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 generally a better measure of the goodness of fit for a regression model than the residual sum of squares (RSS).

The reason why R^2 is often preferred over RSS as a measure of goodness of fit is due to its standardized nature:

1. Scalability: R^2 is scale-invariant, meaning it does not change if the scale of the data changes, whereas RSS is affected by the scale of the dependent variable. This makes R^2 a better choice when comparing models fitted on different scales.

2. Interpretability: R^2 has an intuitive interpretation as the proportion of variance explained, which is easier to understand than the sum of squared residuals. An R^2 of 0.75 means that 75% of the variance in the dependent variable is explained by the model, which is a straightforward interpretation.

3. Benchmarking: R^2 provides a clear benchmark. An R^2 of 0 indicates that the model explains none of the variability in the response data around its mean, while an R^2 of 1 indicates that the model explains all the variability.

4. Adjustment for model complexity: Adjusted R^2 takes into account the number of predictors in the model, which helps in assessing whether the addition of a new predictor really improves the model or is just adding complexity without significantly improving the fit.

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.

In statistics, the explained sum of squares (ESS), alternatively known as the model sum of squares or sum of squares due to regression (SSR – not to be confused with the residual sum of squares (RSS) or sum of squares of errors), is a quantity used in describing how well a model, often a regression model, represents the data being modelled. In particular, the explained sum of squares measures how much variation there is in the modelled values and this is compared to the total sum of squares (TSS), which measures how much variation there is in the observed data, and to the residual

sum of squares, which measures the variation in the error between the observed data and modelled values.

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,

$Yi = a + b1x1i + b2x2i + \ldots + \varepsilon i,$

Where yi is the I th observation of the response variable

, xji is the I th observation of the j th explanatory variable,

A and bj are coefficients,

I indexes the observations from 1 to n, and εi is the I th value of the error term.

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

Ans: Regularization is a technique used in machine learning to prevent overfitting and improve the generalization performance of models. In essence, regularization adds a penalty term to the loss function, discouraging the model from learning overly complex patterns that may not generalize well to unseen data

4. What is Gini–impurity index?

Ans: The Gini impurity index, also known as the Gini index, is a measure of how mixed up or impure a dataset is. It's commonly used in decision tree algorithms, particularly for classification tasks.

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

Ans: yes, Overfitting in decision tree models occurs when the tree becomes too complex and captures noise or random fluctuations in the training data, rather than learning the underlying patterns that generalize well to unseen data. Other reasons for overfitting include:

Complexity: Decision trees become overly complex, fitting training data perfectly but struggling to generalize to new data.

Memorizing Noise: It can focus too much on specific data points or noise in the training data, hindering generalization.

Overly Specific Rules: Might create rules that are too specific to the training data, leading to poor performance on new data.

Feature Importance Bias: Certain features may be given too much importance by decision trees, even if they are irrelevant, contributing to overfitting.

Sample Bias: If the training dataset is not representative, decision trees may overfit to the training data's idiosyncrasies, resulting in poor generalization.

Lack of Early Stopping: Without proper stopping rules, decision trees may grow excessively, perfectly fitting the training data but failing to generalize well.

6. What is an ensemble technique in machine learning?

Ans: Ensemble techniques in machine learning combine multiple models to make predictions or classifications. The goal is to improve the overall performance and accuracy of the predictions by combining the strengths of the individual models

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

Ans: Bagging and boosting are both machine learning ensemble techniques that improve the accuracy and stability of algorithms. The main difference between the two is how they train and combine base learners:

Bagging

Trains base learners independently in parallel, and then combines their predictions to get the final result. Bagging is best for data with high variance, low bias, and low noise.

Boosting

Trains base learners sequentially, with each learner trying to correct the mistakes of the previous one. Boosting combines the predictions of the base learners using weighted averaging, giving more weight to better models. Boosting is best for data with low variance, high bias, and high noise.

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

Ans: The out-of-bag (OOB) error is a way to measure the prediction error of a random forest model. It's a useful tool for machine learning professionals and data scientists because it provides an accurate estimate of model performance without the need for cross-validation

9. What is K-fold cross-validation?

Ans: K-fold cross validation in machine learning cross-validation is a powerful technique for evaluating predictive models in data science. It involves splitting the dataset into k subsets or folds, where each fold is used as the validation set in turn while the remaining k-1 folds are used for training.

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

Ans: Hyperparameter tuning is the process of finding the best set of hyperparameters for a machine learning model to improve its performance. Hyperparameters are variables that control the model training process and are set before the learning process begins.

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

Ans: Using a large learning rate in gradient descent can cause a number of issues, including:

Overshooting the optimal point: A large learning rate can cause the algorithm to take big steps in the direction of the negative gradient, which can cause it to skip the minimum point.

Divergence: A large learning rate can cause the algorithm to diverge, resulting in poor performance or failure to converge.

Instability: A large learning rate can cause instability in the algorithm.

Oscillations: A large learning rate can cause oscillations in the algorithm's performance.

Lower final performance: A large learning rate can result in lower final performance.

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

Ans: No, logistic regression is not ideal for classifying non-linear data because it assumes a linear relationship between the input variables and the output.

13. Differentiate between Adaboost and Gradient Boosting.

Ans: AdaBoost and Gradient Boosting are both machine learning algorithms that combine multiple weak learners to create a stronger model. However, they differ in several ways, including:

How they focus on errors

AdaBoost focuses on misclassified instances by adjusting their weights, while Gradient Boosting minimizes a loss function.

How they build models

AdaBoost builds a sequence of models by adding stumps, while Gradient Boosting builds a sequence of models by adding trees.

How they handle outliers

Gradient Boosting is more robust to outliers and noise than AdaBoost.

How they handle missing values

AdaBoost and Gradient Boosting require explicit imputation of missing values, while XGBoost has built-in functionality for this.

How they handle multi-class classification

AdaBoost and Gradient Boosting require a One-vs-All approach to solve multi-class problems, while XGBoost can handle them natively.

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

Ans: The bias-variance trade-off is a fundamental concept in machine learning that refers to the balance between a model's ability to represent data patterns and its susceptibility to fluctuations in training data.

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

Ans: The linear, polynomial and RBF or Gaussian kernel are simply different in case of making the hyperplane decision boundary between the classes.

The kernel functions are used to map the original dataset (linear/nonlinear) into a higher dimensional space with view to making it linear dataset.

Usually linear and polynomial kernels are less time consuming and provides less accuracy than the rbf or Gaussian kernels.

The k cross validation is used to divide the training set into k distinct subsets. Then every subset is used for training and others k-1 are used for validation in the entire trainging phase. This is done for the better training of the classification task.