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?

Answer R-squared measures how much of the variation in the data is explained by the model.it is good for understanding overall model performance but can be misleading, especially with many predictors.range is From 0 to 1; closer to 1 means a better fit. And Residual Sum of Squares (RSS) is Total of the squared differences between observed and predicted values.it Doesn't account for the number of predictors, making comparisons between models difficult.in RSS Lower values mean a better fit; it directly measures prediction error. It is better to use them together

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.

Answer TSS is the sum of the squared differences between each observed value and the mean of the observed values.ESS is the sum of the squared differences between the predicted values and the mean of the observed values and RSS is the sum of the squared differences between the observed values and the predicted values. The relationship between these three is:

TSS=ESS+RSS

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

Answer Regularization is used to prevent overfitting and improve the generalization performance of models. Basically regularization adds a penalty term to the loss function, preventing the model from learning overly complex patterns that wont generalize enough to unseen data

. 4.What is Gini—impurity index?

Answer The Gini impurity index is used to measure impurity or disorder in a dataset.0 indicates pure and 1 indicates max impure.In decision tree algorithms, Gini impurity helps to sort out the best feature to split the data at each node. The goal is to minimize the Gini impurity after the split, leading to

more homogenous child nodes.the Gini impurity index helps make decisions about how to split data to improve classification accuracy

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

Answer Yes, unregularized decision trees tend to overfit the data. They can become too complicated by splitting the data until each group is very small, often fitting random noise instead of real patterns. This means that small changes in the data can lead to very different trees, making them less reliable for new data. Without pruning to remove extra branches, the tree stays too complex and captures irrelevant details. To avoid overfitting, it helps to limit how deep the tree can grow, set a minimum number of samples for each split, or prune unnecessary branches.

6. What is an ensemble technique in machine learning?

Answer An ensemble technique in machine learning combines multiple models to improve predictions. It works better than using a single model. There are two main types: bagging, which trains several models on different random samples and averages their predictions (like in Random Forests), and boosting, where models are built one after another, focusing on correcting the previous model's mistakes. Overall, ensemble methods enhance accuracy and reduce errors.

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

Answer Bagging and boosting are techniques to improve machine learning models but work differently. Bagging (Bootstrap Aggregating) reduces errors by training multiple models on different random samples of the data, then combines their predictions by averaging or voting, like in Random Forests. Boosting, on the other hand, trains models sequentially, with each new model correcting the mistakes of the previous one, giving more weight to better performers. Examples include AdaBoost and Gradient Boosting. In summary, bagging focuses on reducing variance, while boosting improves accuracy by reducing both bias and variance.

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

Answer Out-of-bag (OOB) error measures how well a Random Forest model performs without needing a separate test set. Each tree is trained on a random sample of the data, which means some data points are left out—these are called out-of-bag samples. To calculate the OOB error, each of these samples is predicted by the trees that didn’t use them for training. The OOB error shows

the rate of wrong predictions for classification or the average error for regression. This method helps estimate the model's accuracy using the data it has already seen.

9. What is K-fold cross-validation?

Answer K-fold cross-validation is a method used to evaluate how well a machine learning model performs. In this approach, we split the dataset into K equal parts or "folds." For example, if K is 5, we divide the data into 5 groups. The model is then trained and tested K times, where each time one group acts as the test set while the remaining groups are used for training. This way, every data point gets tested once. After completing this for all K folds, we average the results to get a better idea of the model's performance. This technique ensures a fair evaluation and helps reduce the risk of overfitting, making the model more reliable.

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

Answer Hyperparameter tuning in machine learning is about adjusting the settings that control how a model learns, and these settings are not learned from the data itself. For example, this includes things like the learning rate or the number of trees in a Random Forest. Tuning is important because it can enhance the model’s accuracy, help it perform better on new data, and make the training process more efficient. Since different problems require different settings, tuning allows us to customize the model for specific tasks. Common methods for tuning include grid search and random search. In short, hyperparameter tuning is essential for building better-performing models.

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

Answer If we use a large learning rate in Gradient Descent, several problems can happen. First, the model might overshoot the lowest point of the loss function, causing it to diverge instead of finding a solution. This can lead to oscillations, where the updates jump back and forth without settling down. The training process can become unstable, resulting in unpredictable changes in loss values. Even if it doesn’t diverge, the model might take a long time to converge, or it could completely miss the best solution. So, it's important to choose the right learning rate to avoid these issues.

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

Answer Logistic Regression is mainly used for binary classification, but it struggles with non-linear data. This is because it creates a straight-line decision boundary, which is not effective for separating classes that are not linearly related. As a result, it may lead to poor predictions and misclassifications. To deal with non-linear data, we can either transform the features to capture non-linearity or use other models like Decision Trees or Support Vector Machines that can handle complex relationships better. In short, Logistic Regression is not the best choice for non-linear classification without some adjustments.

13. Differentiate between Adaboost and Gradient Boosting.
Answer Adaboost and Gradient Boosting are both methods used to improve machine learning models, but they work differently. Adaboost focuses on giving more weight to misclassified samples, often using simple models like decision stumps. It adjusts these weights in each round to help the model learn from its mistakes. In contrast, Gradient Boosting fits new models to the errors of the previous ones, directly minimizing the loss function. It can use more complex models and is flexible enough for different types of problems. In short, Adaboost corrects mistakes by focusing on tough samples, while Gradient Boosting improves by reducing overall errors.

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

Answer The bias-variance trade-off is an important concept in machine learning that helps us understand model errors. Bias is the error that occurs when a model is too simple and fails to capture the true patterns in the data, leading to underfitting. For example, a linear model trying to fit a complex dataset has high bias. On the other hand, variance is the error that happens when a model is too complex and sensitive to small changes in the training data, resulting in overfitting. The goal is to find a balance between bias and variance to minimize total error. A good model captures the right patterns while still performing well on new data.

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

Linear Kernel:linear kernel is the simplest type, which uses the dot product of two input vectors. It works well when the data can be separated with a straight line. It's efficient and great for high-dimensional data.

RBF (Radial Basis Function) Kernel:RBF kernel is a non-linear option that measures how similar two points are based on their distance. It’s useful for data that isn't linearly separable, allowing for more flexible decision boundaries.

Polynomial Kernel:polynomial kernel calculates similarity by taking the dot product of input vectors raised to a certain power. This allows it to capture interactions between features and create non-linear decision boundaries.