaluate the model's ability when given new data.

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

Ans   In machine learning hyper parameter optimization or tuning is the problem of choosing a set of optimal hyper parameter for a learning algorithm. It is done for maximize the model's performance minimizing a predefined loss function to produce better results with fewer errors.

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

Ans   If learning rate is too large gradient descent can overshoot the minimum. It may fail to converge and even diverge.

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

Ans   No, because it is used to come up with a hyperplane in feature space to separate observations that belong to a class from all other observations that do not belong to that class.

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 approx.. solutions to the additive modelling problem.

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

Ans.   The bias-variance tradeoff is the property of model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameter.

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

Ans