Question 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 and Residual Sum of Squares (RSS) are both measures of the goodness of fit of a regression model, but they serve slightly different purposes.

R-squared is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. In other words, R-squared measures the extent to which changes in the dependent variable can be predicted by changes in the independent variable(s). Higher R-squared values indicate a better fit of the regression model to the data. Therefore, R-squared is often used to compare different models and select the best one.

On the other hand, Residual Sum of Squares (RSS) measures the difference between the observed values of the dependent variable and the predicted values by the model. It represents the sum of the squared differences between the actual and predicted values of the dependent variable. The goal is to minimize the residual sum of squares to obtain a better model fit.

In terms of determining the goodness of fit of a model, R-squared is generally considered a better measure than RSS. This is because R-squared provides an overall measure of the proportion of variance in the dependent variable that is explained by the model, whereas RSS only measures the magnitude of the residuals. Additionally, R-squared is a standardized measure and ranges from 0 to 1, making it easy to compare the fit of different models. In contrast, the magnitude of the RSS value depends on the scale of the dependent variable and can't be easily compared across models.

However, it's worth noting that neither R-squared nor RSS is a perfect measure of model fit. R-squared can be influenced by outliers or data points that don't fit the model well, while RSS doesn't take into account the number of variables or degrees of freedom in the model. Therefore, it's important to consider multiple metrics when evaluating a regression model's goodness of fit.

Question 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 : TSS is the sum of square of difference of each data point from the mean value of all the values of target variable (y).

The explained sum of squares (ESS) is the sum of the squares of the deviations of the predicted values from the mean value of a response variable, in a standard regression

model — for example, $y_i = a + b_1x_{1i} + b_2x_{2i} + ... + \varepsilon_i$, where y_i is the i^{th} observation of the response variable, x_{ji} is the i^{th} observation of the j^{th} explanatory variable, a and b_j are coefficients, i indexes the observations from 1 to n, and ε_i is the i^{th} value of the error term. In general, the greater the ESS, the better the estimated model performs.

If

$a\hat{}$

and

$b\hat{}i$

are the estimated coefficients, then

$$y\hat{}i=a\hat{}+b\hat{}1x1i+b\hat{}2x2i+\cdots$$

is the i^{th} predicted value of the response variable. The ESS is then:

$$ESS=\sum i=1n(y\hat{}i-\bar{y})2$$

where

$y\hat{}i$ is the value estimated by the regression line

RSS : Residual Sum of Squares (RSS) measures the difference between the observed values of the dependent variable and the predicted values by the model

Total sum of squares (TSS) = explained sum of squares (ESS) + residual sum of squares (RSS).

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

Ans : The primary goal of regularization is to reduce the model's complexity to make it more generalizable to new data, thus improving its performance on unseen datasets.

Question 4. What is Gini–impurity index?
Ans : Gini Impurity measures how well does a node splits the data set between the two outcomes. It aims to reduce the impurity score from the root node of the tree to the leaf node.

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

Ans : Yes, Decision trees, by their very nature, are prone to overfitting, especially when they are deep. Overfitting occurs when a model captures noise or fluctuations in the training data that do not represent the underlying data distribution. In the context of decision trees, overfitting can mean creating too many branches based on outliers or anomalies in the training data.

Question 6. What is an ensemble technique in machine learning?
Ans : Ensemble techniques in machine learning involve combining multiple models to improve performance.

Question 7. What is the difference between Bagging and Boosting techniques?
Ans : One common ensemble technique is bagging, which uses bootstrap sampling to create multiple datasets from the original data and trains a model on each dataset. Another technique is boosting, which trains models sequentially, each focusing on the previous models' mistakes.

Question 8. What is out-of-bag error in random forests?
Ans : Out-of-bag (OOB) error, also called out-of-bag estimate, is a method of measuring the prediction error of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating (bagging).

Question 9. What is K-fold cross-validation?
Ans : In K-fold cross-validation, the data set is divided into a number of K-folds and used to assess the model’s ability as new data become available. K represents the number of groups into which the data sample is divided. For example, if you find the k value to be 5, you can call it 5-fold cross-validation.

Question 10. What is hyper parameter tuning in machine learning and why it is done?
Ans : Hyperparameters directly control model structure, function, and performance. Hyperparameter 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.

Question 11. What issues can occur if we have a large learning rate in Gradient Descent?
Ans : Large learning rate can lead to exploding or oscillating performance over the training epochs and to a lower final performance.

Question 12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?
Ans : Logistic Regression is a linear classifier, which means it's designed to classify data points using a linear boundary. This linear boundary is defined by a linear combination of the feature variables.

When it comes to non-linear data, Logistic Regression may not perform well because it cannot capture non-linear relationships between features. If the decision boundary that separates different classes in the data is non-linear, Logistic Regression will not be able to model it accurately.

Question 13. Differentiate between Adaboost and Gradient Boosting?
Ans : The most significant difference is that gradient boosting minimizes a loss function like MSE or log loss while AdaBoost focuses on instances with high error by adjusting their sample weights adaptively.

Gradient boosting models apply shrinkage to avoid overfitting which AdaBoost does not do. Gradient boosting also performs subsampling of the training instances while AdaBoost uses all instances to train every weak learner.

Overall gradient boosting is more robust to outliers and noise since it equally considers all training instances when optimizing the loss function. AdaBoost is faster but more impacted by dirty data since it fixates on hard examples.

Question14. What is bias-variance trade off in machine learning?
Ans : In statistics and machine learning, the bias–variance tradeoff 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.

Question 15. Give short description each of Linear, RBF, Polynomial kernels used in SVM
Ans : The most common SVM kernels are linear, good for straight-line data, polynomial, and useful for curves. Radial basis function (RBF), is great for complex patterns. Also, sigmoid can handle different kinds of data changes.

MCQ

1. Using a goodness of fit,we can assess whether a set of obtained frequencies differ from a set of frequencies.
Ans: d) Expected

2. Chisquare is used to analyse
Ans: c) Frequencies

3. What is the mean of a Chi Square distribution with 6 degrees of freedom?
Ans c) 6

4. Which of these distributions is used for a goodness of fit testing?
Ans b) Chisqared distribution

5. Which of the following distributions is Continuous
Ans: c) F Distribution

6. A statement made about a population for testing purpose is called?
Ans : b) Hypothesis
7. If the assumed hypothesis is tested for rejection considering it to be true is called?
Ans: a) Null Hypothesis

8. If the Critical region is evenly distributed then the test is referred as?
Ans :a) Two tailed

9. Alternative Hypothesis is also called as?
Ans: b) Research Hypothesis

10. In a Binomial Distribution, if ‘n’ is the number of trials and ‘p’ is the probability of success, then the mean value is
given by
Ans : a) np