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?

R-squared, The residual sum of squares (RSS) is the absolute amount of explained variation, whereas R-squared is the absolute amount of variation as a proportion of total variation.

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.

RSS: Residual is used to find the best fit line. Every data point has a residual value which is the difference between the actual value and the predicted value (the value of point on line).

TSS: TSS is the sum of square of difference of each data point from the mean value of all the values of target variable

3. What is the need of regularization in machine learning?
Constraining a model to make it simpler and reduce the risk of overfitting is called regularization.
Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

4. What is Gini–impurity index?
The Entropy and Information Gain method focuses on purity and impurity in a node. The Gini Index or Impurity measures the probability for a random instance being misclassified when chosen randomly. The lower the Gini Index, the better the lower the likelihood of misclassification.

5. Are unregularized decision-trees prone to overfitting? If yes, why?
Decision Trees make very few assumptions about the training data If left unconstrained, the tree structure will adapt itself to the training data, fitting it very closely, and most likely overfitting it. To avoid overfitting the training data, you need to restrict the Decision Tree's freedom during training.

6. What is an ensemble technique in machine learning?
The ensemble methods in machine learning combine the insights obtained from multiple learning models to facilitate accurate and improved decisions. These methods follow the same principle as the example of buying an air-conditioner cited above.

In learning models, noise, variance, and bias are the major sources of error. The ensemble methods in machine learning help minimize these error-causing factors, thereby ensuring the accuracy and stability of machine learning (ML) algorithms.

7. What is the difference between Bagging and Boosting techniques?
Bagging and boosting are different ensemble techniques that use multiple models to reduce error and optimize the model. The bagging technique combines multiple models trained on different subsets of data, whereas boosting trains the model sequentially, focusing on the error made by the previous model

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

9. What is K-fold cross-validation?
In this approach, called *k*-fold CV, the training set is split into *k* smaller sets The following procedure is followed for each of the *k* "folds":

- A model is trained using $k-1$ of the folds as training data;
- the resulting model is validated on the remaining part of the data (i.e., it is used as a test set to compute a performance measure such as accuracy).

The performance measure reported by *k*-fold cross-validation is then the average of the values computed in the loop.

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

A Machine Learning model is defined as a mathematical model with several parameters that need to be learned from the data. By training a model with existing data, we can fit the model parameters.
However, there is another kind of parameter, known as Hyperparameters, that cannot be directly learned from the regular training process. They are usually fixed before the actual training process begins. These parameters express important properties of the model such as its complexity or how fast it should learn.

11. What issues can occur if we have a large learning rate in Gradient Descent?
if the learning rate is too high. This might make the algorithm diverge, with larger and larger values, failing to find a good solution

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

Logistic regression is an algorithm that learns a model for binary classification. A nice side-effect is that it gives us the probability that a sample belongs to class 1 (or vice versa: class 0)

13. Differentiate between Adaboost and Gradient Boosting.

AdaBoost :The method automatically adjusts its parameters to the data based on the actual performance in the current iteration. Meaning, both the weights for re-weighting the data and the weights for the final aggregation are re-computed iteratively.

Gradient Boosting works by sequentially adding predictors to an ensemble, each one correcting its predecessor. However, instead of tweaking the instance weights at every iteration like AdaBoost does, this method tries to fit the new predictor to the residual errors made by the previous predictor.

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

If our model is too simple and has very few parameters then it may have high bias and low variance. On the other hand if our model has large number of parameters then it’s going to have high variance and low bias. So we need to find the right/good balance without overfitting and underfitting the data.

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

Linear SVMs are used with linearly separable data; this means that the data do not need to undergo any transformations to separate the data into different classes.