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 ofgoodness of fit model in regression and why?**

Ans. Both R-squared and RSS gives a better measure of goodness of fit model in regression in there own ways , but R-squared is better and RSS because RSS doesn't provide a normalized measure like R-squared ,making the RSS difficult to compare models with different scales.

2. **What are TSS (Total Sum of Squares), ESS (Explained Sum of Squares) and RSS (Residual Sumof Squares) in regression. Also mention the equation relating these three metrics with each other.**

Ans.

a) RSS (Residual Sumof Squares):It is difference between the observed values and the predicted values.
b) ESS (Explained Sum of Squares): It is difference between the predicted values and the mean of all the predicted values.
c) TSS (Total Sum of Squares):It is the sum of squared differences between each observed value and the mean of actual values.

TSS=RSS+ESS

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

Ans. Regularization is required to avoid the the underfitting or overfitting of the model.

4. **What is Gini–impurity index?**

Ans. The Gini impurity index is a measure used in decision tree algorithms to assess the impurity or disorder of a set of data points within a node, aiding in the construction of an effective decision tree for classification tasks.

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

Ans.
Yes, unregularized decision trees are prone to overfitting. Decision trees, with their ability to create complex and deep structures to perfectly fit the training data, are particularly susceptible to overfitting when not regularized.

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

Ans.
An ensemble technique in machine learning is a method that combines the predictions of multiple individual models to improve overall performance or accuracy.

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

Ans.

a) Bagging : Each model is trained independently and produces its own prediction. The final prediction is often an average or a voting mechanism across all the individual predictions.
b) Boosting: Here the models are tested in a sequential order and the recovering the mistakes in the previous model.The final prediction is the weighted sum of individual model predictions.

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

Ans. In a Random Forest algorithm, the out-of-bag (OOB) error is a way to estimate the model's performance without the need for a separate validation set.

9. What is K-fold cross-validation?

Ans. K-fold cross validation is a technique in machine learning in which the datasets is divided into k folds of subsets and one of those subset is used for predicting and others for training .In this way multiple iterations are done and the average of all iterations are taken to get a better result.

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

Ans. Hyperparameter tuning, also known as hyperparameter optimization, is the process of finding the best set of hyperparameters for a machine learning model. Hyperparameters are configuration settings that are external to the model and are not learned from the data. They are set prior to the training process and significantly influence the performance of the model.

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

Ans. Having a large learning rate in Gradient Descent can lead to Divergence,Oscillations,Instability.

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

Ans. Logistic Regression cannot be applied for classification of Non-Linear Data, because the linear model is primarily designed for binary classification tasks, where the decision boundary is a linear function of the input features.

13. Differentiate between Adaboost and Gradient Boosting.

Ans. Adaboost and Gradient Boosting are both ensemble learning techniques that aim to improve the performance of weak learners by combining them into a strong predictive model. Despite sharing the boosting framework, they differ in their optimization strategies, loss functions, and how they assign weights to data points during training.

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

Ans. The bias-variance tradeoff is a fundamental concept in machine learning that describes the tradeoff between the model's ability to fit the training data accurately and its ability to generalize to new, unseen data

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

Ans.

a) The Linear kernel is appropriate for linearly separable data and provides a straightforward computation of the dot product.

b) The RBF kernel is versatile and effective for capturing complex non-linear patterns. Its flexibility makes it a popular choice in practice, although it might require careful tuning of hyper parameters.

c) The Polynomial kernel introduces non-linear features based on polynomial functions and is suitable for capturing various degrees of complexity in the data.