Machine Learning Assignment-5

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?

R-Squared is generally considered a better measure of goodness of fit in regression compared to Residual Sum of Squares.

R-Squared that represents the amount variation in the dependent variable that is explained by the independent variable. R-squared useful for comparing the different models or assessing the overall fit of a model.

RSS represents the sum of squared differences between observed and predicted values. RSS is useful measure of overall fit of the model and does not provide the standardized measure.

Overall, R-Squared is preferred for assessing goodness of fit because it is not only measures the amount of variance explained by the model but also provide the standard metric for comparison.

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.

Total Sum of Squares (TSS), Explained Sum of Squares (ESS), and Residual Sum of Squares (RSS) are metrics used to assess the variance in the dependent variable in Regression analysis.

1.Total Sum of Squares (TSS): sum of the squared differences between each observed dependent variable value and the overall mean of the dependent variable.

$TSS = \sum(Y_i - \bar{Y})^2$, where $\bar{Y}$ is the mean of Y.

2. Explained Sum of Squares (ESS): sum of the squared differences between the predicted values (obtained from the regression model) and the overall mean of the dependent variable.

: $ESS = \sum(\hat{Y}_i - \bar{Y})^2$, where $\hat{Y}_i$ is the predicted value for observation i.

3. Residual Sum of Squares (RSS): sum of the squared differences between each observed dependent variable value and its corresponding predicted value from the regression model.

$RSS = \sum(Y_i - \hat{Y}_i)^2$, where Y_i is the observed value for observation i,

and $\hat{Y}_i$ is the predicted value for observation i.

TSS=ESS+RSS

This equation signifies that the total variability in the dependent variable (TSS) can be decomposed into the explained variability (ESS) and the unexplained variability (RSS) by the regression model.

R-Squred:ESS/TSS , R-squared represents the proportion of the total variability explained by Regression model.

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

Regularization is a technique to prevent the overfitting and improve the generalization of models.

Regularization method (Ridge regression) helps to stabilize the estimates of the regression coefficients when multicollinearity is present.

4. What is Gini–impurity index?

Gini–impurity index is a measure of the impurity or disorder in a node and used it decision algorithms to take a decision about how to split the data at each node during tree-building process.

Gini–impurity index used in decision tree algorithms to evaluate how well a given node separates the data into different classes. It is used as criteria to split the node during construction of decision trees in classification problems.

Gini–impurity takes values between 0 and 1.0 indicates that node contains instances only from one class (perfect pure) and 1 indicates that instances are evenly distributed across different classes (maximum impurity).

When we constructing the decision tree, the algorithm aims to minimize the Gini–impurity in each node.

Gini–impurity index is a measure of the impurity or disorder in a node and used it decision algorithms to take a decision about how to split the data at each node during.

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

6. What is an ensemble technique in machine learning?

Ensemble is a technique in machine learning that refers the combination of multiple individual's models to create a stronger model instead of relying on the prediction of single model. ensemble methods leverage the diversity of multiple models to improve overall performance and generalization.

Ensemble is a popular approach to enhance the model predictive accuracy and reduce overfitting.

Popular ensemble methods include Random Forest, Gradient Boosting , AdaBoost, etc..

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

Bagging and Boosting are both ensemble techniques used in machine learning to improve the performance of predictive models.

Bagging	Boosting
Independent and parallel training of multiple models on different bootstrap samples	Sequential training with a focus on correcting errors made by previous models
Equal weight is given to all instances in the training data	Instances are assigned weights during training, with higher emphasis on misclassified instances
Averaging or majority voting is used to combine predictions.	Predictions are combined through a weighted sum of individual model predictions

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

In a Random Forest ensemble, the out-of-bag (OOB) error is a measure of the model's performance on unseen data, and it is calculated using the data points that were not included in the bootstrap samples used to train each individual tree.

The key advantage of using the out-of-bag error is that it provides an unbiased estimate of the model's performance on unseen data without the need for a separate validation set.it is useful when we working with limited data.

9. What is K-fold cross-validation?

K-fold cross-validation is a popular technique used in Machine learning to assess the performance and generalization ability of the model.

The basic idea behind K-fold cross-validation is to partition the original dataset into K equally sized folds (or subsets).

The model is trained and evaluated K times, each time using a different fold as the test set and the remaining K-1 folds as the training set.

Below steps involved in K-fold cross-validation:

a. The original dataset is divided into K equally sized folds.
b. The model is trained K times, with each iteration using a different fold as the test set and the remaining K-1 folds as the training set. For each iteration, the model is trained on the training set and evaluated on the corresponding test set
c. Each iteration recorded with performance metrics
d. The recorded performance metrics are averaged to obtain a single performance estimate for the model

K-fold cross-validation helps in addressing issues related to variability in the training and testing data splits, providing a more robust evaluation of the model's performance.

K-fold cross-validation is valuable for detecting overfitting and understanding how well the model generalizes to different subsets of the data

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

a. Hyperparameter tuning, also known as hyperparameter optimization or model selection, is the process of selecting the best set of hyperparameters for a machine learning model.

b. Hyperparameters are external configuration settings that are not learned from the data but are set before the training process begins.

c. They influence the behaviour of the training algorithm and, consequently, the performance of the model

why it is done:

==========

a. Hyperparameter tuning aims to find the combination that results in the best overall performance on unseen data
b. Proper hyperparameter tuning helps strike a balance between overfitting and underfitting.
c. Hyperparameter tuning contributes to the model's ability to capture underlying patterns in the data
d. hyperparameter tuning is a crucial step in the model development process, aiming to find the best configuration of hyperparameters to achieve optimal model performance and generalization

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

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

a. Logistic Regression is a linear model that is typically used for binary classification problems. It models the relationship between the independent variables and the probability of a particular outcome using a logistic function.

b. Logistic Regression may not perform well on non-linear data or datasets with complex decision boundaries. The reason is that the logistic function used in Logistic Regression introduces linearity into the model, and it assumes that the relationship between the input features and the log-odds of the output is linear.

c. If the underlying patterns in the data are non-linear, Logistic Regression may struggle to capture those patterns effectively and it will result the poor classification performance and a model that fails to generalize well to unseen data.

13. Differentiate between Adaboost and Gradient Boosting.

AdaBoost (Adaptive Boosting) and Gradient Boosting are both ensemble learning techniques used for improving the performance of base models.

AdaBoost (Adaptive Boosting)	Gradient Boosting
Adjusts data point weights based on classification accuracy	Gradient Boosting fits models to the residual errors

Assigns weights to the models during the final prediction based on their accuracy	Uses a shrinkage parameter to control the contribution of each model
AdaBoost uses an exponential loss function to update the weights of misclassified data points, placing more emphasis on correcting the errors	Minimizes a loss function using gradient descent. The optimization process involves adjusting the parameters of the models in the direction that reduces the loss.
Weak learners are trained sequentially, with each subsequent model adjusting its focus to correct the errors made by the previous models	Weak learners are trained sequentially. However, in Gradient Boosting, each model focuses on reducing the residual errors

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

a. Bias:

The bias is defined as the difference between the ML model's prediction of the values and the correct value. Biasing causes a substantial inaccuracy in both training and testing data. To prevent the problem of underfitting, it is advised that an algorithm be low biassed at all times.

b. Variance:

The variance of the model is the variability of model prediction for a particular data point, which tells us about the dispersion of our data. The model with large variance has a very complicated fit to the training data and so is unable to fit correctly on new data. As a result, while such models perform exceptionally well on training data, they have substantial error rates on test data.

The model basically tries to memorize the patterns in the training data and performs poorly when unseen data is presented. When a model has a large variance, this is referred to as Overfitting of Data.

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

Support Vector Machines (SVMs) are powerful supervised learning algorithms commonly used for classification and regression tasks. SVMs use kernel functions to transform input data into higher-dimensional spaces, allowing them to find non-linear decision boundaries.

a. Linear Kernel: The linear kernel is the simplest and most used kernel in SVM. It represents a linear relationship between the input features and is suitable for linearly separable data.
b. RBF (Radial Basis Function) Kernel: Known as the Gaussian kernel, is a popular choice for handling non-linear relationships in SVM. It introduces a non-linearity by mapping the input features into an infinite-dimensional space. Effective for capturing complex non-linear relationships.
c. Polynomial Kernel: The polynomial kernel introduces non-linearity by mapping the input features into a higher-dimensional space using polynomial functions. It is characterized by two parameters: the degree of the polynomial and an optional coefficient. It is suitable for capturing non-linear patterns.