

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 is better measure of goodness of fit model in regression because it indicates a higher amount of variability which is used by model.

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 is the combination of ESS and RSS. ESS is the sum of difference between the estimated value and the mean of dependent value. RSS is used to measure the variance level in error term in the regression model.

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

Answer: The need of regularization in machine learning is used to measure machine learning models in order to minimize the adjusted loss function and to prevent the overfitting.

4. What is Gini-impurity index?

Answer: Gini-impurity index is a measurement which is used to check the impurity and disorders in dataset in decision tree algorithms.

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

Answer. Yes, unregularized decision trees prone to overfitting because it requires a way to turn numeric data into single rule with decision.

6. What is an ensemble technique in machine learning?

Answer: Ensemble method is a technique which is used to check the accuracy by combining many models instead of single model.

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

Answer:

i) Boosting tries to reduce bias whereas, bagging attempts to tackle the overfitting or underfitting.

ii) If the classifier is high bias then we use boosting whereas, if the classifier is high variance we use bagging.

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

Answer: Out of bag error is a method to measure the predict error in random forest, specially in random forest algorithm that were not included in the construction of each decision tree.

9. What is K-fold cross-validation?

Answer: K- fold cross validation is technique for assess predictive models. The dataset is divided into folds. The model is used k times to, by using different types of folds as the validation set each time.

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

Answer: Hyper parameter tuning is the process to finding the optimal hyper parameter tuning for machine learning model.

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

Answer: If the learning rate is too high then the algorithm overshoots the minimum and if it is too low the algorithm takes too long to coverage.

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

Answer: Yes, we can use Logistic Regression for classification of Non-Linear Data.

13. Differentiate between Ada boost and Gradient Boosting.

Answer: Ada boost is a learning algorithm which is used to classifying data by combing multiple weak learners into strong learners. Gradient boosting is a algorithm which works as a tabular data with a target and a set of data.

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

Answer: Bias-variance tradeoff is used to describe the relation between model complexity the accuracy of its prediction and how accurately its can predict on previously unseen and unused data.

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

Answer:

Linear kernel is most commonly used and simplest kernel function. It defines the dot product between the original feature space and the input vectors.

RBF kernel stands for radial basis function. It is a popular kernel function used to kernelized learning algorithm. It can create smooth and circular decision boundaries.