R-squared (R^2) and Residual Sum of Squares (RSS) are both measures used to assess the goodness of fit of a regression model, but they capture different aspects of model performance.

Ans. R-squared (R^2) and Residual Sum of Squares (RSS) are both measures used to assess the goodness of fit of a regression model, but they capture different aspects of model performance.

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. TSS (Total Sum of Squares) represents the total variability in the dependent variable (response variable) without considering the regression model. It is the sum of the squared differences between each observed value of the dependent variable and the overall mean of the dependent variable.

ESS (Explained Sum of Squares) represents the variability in the dependent variable that is explained by the regression model. It is the sum of the squared differences between the predicted values from the regression model and the overall mean of the dependent variable.

RSS (Residual Sum of Squares) represents the unexplained variability in the dependent variable, i.e., the variability that is not accounted for by the regression model. It is the sum of the squared differences between the observed values of the dependent variable and the predicted values from the regression model.

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 a model. Overfitting occurs when a model learns the training data too well, capturing noise and random fluctuations rather than the underlying patterns. Regularization helps address this issue by adding a penalty term to the model's objective function, discouraging complex models with overly large coefficients.

What is Gini–impurity index?

Ans. The Gini impurity index is a measure of impurity or disorder in a set, commonly used in decision tree algorithms for classification problems. It quantifies the likelihood of misclassifying an element randomly chosen from the set based on the distribution of classes within that set. The Gini impurity of a set is calculated as the sum of the squared probabilities of each class being chosen, subtracted from 1. A lower Gini impurity indicates a purer set, and decision trees use this index to determine optimal splits

during the construction of the tree, ultimately leading to more homogeneous subsets with respect to the target variable.

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

Ans. Yes, unregularized decision trees are indeed prone to overfitting. Decision trees have the capacity to grow very deep, creating complex structures that perfectly fit the training data, including its noise and outliers. This overfitting phenomenon occurs because an unregularized decision tree will keep splitting nodes until each leaf contains very few instances, potentially resulting in a tree that memorizes the training data rather than learning its underlying patterns.

What is an ensemble technique in machine learning?

Ans. Ensemble techniques in machine learning involve combining the predictions of multiple models to create a more accurate and robust prediction than any individual model. The idea is to harness the collective wisdom of diverse models to improve overall performance. Ensemble methods include approaches such as bagging, boosting, stacking, and voting, and they are widely used to enhance generalization, reduce overfitting, and increase the stability and reliability of machine learning models. Popular ensemble algorithms include Random Forests, AdaBoost, Gradient Boosting, and various voting classifiers/regressors. Ensembles are effective when individual models have complementary strengths and weaknesses, and they have achieved significant success in various machine learning applications.

What is the difference between Bagging and Boosting techniques?

Ans. Bagging stands for Bootstrap aggregating, which combines several models for better predictive results. In statistical classification and regression, bagging improves the stability and accuracy of machine learning algorithms by decreasing the variance and reducing the chances of overfitting.

Boosting involves building a strong classifier from several weak classifiers using the weak models in series. The first step is to build a model from the training set. Then we create the second model, which tries to correct the error incurred while training the first. This process is continued while adding new models until the maximum number of models is reached, or the training data is finished.

Both are ensemble methods that improve the stability of the machine learning model.

Both generate one learner from multiple learners.

The final decision is by combining the predictions of the N learners.

Both algorithms help in dealing with the bias-variance trade-off.

Both can be used to solve classification as well as regression problems.

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

Ans. The out-of-bag (OOB) error in Random Forests is a measure of the model's performance on data points that were not used during the training of individual trees within the ensemble. It is calculated by leveraging the fact that each tree is trained on a different random subset of the data, allowing for an unbiased estimate of the model's ability to generalize to new, unseen data. The out-of-bag error is a useful validation metric, providing an estimate of performance without the need for a separate validation set. It is commonly used for hyperparameter tuning and contributes to the overall robustness and interpretability of Random Forest models.

What is K-fold cross-validation?

Ans. K-fold cross-validation is a model evaluation technique used in machine learning to assess the performance of a model on a dataset. The process involves partitioning the dataset into K equally sized folds or subsets. The model is trained and validated K times, each time using a different fold as the validation set and the remaining data as the training set.

Partitioning Data

Training and Validation:

Performance Metric:

Average Performance:

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

Ans. Hyperparameter tuning in machine learning is the process of selecting the optimal set of hyperparameters for a model before training. Hyperparameters are external configurations that influence a model's behavior and performance.

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

Ans. If the learning rate is too high, the algorithm may overshoot the minimum, and if it is too low, the algorithm may take too long to converge. Overfitting: Gradient descent can overfit the training data if the model is too complex or the learning rate is too high.

Differentiate between Adaboost and Gradient Boosting.

Ans. AdaBoost or Adaptive Boosting is the first Boosting ensemble model. The method automatically adjusts its parameters to the data based on the actual performance in the current iteration. Meaning, both the weights for re-weighting the data and the weights for the final aggregation are re-computed iteratively**.**

Gradient Boost is a robust machine learning algorithm made up of Gradient descent and Boosting. The word 'gradient' implies that you can have two or more derivatives of the same function. Gradient Boosting has three main components: additive model, loss function and a weak learner.

What is bias-variance trade off in machine learning?

Ans. The bias-variance trade-off in machine learning is the delicate balance between bias and variance in a predictive model.

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

Ans. The linear kernel is the simplest SVM kernel and is used for linearly separable data. It computes the dot product between the input features and is equivalent to using a linear decision boundary in the original feature space.

The RBF kernel, also known as the Gaussian kernel, transforms the data into an infinite-dimensional space. It considers the similarity (or distance) between data points and assigns higher weights to nearby points. The RBF kernel is versatile and can capture complex, non-linear relationships in the data.

The polynomial kernel transforms the data into a higher-dimensional space using polynomial functions. It allows SVMs to capture polynomial relationships between features. The degree parameter controls the degree of the polynomial.