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:- R-squared is generally a better measure of the goodness of fit for a regression model than the residual sum of squares (RSS). , Because it is a statistical measure that represents the proportion of the variance for the dependent variable that's explained by the independent variables in the 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:- In statistics, the explained sum of squares (ESS), alternatively known as the model sum of squares or sum of squares due to regression (SSR – not to be confused with the residual sum of squares (RSS) or sum of squares of errors), is a quantity used in describing how well a model, often a regression model, represents the data being modelled. In particular, the explained sum of squares measures how much variation there is in the modelled values and this is compared to the total sum of squares (TSS), which measures how much variation there is in the observed data, and to the residual sum of squares, which measures the variation in the error between the observed data and modelled values.
   Proof of Sum of Squares TSS = ESS + RSS in Econometrics Mathematical Explanation.
   TSS (total sum of squares) is equal to ESS (explained sum of squares) plus RSS (residual sum of squares).

3. What is the need of regularization in machine learning?
   Answer:- Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it. Sometimes the machine learning model performs well with the training data but does not perform well with the test data.

4. What is Gini–impurity index?
   Answer:- Gini impurity index measures how often a randomly chosen element of a set would be incorrectly labeled if it were labeled randomly and independently according to the distribution of labels in the set.

5. Are unregularized decision-trees prone to overfitting? If yes, why?
   Answer:- If the training dataset is not representative, decision trees may overfit to the training data's idiosyncrasies, resulting in poor generalization.
   Without proper stopping rules, decision trees may grow excessively, perfectly fitting the training data but failing to generalize well.

6. What is an ensemble technique in machine learning?
   Answer:-Ensemble learning combines multiple learners to improve predictive performance. It has been adopted in response to issues resulting from limited datasets.

7. What is the difference between Bagging and Boosting techniques?
   Answer:- Bagging: It is a homogeneous weak learners' model that learns from each other independently in parallel and combines them for determining the model average.
   Bagging decreases variance.

   Boosting: It is also a homogeneous weak learners' model but works differently from Bagging. In this model, learners learn sequentially and adaptively to improve model predictions of a learning algorithm.
   Boosting decreases bias.

8. What is out-of-bag error in random forests?
   Answer:- 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). Bagging uses subsampling with replacement to create training samples for the model to learn from.

9. What is K-fold cross-validation?
   Answer:-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 and why it is done?
Answer:- 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.

11. What issues can occur if we have a large learning rate in Gradient Descent?
    Answer:- When the learning rate is too large, gradient descent can suffer from divergence. This means that weights increase exponentially, resulting in exploding gradients which can cause problems such as instabilities and overly high loss values.

12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?
    Answer:- Logistic and softmax classification can be turned into non-linear classifiers simply by changing the representation of the input. Instead of passing the inputs xn directly to the algorithm, we can pass non-linear functions of the input Φ(xn).

13. Differentiate between Adaboost and Gradient Boosting.
    Answer:- Basic Concept

    AdaBoost (Adaptive Boosting):
    Focuses on adjusting the weights of misclassified instances in the training set. It combines multiple weak learners (often decision stumps) to create a strong classifier.
    It sequentially adds classifiers, giving more weight to instances that were misclassified by previous classifiers.
    Gradient Boosting:
    Builds models in a stage-wise fashion, where each new model is trained to correct the errors made by the previous models.
    It optimizes a loss function by fitting new models to the residual errors of the existing ensemble.

14. What is bias-variance trade off in machine learning?
Answer:-In machine learning, the bias–variance tradeoff describes the relationship between a model's complexity, the accuracy of its predictions, and how well it can make predictions on previously unseen data that were not used to train the model. In general, as we increase the number of tunable parameters in a model, it becomes more flexible, and can better fit a training data set. It is said to have lower error, or bias.

15. Give short description each of Linear, RBF, Polynomial kernels used in SVM.
Answer:- In machine learning, the radial basis function kernel, or RBF kernel, is a popular kernel function used in various kernelized learning algorithms. In particular, it is commonly used in support vector machine classification.

The linear kernel produces a decision boundary that is a hyperplane in the feature space. This hyperplane separates data points from different classes in a linear fashion.

In machine learning, 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.

FLIP ROBO