

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?

Answer: The residual sum of squares (RSS) is the absolute amount of explained variation, whereas R-squared is the absolute amount of variation as a proportion of total variation.

R-Squared has an advantage that it can prevent you from over fitting a model. An overfit model contains too many predictors and it starts to model the random noise. Because it is impossible to predict random noise, the predicted R-squared must drop for an overfit 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 sum of square of difference of each data point from the mean value of all the values of target variable (y).

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

The residual sum of squares (RSS) measures the level of variance in the error term, or residuals, of a regression model.

$TSS = ESS + RSS$, where TSS is Total Sum of Squares, ESS is Explained Sum of Squares and RSS is Residual Sum of Squares

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

Answer: Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting. Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

4. What is Gini-impurity index?

Answer: Gini Impurity is a measurement used to build Decision Trees to determine how the features of a dataset should split nodes to form the tree.

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

Answer: Yes, unregularized decision-trees prone to overfitting., especially when a tree is particularly deep. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions. This small sample could lead to unsound conclusions.

6. What is an ensemble technique in machine learning?

Answer: Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model . To better understand this definition lets take a step back into ultimate goal of machine learning and model building

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

Answer: Bagging is a technique for reducing prediction variance by producing additional data for training from a dataset by combining repetitions with combinations to create multi-sets of the original data. Boosting is an iterative strategy for adjusting an observation's weight based on the previous classification

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

Answer: The out-of-bag (OOB) error is the average error for each calculated using predictions from the trees that do not contain in their respective bootstrap sample. This allows the RandomForestClassifier to be fit and validated whilst being trained.

9. What is K-fold cross-validation?

Answer: K-fold Cross-Validation is when the dataset is split into a K number of folds and is used to evaluate the model's ability when given new data. K refers to the number of groups the data sample is split into. For example, if you see that the k-value is 5, we can call this a 5-fold cross-validation.

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

Answer: Hyperparameter tuning consists of finding a set of optimal hyperparameter values for a learning algorithm while applying this optimized algorithm to any data set. That combination of hyperparameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors

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

Answer: If learning rate is too large, gradient descent can overshoot the minimum. It may fail to converge and even diverge.

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

Answer: No, logistic regression only forms linear decision surface, and in Non-Linear data the target label has no linear correlation with the features. In such cases, logistic regression (or linear regression for regression problems) can't predict targets with good accuracy (even on the training data).

13. Differentiate between Adaboost and Gradient Boosting.

Answer: AdaBoost is the first designed boosting algorithm with a particular loss function. On the other hand, Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem. This makes Gradient Boosting more flexible than AdaBoost.

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

Answer: In statistics and machine learning, the bias-variance tradeoff is the property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameters.

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

Answer: Linear Kernel is used when the data is Linearly separable, that is, it can be separated using a single Line. It is one of the most common kernels to be used. It is mostly used when there are a Large number of Features in a particular Data Set.

The polynomial kernel is a kernel function commonly used with support vector machines (SVMs) and other kernelized models, that represents the similarity of vectors (training samples) in a feature space over polynomials of the original variables, allowing learning of non-linear models.

RBF (Radial Basis Function) is more complex and efficient at the same time that it can combine multiple polynomial kernels multiple times of different degrees to project the non-linearly separable data into higher dimensional space so that it can be separable using a hyperplane.
