ASSIGNMENT ON 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 is the standard goodness-of-fit measure for linear models. Its formula incorporates RSS. Typically, you'll evaluate R 2 rather than the RSS because it avoids some of RSS's limitations, making it much easier to interpret.

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 :- the sum of squares total, measures the total variability in the dependent variable (the outcome we're trying to predict).

ESS :- the sum of squares due to regression, represents the variability explained by our regression model.

RSS :- the sum of squares error, captures the unexplained variability in the data.

Relationship Between TSS, ESS, and RSS:-

$$TSS = ESS + RSS$$

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

Ans :- Overfitting Prevention :- Overfitting occurs when a model learns to perform exceptionally well on the training data but fails to generalize to unseen data.

Bias-Variance Trade-off :- Regularization strikes a balance between bias and variance.

4. What is Gini–impurity index?

Ans :- Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen.

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

Ans :- Yes, unregularized decision trees are prone to overfitting. Decision trees tend to overfit when the available training data is limited, as they attempt to extract patterns even from noise Decision trees can grow to a considerable depth, resulting in intricate decision boundaries. As the tree becomes deeper, it becomes more susceptible to overfitting

6. What is an ensemble technique in machine learning?

Ans :- Ensemble learning is a machine learning technique that combines the predictions from multiple individual models to obtain a better predictive performance than any single model.

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

Ans :- Bagging: It is a homogeneous weak learners' model that learns from each other independently in parallel and combines them for determining the model average.

Boosting: It is also a homogeneous weak learners' model but works differently from Bagging. In this model, learners learn sequentially and adaptively to improve model predictions of a learning algorithm.

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

Ans :- OOB (out-of-bag) score is a performance metric for a machine learning model, specifically for ensemble models such as random forests. It is calculated using the samples that are not used in the training of the model, which is called out-of-bag samples. These samples are used to provide an unbiased estimate of the model's performance, which is known as the OOB score.

9. What is K-fold cross-validation?

Ans :- K-fold cross-validation is a statistical method used to estimate the skill of machine learning models. It is commonly employed in applied machine learning to compare and select a model for a given predictive modeling problem.

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

Ans :- in machine learning, hyperparameters are configuration variables that are set before the training process of a model begins.Unlike regular model parameters (such as weights and biases), hyperparameters cannot be directly learned from the data.They control the learning process itself and express important properties of the model, such as its complexity or how fast it should learn. It done because of The goal of hyperparameter tuning is to find the optimal values for these hyperparameters.By selecting the right hyperparameters, we can improve the model's performance, accuracy, and generalization.

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

Ans :- If the learning rate is too large in Gradient Descent, it can "jump over" the minima we are trying to reach, leading to osculations around the minimum or in some cases to outright divergence. If the learning rate is too high, the steps taken will be large and we can miss the minima, causing the model to fail to converge. On the other hand, if the learning rate is too small, the model will take too much time to reach the minimum cost with only small steps.

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

Ans :- Overall, logistic regression is a popular and effective method for binary classification problems. However, it may not be suitable for more complex classification problems where there are multiple classes or nonlinear relationships between the input variables and the outcome.

13. Differentiate between Adaboost and Gradient Boosting.

Ans :- AdaBoost is a boosting algorithm, which also works on the principle of the stagewise addition method where multiple weak learners are used for getting strong learners. Unlike Gradient Boosting in XGBoost, the alpha parameter I calculated is related to the errors of the weak learner, here the value of the alpha parameter will be indirectly proportional to the error of the weak learner.

Gradient Boosting is the boosting algorithm that works on the principle of the stagewise addition method, where multiple weak learning algorithms are trained and a strong learner algorithm is used as a final model from the addition of multiple weak learning algorithms trained on the same dataset.

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

Ans :- The bias-variance tradeoff is a fundamental concept in machine learning and statistics. It refers to the delicate balance between two sources of error in a predictive model: bias and variance. Bias represents the error due to overly simplistic assumptions in the learning algorithm.

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

Ans :- Linear Kernel :- The linear kernel is employed when the data is linearly separable. It assumes that a single straight line can effectively separate the data points.Mathematically, it computes the inner product between the input data points in the original feature space.Linear SVMs using this kernel are efficient and work well for linearly separable datasets.

Radial Basis Function (RBF) Kernel :-The RBF kernel, also known as the Gaussian kernel, is a powerful and widely used kernel in SVM.Unlike linear or polynomial kernels, RBF is more complex and efficient. It can handle non-linearly separable data by mapping it into a higher-dimensional space.The RBF kernel works by finding dot products and squares of all features in the dataset, allowing classification using the basic idea of Linear SVM.

Polynomial Kernel:The polynomial kernel captures the similarity of vectors in the training data by transforming them into a higher-dimensional space using polynomials of the original variables.It allows

SVMs to learn non-linear decision boundaries.The degree of the polynomial determines the complexity of the model. Higher degrees can fit more complex data.