Worksheet_Set_2

Q1:- 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 generally a better measure of the goodness of fit for a regression model than the residual sum of squares(RSS).

Q2:- 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:- The sum of squares is a statistical measure of variability. It indicates the dispersion of data points around the mean and how much the dependent variable deviates from the predicted values in regression analysis.

TSS:- The Total sum of Squares (TSS) is the sum of squared differences between the observed dependent variables and the overall mean. Think of it as the dispersion of the observed variables around the Mean—similar to the Variance in descriptive statistics.

ESS:- The Explained Sum Of Squares (ESS) is the sum of the differences between the predicted value and the mean of the dependent variable. In other words, it describes how well our line fits the data.

RSS:- The Residual Sum Of Squares(RSS) where residual means remaining or unexplained is the difference between the *observed* and *predicted* values.

Q3:- What is the need of regularization in machine learning?

Ans:- The primary goal of regularization is to reduce the model's complexity to make it more generalizable to new data, thus improving its performance on unseen datasets.

Q4:- What is Gini–impurity index?

Ans:- Gini impurity is a measure used in decision tree algorithms to quantify a dataset's impurity level or disorder.

Q5:- Are unregularized decision-trees prone to overfitting? If yes, why?

Ans:- Decision Trees are prone to over-fitting. A decision tree will always overfit the training data if we allow it to grow to its max depth. Overfitting in decision trees occurs when the tree becomes too complex and captures the noise in the training data, rather than the underlying pattern. This can lead to poor generalization performance on new unseen data.

Q6:- What is an ensemble technique in machine learning?

Ans:- A machine learning approach where several models are trained to address a common problem, and their predictions are combined to enhance the overall performance.

Q7:- What is the difference between Bagging and Boosting techniques?

Ans:- Bagging and boosting are ensemble learning techniques. Bagging (Bootstrap Aggregating) reduces variance by averaging multiple models, while boosting reduces bias by combining weak learners sequentially to form a strong learner.

Q8:- What is out-of-bag error in random forests?

Ans:- Out-of-bag (OOB) error, also called out-of-bag estimate, is a method of measuring the prediction error of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating (bagging). Bagging uses subsampling with replacement to create training samples for the model to learn from.

Q9:- What is K-fold cross-validation?

Ans:- In K-fold cross-validation, the data set is divided into a number of K-folds and used to assess the model's ability as new data become available. K represents the number of groups into which the data sample is divided. For example, if you find the k value to be 5, you can call it 5-fold cross-validation.

Q10:- What is hyper parameter tuning in machine learning and why it is done?

Ans:- Hyperparameter tuning is the process of selecting the optimal values for a Machine Learning model's hyperparameters. Hyperparameters are settings that control the learning process of the model, such as the learning rate, the number of neurons in a neural network, or the kernel size in a support vector machine. The goal of hyperparameter tuning is to find the values that lead to the best performance on a given task.

Q11:- . What issues can occur if we have a large learning rate in Gradient Descent?

Ans:- If the learning rate is too high, the algorithm may overshoot the minimum, and if it is too low, the algorithm may take too long to converge. Overfitting: Gradient descent can overfit the training data if the model is too complex or the learning rate is too high.

Q12:- Can we use Logistic Regression for classification of Non-Linear Data? If not, why?

Ans:- Logistic regression is simple and easy to implement, but it also has some drawbacks. One of them is that it assumes a linear relationship between the input features and the output. This means that it cannot capture the complexity and non-linearity of the data.

Q13:- Differentiate between Adaboost and Gradient Boosting.

Ans:- Overall gradient boosting is more robust to outliers and noise since it equally considers all training instances when optimizing the loss function. AdaBoost is faster but more impacted by dirty data since it fixates on hard examples.

Q14:- What is bias-variance trade off in machine learning?

Ans:- In statistics and machine learning, the bias–variance trade off 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.

Q15:- Give short description each of Linear, RBF, Polynomial kernels used in SVM?

Ans:- SVM algorithms use a set of mathematical functions that are defined as the kernel. The function of kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. These functions can be different types. For example linear, nonlinear, polynomial, radial basis function (RBF*), and sigmoid.*
Introduce Kernel functions for sequence data, graphs, text, images, as well as vectors. The most used type of kernel function is RBF. Because it has localized and finite response along the entire x-axis.

The kernel functions return the inner product between two points in a suitable feature space. Thus by defining a notion of similarity, with little computational cost even in very high-dimensional spaces.