

MACHINE LEARNING

Q1 to Q15 are subjective answer type questions, Answer them briefly.

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?

R-squared is generally a better measure of the goodness of fit for a regression model than the residual sum of squares (RSS). , is a statistical measure that represents the proportion of the variance for the dependent variable that's explained by the

2.what are tss rss ess in regression also mention the metrics with each other.

TSS = ESS + RSS, where TSS is Total Sum of Squares, ESS is Explained Sum of Squares and RSS is Residual Sum of Squares. The aim of Regression Analysis is explain the variation of dependent variable Y. Here “explained” should be apperantly explained.

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

Regularization is a critical technique in machine learning to reduce overfitting, enhance model generalization, and manage model complexity. Several regularization techniques are used across different types of models.

4.what is gini impurity index?

In machine learning, it is utilized as an impurity measure in decision tree algorithms for classification tasks. The Gini Index measures the probability of a haphazardly picked test being misclassified by a decision tree algorithm, and its value goes from 0 (perfectly pure) to 1 (perfectly impure).

5 an unregulated decision tree is prone to overfitting.

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.

6. what is an ensemble technique in machine learning

Ensemble learning refers to a machine learning approach where several models are trained to address a common problem, and their predictions are combined to enhance the overall performance.

7.what is the difference between bagging and boosting.

Bagging: Reduces variance by averaging predictions from models trained on different subsets of data. Effective for models with high variance. Boosting: Reduces bias by sequentially training models that focus on errors of previous models. Effective for models with high bias.

8.what is out of bag error in random forest.

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).

9. what is k fold cross validation in machine learning.

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.

10.what is hyper parameter tuning in machine learning

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.

what is the issue we can occur if we use lagre learning grate in gradiant decent

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.

can we use logistic regression for nonlinear data

A logistic regression means that you apply a logarithmic function to your variables. This models (some specific) non-linear relationships but the relationship is still monotonic, more of one variable always implies more of the other variable.

difference between adaboost and gradient boosting

Distinguishing Features of Gradient Boosting vs

AdaBoostOverall 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.

what is the bias-variance trade off in machine learning

In With kernels, SVMs can handle all kinds of relationships between data

points, making accurate predictions. Common types of kernels include linear, polynomial, RBF, and sigmoid, each good for different kinds of data. Picking the right kernel is super important because it decides how well the SVM will work.

give short description each of linear rbf polynomial kernels used in svm

With kernels, SVMs can handle all kinds of relationships between

data points, making accurate predictions. Common types of kernels include linear, polynomial, RBF, and sigmoid, each good for different

kinds of data. Picking the right kernel is super important because it decides how well the SVM will work.