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?

Typically, however, a smaller or lower value for the RSS is ideal in any model since it means there's less variation in the data set. In other words, the lower the sum of squared residuals, the better the regression model is at explaining the data.

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 squares (TSS) measures how much variation there is in the observed data,
while **the residual sum of squares** measures the variation in the error between the observed data and modeled values.
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, in a standard regression model
__RELATIONSHIP-__ **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?

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**.

4. What is Gini–impurity index?

Gini Index, also known as Gini impurity, **calculates 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.

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

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?

Ensemble methods are **techniques that aim at improving the accuracy of results in models by combining multiple models instead of using a single model**. The combined models increase the accuracy of the results significantly. This has boosted the popularity of ensemble methods in machine learning.

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

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?

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?

k-fold cross validation is **a procedure used to estimate the skill of the model on new data**.

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

Hyperparameters are **the parameters that are explicitly defined to control the learning process before applying a machine-learning algorithm to a dataset**. These are used to specify the learning capacity and complexity of the model.

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

This 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. If it is too small we will need too many iterations to converge to the best values. So using a good learning rate is crucial.

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

Logistic Regression has traditionally been used as a linear classifier, i.e. when the classes can be separated in the feature space by linear boundaries. The blog is about using Logistic Regression when the separation/decision boundaries may be nonlinear.

13. Differentiate between Adaboost and Gradient Boosting.

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?

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.

SVM algorithms use a set of mathematical functions that are defined as the kernel. The function of kernel is to take data as input and transform it into the required form.

Different SVM algorithms use different types of kernel functions. These functions can be different types. For example *linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid.*

The most used type of kernel function is **RBF.** Because it has localized and finite response along the entire x-axis.