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 (R^2):** This is the proportion of the variance in the dependent variable that is predictable from the independent variable(s). It provides a measure of how well the model explains the variability of the outcome data.

Residual Sum of Squares (RSS): This is the sum of the squared differences between the observed and predicted values. It measures the discrepancy between the data and the estimation model.

Which is Better: -

R-squared is generally considered a better measure of goodness of fit because it provides a normalized measure of how well the regression model fits the data. RSS can be useful, but its value depends on the scale of the data and isn't easily interpretable in the context of model fit quality.

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 (TSS):** Measures the total variance in the observed data and is calculated as the sum of the squares of the differences between each observation and the overall mean of the observations.

Explained Sum of Squares (ESS): Measures the variance explained by the regression model and is calculated as the sum of the squares of the differences between the predicted values and the overall mean of the observations.

Residual Sum of Squares (RSS): Measures the variance not explained by the model and is calculated as the sum of the squares of the differences between the observed values and the predicted values.

Equation:

TSS=ESS+RSS

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

Ans: **Reduce model complexity:** By penalizing large coefficients, regularization discourages the model from fitting the noise in the training data.

Improve generalization: It helps the model to generalize better to unseen data by avoiding overfitting to the training data.

4. What is Gini–impurity index?

Ans :The Gini-impurity index is a measure used in decision trees to evaluate the purity of a node. It measures the likelihood of an incorrect classification of a randomly chosen element if it was randomly labeled according to the distribution of labels in the node. It is calculated as: Gini=1−∑i=1npi2Gini = 1 - \sum_{i=1}^{n} p_i^2Gini=1−∑i=1npi2 where pip_ipi is the probability of an element being classified into class iii.

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

Ans :Yes, unregularized decision trees are prone to overfitting because they can grow very deep and create complex models that perfectly fit the training data, including its noise and outliers. This results in poor performance on new, unseen data.

6. What is an ensemble technique in machine learning?

Ans :An ensemble technique combines multiple machine learning models to produce a single model that performs better than any of the individual models. It leverages the strengths of different models and mitigates their weaknesses.

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

Ans : **Bagging (Bootstrap Aggregating):** It involves training multiple models independently on different random subsets of the data and then averaging their predictions. It reduces variance and helps prevent overfitting.

Boosting: It involves training models sequentially, each model correcting the errors of the previous ones. It focuses on reducing bias and improving model accuracy

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

Ans: Out-of-bag (OOB) error is an estimate of the model's prediction error for unseen data. In random forests, each tree is trained on a bootstrap sample of the data, and the OOB error is calculated using the observations that were not included in the bootstrap sample (out-of-bag observations).

9. What is K-fold cross-validation?

Ans: K-fold cross-validation is a technique to evaluate the performance of a model. It involves partitioning the data into K equally sized folds, training the model on K-1 folds, and testing it on the remaining fold. This process is repeated K times, each time with a different fold as the test set. The final performance is averaged over the K iterations.

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

Ans: Hyperparameter tuning involves selecting the optimal set of hyperparameters for a machine learning model. It is done to improve model performance, as hyperparameters significantly affect the learning process and final accuracy of the model.

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

Ans: A large learning rate in Gradient Descent can cause the algorithm to overshoot the optimal solution, leading to divergence or oscillations around the minimum rather than convergence. It may result in failing to find the minimum error.

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

Ans: Logistic Regression is inherently a linear classifier, meaning it is not suitable for classifying non-linear data unless the data is transformed using techniques like polynomial features or kernel methods to introduce non-linearity.

13. Differentiate between Adaboost and Gradient Boosting.

Ans: **Adaboost:** Focuses on improving the accuracy of the weak classifiers by adjusting the weights of incorrectly classified instances so that subsequent classifiers focus more on difficult cases.

Gradient Boosting: Builds the model sequentially by optimizing the loss function. Each new model tries to correct the errors of the combined ensemble of previous models by fitting to the residual errors.

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

Ans: **Bias:** Error due to overly simplistic models that underfit the data.

Variance: Error due to overly complex models that overfit the data. The goal is to find a model with an optimal level of complexity that minimizes both bias and variance for the best generalization to unseen data.

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

Ans: **Linear Kernel:** The simplest kernel used when the data is linearly separable. It computes the dot product between the input vectors.

RBF (Radial Basis Function) Kernel: A popular non-linear kernel that measures the distance between the input vectors in a feature space transformed by the Gaussian function. It is useful for non-linear classification tasks.

Polynomial Kernel: Computes the similarity of vectors in a feature space over polynomials of the original variables, allowing the SVM to fit more complex, non-linear boundaries.

############################## FINISHED ############################