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. **RSS** :- Residual Sum of Square (RSS) is the sum of squares of error terms and this is an absolute number. This number can vary greatly based on the number of data points. RSS will increase if the data points are more and the value will decrease If the data points are less.

R-squared *R-square* is a measure of variance for dependent variables. That is Variance in the output that is explained by the small change in input.

The value of *R-square* is always between 0 (0%) and 1 (100%). The bigger the value better the fit.

—

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. **TSS** :- The total sum of squares (TSS) is a variation from the actual value from the Mean value of data points.

ESS :- Sum of squared difference between the predicted Y and the mean of Y.

ESS

tells us how much of the variation in the dependent variable our model explained.

RSS :- Residual Sum of Square (RSS) is the sum of squares of error terms and this is an absolute number. This number can vary greatly based on the number of data points RSS will increase if the data points are more and the value will decrease If data points are less.

Equation :-..

/

TSS = ESS + RSS

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

Ans. When training a machine learning model, the model can be easily overfitted or under fitted. To avoid this, we use regularization in machine learning to properly fit the model to our test set. Regularization techniques help us to reduce the overfitting problem and help us obtain an optimal model

Regularization techniques:

1. Lasso
2. Ridge
3. Elasticnet

4. What is Gini–impurity index?

Ans. It will check every feature. How much impurity each feature have?

Less the impurity more the information.

Gini indexing find the gini impurity.if the feature have less impurity.it selects the root

Node

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

Ans. Decision trees are 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?

Ans. Ensembles techniques has similar underlying idea where we aggregate predictions from group of predictors,which may be regressor or classifier and most of the time prediction is better than one obtained using a single predictor.such algorithms are called ensemble techniques.and such predictors are called ensembles.

There are two types of ensemble techniques:

1. Bagging
2. Boosting

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

Ans. **Bagging** :- 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 :- Boosting is an iterative strategy for adjusting an observation's weight based on the previous classification. It attempts to increase the weight of

an observation if it was erroneously categorized. Boosting creates good predictive models in general.

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

Ans. 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?

Ans. K-Fold Cross Validation

In this method, **we split the data-set into k number of subsets(known as folds) then we perform training on the all the subsets but leave one(k-1) subset for the evaluation of the trained model**. In this method, we iterate k times with a different subset reserved for testing purpose each time

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

Ans. 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?

Ans. A learning rate that is too large can **cause the model to converge too quickly to a suboptimal solution**, whereas a learning rate that is too small can cause the process to get stuck.

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

Ans. It can only be used to predict discrete functions. Hence, the dependent variable of Logistic Regression is bound to the discrete number set. It is very fast

at classifying unknown records. **Non-linear problems can't be solved with logistic regression because it has a linear decision surface**.

13. Differentiate between Adaboost and Gradient Boosting.

Ans. 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?

Ans. 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.

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. It is mostly used when there are a Large number of Features in a particular Data Set.

RBF Kernel is popular because of its similarity to K-Nearest Neighborhood Algorithm. It has the advantages of K-NN and **overcomes the space complexity problem** as RBF Kernel Support Vector Machines just needs to store the support vectors during training and not the entire dataset.

polynomial kernel is a general representation of kernels with a degree of more than one. It’s useful for image processing.