MACHINE LEARNING

ASSIGNMENT NO - 5

Q1. R-squared is a better measure of goodness of fit in regression compared to Residual Sum of Squares (RSS). R-squared is a statistical measure that represents the proportion of the variance in the dependent variable that is explained by the independent variables in the model. It ranges from 0 to 1, with a value of 1 indicating a perfect fit. R-squared provides an overall summary of how well the model fits the data, making it a useful tool for comparing multiple regression models.

On the other hand, Residual Sum of Squares (RSS) is the sum of the squared differences between the observed and predicted values of the dependent variable. A lower value of RSS indicates a better fit of the model. However, RSS does not provide information about the proportion of variance explained by the independent variables and is therefore not as useful for comparing different models.

In conclusion, R-squared is a more comprehensive measure of goodness of fit in regression compared to RSS.

Q2. In regression analysis, the Total Sum of Squares (TSS) is the sum of the squared differences between the observed values of the dependent variable and its mean. It represents the total variation in the dependent variable.

The Explained Sum of Squares (ESS) is the sum of the squared differences between the predicted values of the dependent variable and its mean. It represents the variation in the dependent variable that can be explained by the independent variables in the regression model.

The Residual Sum of Squares (RSS) is the sum of the squared differences between the observed and predicted values of the dependent variable. It represents the variation in the dependent variable that cannot be explained by the independent variables in the regression model.

The following equation relates these three metrics with each other:

TSS = ESS + RSS

TSS represents the total variation in the dependent variable, and ESS and RSS represent the explained and residual variation, respectively.

Q3. Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting. Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

Q4. The Gini Impurity of a dataset is a number between 0-0.5, which indicates the likelihood of new, random data being misclassified if it were given a random class label according to the class distribution in the dataset.

Q5. Yes, unregularized decision trees are prone to overfitting. Overfitting occurs when a model fits the training data too closely, memorizing the noise and random variations in the data, and as a result, it performs poorly on unseen data. Unregularized decision trees can be highly prone to overfitting, as they have no constraints on their growth and can become overly complex, fitting the training data too closely and not generalizing well to new data.

Q6. Ensemble methods are techniques that aim at improving the accuracy of results in models by combining multiple models instead of using a single model. The combined models increase the

accuracy of the results significantly. This has boosted the popularity of ensemble methods in machine learning.

Q7. The main difference between Bagging and Boosting is that Bagging focuses on reducing the variance of the model and improving its stability, while Boosting focuses on reducing the bias of the model and improving its accuracy. Another key difference is that Bagging trains multiple instances of the same base model on different random subsets of the training data, while Boosting trains multiple instances of the same base model with a focus on correcting the mistakes made by the previous models.

Q8. The out-of-bag (OOB) error is the average error for each calculated using predictions from the trees that do not contain in their respective bootstrap sample. This allows the Random Forest Classifier to be fit and validated whilst being trained.

Q9. K-fold cross-validation is a technique for evaluating the performance of a machine learning model by dividing the data into K parts and using each part for validation once, with the performance of the model being averaged over all K iterations. It provides a more robust and less-variant estimate of the model's performance than a single validation set.

Q10. Hyperparameter tuning is the process of selecting the best set of hyperparameters for a machine learning model to optimize its performance. It is important because the performance of a model is strongly influenced by the choice of hyperparameters and selecting the optimal hyperparameters can lead to improved performance. Hyperparameter tuning can be done using various techniques such as grid search, random search optimization.

Q11. A large learning rate in gradient descent can result in oscillations, divergence, slow convergence, and non-convergence, all of which can prevent the algorithm from finding the optimal solution or cause it to converge to a suboptimal solution .

Q12. Yes, Logistic Regression can be used for classification of non-linear data, but it may not always produce the best results. Logistic Regression is a linear model, meaning it makes a prediction based on a linear combination of input features. When the relationship between the input features and target variable is non-linear, a linear model may not accurately capture the complex relationship. In such cases, other non-linear models such as decision trees, random forests, or support vector machines (SVMs) might perform better.

Q13. AdaBoost (Adaptive Boosting) and Gradient Boosting are both ensemble learning methods used for classification and regression problems. However, they have some key differences:

- Conceptual difference: AdaBoost is based on the concept of boosting weak learners, while Gradient Boosting is based on the idea of boosting the loss function.
- Algorithm difference: AdaBoost assigns weights to each data point in the training set based on the performance of previous learners and then trains the next learner to focus more on the misclassified data points. In Gradient Boosting, weak learners are added one by one to improve the overall prediction by focusing on the residual errors made by the previous learner.
- Complexity: AdaBoost is a relatively simple algorithm and computationally fast, while Gradient Boosting is more complex and computationally slower.
- Handling of Outliers: AdaBoost is sensitive to outliers, while Gradient Boosting handles outliers better as it minimizes the overall loss function.

- Regularization: Gradient Boosting has the option of adding regularization to prevent overfitting, while AdaBoost does not have this option.

In summary, both methods can be useful, but Gradient Boosting is typically more powerful and flexible and is widely used in Kaggle competitions and industry.

Q14. The bias-variance trade-off is a fundamental concept in machine learning that refers to the balance between the model's ability to fit the training data well (low bias) and its ability to generalize to new, unseen data (low variance). A model with high bias will over-simplify the training data and produce under-fit models that don't capture the underlying pattern. On the other hand, a model with high variance will over-fit the training data and produce over-complex models that perform well on the training data but poorly on unseen data. The goal of machine learning is to find the right balance between bias and variance, so that the model can generalize well to unseen data while still fitting the training data .

Q15. Support Vector Machines (SVM) are a type of machine learning algorithm that can be used for both classification and regression tasks. In SVM, the goal is to find the hyperplane that best separates the data into different classes or predicts the target variable. The choice of the kernel function is an important aspect of SVM that determines the form of the hyperplane.

- Linear kernel: The linear kernel is the simplest kernel function and models the decision boundary as a straight line or hyperplane. This kernel is used when the data is linearly separable.
- RBF (Radial Basis Function) kernel: The RBF kernel is a non-linear kernel function that models the decision boundary as an ellipse or a circle. It maps the input data into a higher dimensional space where the data may become linearly separable.
- Polynomial kernel: The polynomial kernel is another non-linear kernel function that models the decision boundary as a polynomial. It can model more complex decision boundaries than the linear and RBF kernels, but can also lead to over-fitting if the degree of the polynomial is too high.