Machine Learning

1. R-squared and Residual sum of Squares(RSS) are two measures used to evaluate the goodness of fit of a regression model.
   R-squared is statistical find how the data are to the fitted regression line. It ranges from 0 to 1, with higher value indicate a better fit model.
   Residual sum of square is measure of the difference between the data and predicted value.it is calculated as sum of the squared differences between the predicted values and the actual values. lower RSS indicates a better fit.
   R-squared is commonly used in practice because it is easy to interpret.

2. TSS: Total Sum of Squares (TSS or SST) is defined as the sum of overall observations, of the squared differences of each observation from the overall mean.
   ESS: Explained sum of square (ESS) is a statistical quantity used in modeling of a process. It tell how much of the variation between observed data and predicted data is being explained by the model proposed.
   RSS: The residual sum of squares(RSS) calculates the degree of variance in regression model. It's estimates the erro in model prediction. Smaller the residual sum of squares, the better your model fits your data.

   TSS = ESS + RSS
   This equation states that the total variation in the dependent variable is the sum of the variation explained by the model and the variation not explained by the model.

3. Regularization is a technique used to reduce the error by fitting the function appropriately on the given training dataset and avoid overfitting. Overfitting can cause poor generalization performance on new, unseen data. regularization reduce the risk of overfitting by adding a penalty term to the loss function. There are 2 type of regularization techniques L1 and L2.
   L1(LASSO): Lasso Regression adds "absolute value of magnitude" of coefficient as penalty term to the loss function

L2(Ridge Regression): Ridge regression adds “squared magnitude” of coefficient as penalty term to the loss function

4. Gini impurity is a measure of the impurity of a set of samples. It is commonly used in decision tree algorithms as a measure of the quality of a split in the data. Gini Impurity tells us what is the probability of misclassifying an observation. It helps to find out the root node, intermediate nodes and leaf node to develop the decision tree.

5. Are unregularized decision-trees prone to overfitting. Decision trees are type of model that splits the data into subset on the value of the input features.

6. Ensemble learning is a technique in machine learning which takes the help of several base models and combines their output to produce an optimized model. This type of machine learning algorithm helps in improving the overall performance of the model
   There are different ensemble methods:
   - Bagging
   - Boosting

7. Bagging: The primary goal of bagging ensemble method is to minimize variance error in decision trees. Randomly create subsets of training dataset with replacement. The subsets are then used for training decision tree. Which reduce variance ,as the average prediction generated from different model is much more reliable an robust than single model.
   Boosting: an iterative ensemble technique, it’s adjust an observation weight based on it’s last classification. In case observation is incorrectly classified, boosting increases the weight and vice versa. Boosting reduce bias errors and produce superior predictive models.

8. Random Forest is one of the machine learning algorithms that use bootstrap aggregation. Random Forest aggregates the result of several decision trees. Decision Trees are known to work well when they have

small depth otherwise they overfit. The prediction of the Random Forest is then a combination of the individual prediction of the decision trees. The out-of-bag error is the average error for each predicted outcome calculated using predictions from the trees that do not contain that data point in their respective bootstrap sample. This way, the Random Forest model is constantly being validated while being trained.

9. Cross-validation is a resampling technique used to evaluate machine learning models on a limited data sample. The technique has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the technique is called as k-fold cross-validation. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as k=10 becoming 10-fold cross-validation

   General process of K-fold cross validation:
   - The whole dataset is randomly split into independent k-folds without replacement.
   - k-1 folds are used for the model training and one fold is used for performance evaluation.
   - This procedure is repeated k times (iterations) so that we obtain k number of performance estimates for each iteration.
   - Then we get the mean of k number of performance estimates.

10. Hyperparameter tuning is the process of selecting the best set of parameters for machine learning model. Hyperparameters, that cannot be directly learned from the regular training process. They are usually fixed before the actual training process begins.
    - Grid Search: set up a grid made up of hyperparameters and their different values. For each possible combination, a model is trained and a score is produced on the validation data.
    - Random Search: similar to grid search, but instead of training and scoring on each possible hyperparameter combination, random combinations are selected. You can set the number of search iterations based on time and resource constraints.

11. a gradient and a learning rate. The learning rate gives you control of how big (or small) the updates are going to be. A bigger learning rate means bigger updates and, hopefully, a model that learns faster.

12.

Ada Boost	Gradient Boosting
AdaBoost, it minimize the exponential loss function that can make the algorithm sensitive to the outliers.	Gradient Boosting, any differentiable loss function can be utilised.
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 modelling problem.
AdaBoost minimize loss function related to any classification error and is best used with weak learners	Gradient Boosting is used to solve the differentiable loss function problem.
The method was mainly designed for binary classification problems and can be utilised to boost the performance of decision trees.	The technique can be used for both classification and regression problems.
AdaBoost is considered as a special case of Gradient boost in terms of loss function, in which exponential losses.	Concepts of Grtdients are more general in nature

13. Logistic Regression is used for Binary Classification, It is linear model. If data data is non liner logistic regression may not be predict accurately.

14. Bias Variance Tradeoff is a design consideration when training the machine learning model. Certain algorithms inherently have a high bias and low variance and vice-versa. In this one, the concept of bias-variance tradeoff is clearly explained so you make an informed decision when training your ML models

Bias is the inability of a machine learning model to capture the true relationship between the data variables. It is caused by the erroneous assumptions that are inherent to the learning algorithm.
You need to find a good balance between the bias and variance of the model we have used. This tradeoff in complexity is what is referred to as bias and variance tradeoff. An optimal balance of bias and variance should never overfit or underfit the model.

15. Support Vector Machine (SVM) is a relatively simple Supervised Machine earning Algorithm used for classification and/or regression. It is more preferred for classification but is sometimes very useful for regression as well.
    - Linear kernel: Linearly Separable Data is any data that can be plotted in a graph and can be separated into classes using a straight line. The linear kernel is the simplest kernel function.
    - RBF kernel: RBF is used when data is not linearly separatable. The similarity between two points in the transformed feature space is an exponentially decaying function of the distance between the vectors and the original input space as shown below. RBF is the default kernel used in SVM.
    - Polynomial kernel: The Polynomial kernel takes an additional parameter, ‘degree’ that controls the model’s complexity and computational cost of the transformation