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?

Ans-

R2 is better than RSS because it is scale invariant. RSS value might vary according to scale of target. Eg: hypothetically, if target variable is distance covered and scale taken was in hundred of meter i.e. target 1,2,3 then we might get rss about 0.2, but if the scale was taken in meter i.e. target 100,200,300 and so on, then we might get larger rss of 200. R2 gives us the degree of variability in the target variable i.e. explained by the model or independent variables. Eg: If the value is 0.7, then it means that the independent variables explain 70% of the variation in the target variable. Since value always lies between 0 and 1, it becomes more easier than RSS to determine whether the model actually fits better or not.

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.

Ans-

The total sum of squares (TSS) measures how much variation there is in the observed data, while the residual sum of squares(RSS) measures the variation in the error between the observed data and modeled values. The explained sum of squares (ESS) measures how much variation there is in the modeled values. Total sum of squares (TSS) = explained sum of squares (ESS) + residual sum of squares (RSS).

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

Ans –

Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.

4. What is Gini–impurity index?

Ans –

Gini Index is the amount of probability of a specific feature that is classified incorrectly when selected randomly. If all the elements are linked with a single class then it can be called pure. It is calculated by subtracting the sum of the squared probabilities of each class from one.

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

Ans-

Unregularized decision trees are prone to overfitting, when the tree runs deep because overfitting happens when the learning processing overly optimizes the training error at the expense of test error. In order to overcome this issue, we can prune the decision tree either by setting the max depth of the tree or by setting minimum data points in each node.

6. What is an ensemble technique in machine learning?

Ans-

Ensemble technique is a process of combining several machine learning models into one model thus increasing its overall prediction accuracy.

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

Ans-

In bagging algorithm weak model learns independently in parallel from each other and then their learning is combined to give an average output, whereas in boosting algorithm weak model learns sequentially which means learning of one model is fed to the next model and so on

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

Ans –

Out-of-bag (OOB) error is a method of measuring the prediction error in random forest. To calculate OOB error, every data point is passed for prediction to tree where it would be behaving as OOB. After aggregated prediction is recorded for each row, the OOB score is computed as the number of correctly predicted rows from the out-of-bag sample whereas OOB Error is calculated as the number of wrongly classified OOB Sample.

9. What is K-fold cross-validation?

Ans-)

K-fold cross-validation is a method of resampling the data set in order to evaluate a model where the parameter K refers to the number of different subsets that the given data set is to be split into. K-1 subsets are used to train the model and the left out subsets are used as a validation set

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

Ans –

Hyperparameter tuning is used to find a set of optimal hyperparameter values for a learning algorithm and then applying this optimized algorithm on the data set. It maximizes the accuracy of the model.

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

Ans –

Learning rate parameter determines how fast or slow we will move towards the optimal weights. If the learning rate is very large we will skip the optimal solution.

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

Ans –)

The Logistic regression for classifier uses hyperplane to separate observations that belong to a class from all the other observations that do not belong to that class. The decision boundary used for this purpose is linear in nature. Hence non linear data cannot be perfetly classified using this approach.

13. Differentiate between Adaboost and Gradient Boosting.

Ans –

Adaboost trees are usually grown as decision stumps whereas gradient boost trees are grown to a greater depth usually ranging from 8 to 32 terminal nodes. Each classifier in Adaboost has different weights assigned to the final prediction based on its performance whereas in gradient boost all classifiers are weighed equally.

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

Ans –

When building a machine learning model, it is required to make a balance between bias and variance errors in order to avoid overfitting and underfitting and this balance between the bias error and variance error is known as the Bias-Variance trade-off.

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

Ans -

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.
Radial Basis Function (RBF) kernel function uses radial basis method to improve the transformation when there is no prior knowledge about data.
Polynomial kernel looks at the given features of input samples to determine their similarity, and also combinations of these features. It represents the similarity of vectors in the training set of data in a feature space over polynomials of the original variables used in the kernel.