MACHINE LEARNING

Q1 to Q15 are subjective answer type questions, Answer them briefly.

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 a goodness – of – fit measure for linear regression models. The statistic indicates the percentage of the variance in the dependent variables explain collectively.

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=ESS+RSS , where TSS is total sum of squares,ESS is explained sum of squares and RSS is Residual sum of squares.

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

Ans:- Regularization is a technique used in machine learning to reduce errors and avoid overfitting or underfitting

4. **What is Gini–impurity index?**

Ans:- The Gini-impurity index is the additional approach to dividing a decision tree.

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

Ans:-

6. **What is an ensemble technique in machine learning?**

Ans:- Ensemble technique is a machine learning technique that combine 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 reduces variance and avoids overfitting . Boosting reduces bias and iteratively improves accuracy .

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

Ans:- The OOB Error provides an unbiased estimate of the model's performance without the need for a separate validation set.

9. **What is K-fold cross-validation?**

Ans:- The K-fold cross validation signifies he data set splits into a K number . It divides the data set at the point where the testing set utilizes each fold .

R

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

Ans:- Hyper parameter tuning allows data scientists to tweak model performance for optimal results. This process is an essential part of machine learning and choosing appropriate hyperparameter values is crucial for success.

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

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

12. **Can we use Logistic Regression for classification of Non-Linear Data? If not, why?**
Ans:- Longistic Regression has traditionally been used to come up with a hyperplane that separates the feature space into classes. But if we suspect that the decision boundary is non linear we may get better results by attempting some non linear functional formsfor the logit function.

13. **Differentiate between Adaboost and Gradient Boosting.**
Ans:- Adaboost is the first designed boosting algorithm with a particular loss function. On the other hand , Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem. This makes Gradient Boosting more flexible than Adaboost.

14. **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.

15. **Give short description each of Linear, RBF, Polynomial kernels used in SVM.**
Ans:-Linear Kernel – The linear kernel produces a decision boundary that is a hyperplane separates data points from different classes in a linear fashion.

Polynomial Kernel – In machine learning, the polynomial kernel is a kernel function commonly used with support vector machines and other kernelized models, that represents the similarity of vectors in a feature space over polynomials of the original variable ,allowing learning of non linear models.

RBF Kernel :- RBF Kernel is a popular kernel function used in various kernelized learning algorithms and support vector machine classification.