

MACHINE LEARNING

1) R square is a better measure of goodness of fit because it gives a normalized value that is easier to interpret and compare across different datasets and models.

2) TSS is the total variability in the observed data.

ESS is the variability given by the regression model.

RSS is the residual variability(not explained).

$TSS = ESS + RSS$ (this helps in calculation of R square).

3) Regularization can help balance the trade off between model bias(underfitting) and model variance(overfitting) in ML leading to improved performance. L1 method of regularization helps to reduce the coefficient's to zero which can also remove multicollinearity.

4) Gini Impurity is a metric in decision tree algorithm that measures the impurity or disorder of a set of elements. A lower Gini impurity suggests a better split.

5) Yes, Unregularized decision tree is prone to overfitting as it grows lots of nodes which captures lot of noises and outliers due to which the model cant predict new data.

6) In Ensemble Technique the large dataset is divided into small subsets and each subset trains homogenous models individually and then their predictions are combined to enhance the performance. There are two types of Ensemble technique: Bagging and Boosting . Bagging trains the model data parallelly whereas Boosting trains sequentially.

7) Bagging aims to reduce the variance of the model by training multiple versions of the model on different subsets of the dataset and then combining their predictions.

Boosting aims to reduce both bias and variance by model in a sequential manner where each subsequent model attempts to correct the errors of the previous one.

8) when we divide the dataset into subsets during bagging ,some data points are left out and not included in the sample. These data points are called Out-of-Bag samples and the error rate of the model while predicting these left out samples is called out-of-bag error.

9) K fold cross validation is a method for evaluating the performance of a ML model where the complete dataset is available for training as well a testing. In this method the dataset is split into k folds and performing training and validation k times. It provides the model accuracy to generalize to new data.

10) Hyper parameter tuning is the technique through which the ML model finds out the best parameter that gives better R^2 score and then we can use the same parameter with the model to find the metrics and improve model's performance, It prevents overfitting and underfitting also. It has two methods: Grid Search and Random Search.

11) Large Learning Rate in Gradient Descent can cause the algorithm to fail to converge and increase the loss values and the parameters may fail to settle at the optimal values.

12) No we can't use Logistic Regression for classification of Non-Linear Data because logistic regression assumes the linear relationship between the input and output variables. and it cannot understand the non-linearity of the data.

13) AdaBoost: Focuses on misclassified instances by adjusting the weights, making weak learners pay more attention to hard to classify instances. It is sensitive to noisy data.

Gradient Boosting: Adds models sequentially to correct residual errors and optimized the loss functions through gradient descent. It is not sensitive to noisy data.

14) The Bias-Variance-Trade off finds balance between bias and variance to minimize total prediction error. High Bias leads to underfitting, model is simple ,model captures noise also. High Variance leads to overfitting, model is complex and captures noise.

15) Linear Kernel: Suitable for linearly separable data. RBF: Effective for non-linear data, maps it to infinite-dimensional space. Polynomial kernel: complex than linear but less than RBF.