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?

ANS: R -squared & Residual sum of Squares (RSS) both are measure of goodness of fit in regression but they both are captured different aspects of the model performance.

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: TSS (Total Sum of Squares): Which measures how much variation there is the observed data.

ESS (Explained Sum of Squares): The square of the deviations of the predicted values from the mean value of response variable.

RSS (Residual Sum of Squares): The residual sum of squares measures the level of variance in the error term, or residuals of a regression model.

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

ANS: The need of regularization in machine learning is to minimize the adjusted loss function & prevent overfitting or underfitting.

4. What is Gini–impurity index?

ANS: Gini-impurity is a measurement used to build a decision tree to determine how to features of a dataset should split nodes to form the tree.

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

ANS: Yes, unlegalized decision -trees prone to overfitting because, it can't help to reduce the complexity and can't avoid overfitting.

6. What is an ensemble technique in machine learning?

ANS: An ensemble technique in machine learning are the techniques that create multiple models & combined them to improved results.

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

ANS: Bagging attempts how to over-fitting issue tackle & Boosting techniques reduce Bias.

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 bootstrap sample.

9. What is K-fold cross-validation?

ANS: K-fold cross-validation is a technique for evaluating predictive models.

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

ANS: Hyper parameter directly controls model structure, function, performance. Hyper parameter tuning allows data scientists to tweak model performance for optimal results.

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

ANS: If the learning rate is large then the algorithm may overshoot the minimum.

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

ANS: If the suspect that the decision boundary is nonlinear then 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 & Gradient boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem.

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

ANS: Describes the relationship between a model's complexity, the accuracy of its predictions, and how well it can make predictions on previously unseen data that were not used to train the model.

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

ANS: Linear SVM: When the data is perfectly linearly separable only then we can use Linear SVM.

RBF SVM: RBF SVM works by mapping the input data into a higher-dimensional feature space, where the classes can be separated by a hyperplane.

Polynomial kernels used in SVM: In machine learning, the polynomial kernel is a kernel function commonly used with support vector machines & other kernelized models, that represents the similarity of vectors in a feature space over polynomials of the original variables, allowing learning of non-linear models.