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?
The residual sum of squares is used to help you decide if a statistical model is a good fit for your data. It measures the overall difference between your data and the values predicted by your estimation model. Total SS is related to the total sum and explained sum with the following formula
[Total SS = Explained SS + Residual Sum of Squares.]
"Goodness-of-fit" is a mathematical model that describes the differences between the observed values and the expected values or how well the model fits a set of observations. This measure can be used in statistical hypothesis testing.

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.
The total sum of square is a variation of values of a dependent variable from the sample mean of dependent variable. The sum of squares quantifies the total variations in sample.

TSS = Total Sum of Squares
n = Number of observation
yi = Value in sample
y-hat = Mean value of sample

Regression sum of squares describe how well a regression model represents the modeled data. A higher regression sum of squares indicates that the model does not fit the data as well. Regression sum of squares also known as sum of square due to regression or explained sum of squares

Explained SS = Σ(Y-Hat – mean of Y)2.

The Residual sum of square essentially measures the variation of modeling error; it depicts how the variation in dependent variable in a regression model cannot be explained by the model. Generally a lower residual sum of squares indicate that the regression model can better explain the data while a higher residual sum of square indicates that the model poorly explain the data.

Residual Sum of Squares = Σ e2

3. What is the need of regularization in machine learning?
Regularization is the technique that is used to reduce the error by fitting a function appropriately on the given training set and avoiding overfitting.

4. What is Gini–impurity index?
It measures the degree or probability of particular variable being wrongly classified when it is randomly chosen.

5. Are unregularized decision-trees prone to overfitting? If yes, why?
Yes, Decision tree are prone to overfitting especially when a tree is particularly deep .This is due to amount of specifisity we look at leading to a smaller sample of events that meet the previous assumption. This small sample could lead to unsound conclusion.

6. What is an ensemble technique in machine learning?
Ensemble methods are techniques that create a multiple model and then combine them to produce improved result. Ensemble methods usually produce more accurate

solution than single model would.

7. What is the difference between Bagging and Boosting techniques?
Bagging is a method of merging same type of prediction and boosting is method of merging different types of prediction. Bagging decreases the variance not bias and solves overfitting issues in a model. Boosting decreases bias not variances. In bagging each model receive equal weight and in boosting weights are based on their performance.

8. What is out-of-bag error in random forests?
The out of bag error is the average error for each, calculated using prediction from the trees that do not contain in their respective bootstrap sample.

9. What is K-fold cross-validation?
Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into Cross-validation, sometimes called rotation estimation or out-of-sample testing, is any of various similar model validation techniques for assessing how the results of a statistical analysis will generalize to an independent data set.

10. What is hyper parameter tuning in machine learning and why it is done?
In machine learning hyperparameter optimization or tuning is the problem of choosing a set of optimal Hyperparameter for a learning algorithm. A hyperparameter is a parameter whose value is used to control the learning process. By contrast, the values of other parameters are learned.

11. What issues can occur if we have a large learning rate in Gradient Descent?
When the learning rate is too large, gradient descent can inadvertently increase rather than decrease the training error.

12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?
Logistic regression is considered a generalized linear model because the outcome always depends on the sum of the inputs and parameters. The output cannot depend on the product of its parameters.

13. Differentiate between Adaboost and Gradient Boosting.
AdaBoost is the first designed boosting algorithm with a particular loss function. Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modeling problem. This makes Gradient Boosting more flexible than AdaBoost.

14. What is bias-variance trade off in machine learning?
Bias is the simplifying assumptions made by the model to make the target function easier to approximate Variance is the amount that the estimate of the target function will change given different training data. Trade-off is tension between the error introduced by the bias and the variance.

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

RBF Kernel - RBF kernel is a popular kernel function used in various kernelized learning algorithms.it is commonly used in support vector machine classification.

Polynomial Kernel - In machine learning, the polynomial kernel is a kernel function commonly used with support vector machines and other kernelized models, that represents the similarity of vectors in a feature space over polynomials of the original variables, allowing learning of non-linear models.