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 used measure of goodness of fit in regression, indicating the proportion of variance explained by the model.

RSS provides an absolute measure of fit but is not as comparable across different model.

Therefor R-squared is a better measure of goodness of fit model in regression.

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. RSS: The residual sum of squares measures the level of variance in the error of a regression model. The smaller the RSS the better your model fits data, the greater the RSS the poorer your model fits.

ESS: The ESS measures how much variation there is in the model values and this is compared to TSS.

TSS: The TSS measures how much variation there is in the observed data and to the RSS.

Relation: TSS=ESS+RSS

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

Ans. When we use regression models to train some data, there is good chance that the model will overfit training data. Therefor Regularization helps to avoid overfitting.

Types of regularization:

1.LASSO (Least Absolute Shrinkage and Selection Operator)

2. Ridge Regression

4. What is Gini–impurity index?

Ans. Gini impurity is a measure of how often a randomly chosen element from the set would be incorrectly labelled if it was randomly labelled according to the distributions of labels in the subset. It is calculated by multiplying the probability that a given observation is classified into the correct class and sum of all the probabilities when that particular observation is classified into the wrong class.

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

Ans. Decision- trees are prone to overfitting. Overfitting can mean creating too many branches based on outliers in the training data.

6. What is an ensemble technique in machine learning?

Ans. Ensemble technique is improving the accuracy of models by combining multiple models.

There are 3 Ensemble techniques:

1. Bagging 2. Boosting 3. Stacking

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

Ans. Bagging**:** -Training data subsets are drawn randomly with replacement from the entire training dataset.

-Bagging attempts to tackle the overfitting issue.

-Every model receives an equal weight.

-Objective to decrease variance, not bias.

-Every model is built independently.

Boosting: -Each new subset contains the components that were misclassified by previous models.

-Boosting tries to reduce bias.

-Models are weighted by their performance

-Objective to decrease bias not variance.

-New models are affected by the performance of the Previously developed model.

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

Ans. In random forests it estimates the prediction error of the model without the need for a separate validation set. Out-of-bag error is computed by evaluating the model's performance on the training data points that were not included in the construction of each individual decision tree in the forest.

9. What is K-fold cross-validation?

An. To tackle the high variance of Hold-Out method, the K-fold method is used. Divide the whole dataset into 'k' sets in equal sizes. Then the 1st set is selected as the test set and the rest 'k-1' sets are used to train the data. Error is calculated for this particular dataset. Then the steps are repeated. Similarly, the process continuous for 'k' times. In the end, the CV error is given as the mean of the total errors.

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

Ans. Hyper parameter tuning is the process of selecting the optimal values for a models hyper parameter. Tuning those parameters is called hyper parameter.

The goal of hyper parameter tuning is to find the values that lead to the best performance on a given task.

To improve accuracy of model.

There are two types of Hyper parameter

1.GridSearchCV 2. RandomisedSearchCV

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

Ans. If we have a large learning rate in Gradient Descent then it can lead to exploding performance over the training time and to a slow final performance.

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

Ans. Logistic Regression is used as linear classifier. When the classes can be separated in the feature space by linear boundaries. That can be remedied however if we happen to have a better idea as to the shape of decision boundary.

13. Differentiate between Ada boost and Gradient Boosting.

Ans. Ada boost: The shift is made by up-weighting the observations that are miscalculated prior.

The trees are called decision stumps.

Gradient Boosting: Gradient Boosted trees use decision trees as estimators. It can work with different loss functions, evaluate its gradient and approximates it with a simple tree.

Gradient Boosting will try to Reduce the Residual Error.

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

Ans. The bias-variance trade off is about finding the right balance between simplicity and complexity in a machine learning model. A model that is too simple has high bias and low variance but lacks the capacity to learn from the data. In contrast, a highly complex model has low bias but high variance, causing it to overfit and perform poorly and new data.

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

Ans. Linear Kernel: It is use when data is linearly separable.

Its produces a decision boundary that is a hyperplane in the feature space.

RBF Kernel: This kernel is used to measure the similarity between pairs of data points in the feature space.

Polynomial kernel: It represents the similarity of vectors in the training set of data in a feature space over polynomials of the original variables used in the kernel.