Answer 1. R-squared measures the goodness of fit of a regression model. Hence, a higher R-squared indicates the model is a good fit while a lower R-squared indicates the model is not a good fit.

Whereas, Adjusted R2 is a corrected goodness-of-fit (model accuracy) measure for linear models. It identifies the percentage of variance in the target field that is explained by the input or inputs. R2 tends to optimistically estimate the fit of the linear regression.

Answer 2. Explained sum of square (ESS) or Regression sum of squares or Model sum of squares is a statistical quantity used in modeling of a process. ESS gives an estimate of how well a model explains the observed data for the process.

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.

Residual Sum of Squares (RSS) is a statistical method used to measure the deviation in a dataset unexplained by the regression model. it measures the variance in the value of the observed data when compared to its predicted value as per the regression model. Hence, RSS indicates whether the regression model fits the actual dataset well or not.

The equation relating these three metrics with each other is :-

$$TSS = RSS + ESS$$

Answer 3. Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it.

Sometimes the machine learning model performs well with the training data but does not perform well with the test data. It means the model is not able to

predict the output when deals with unseen data by introducing noise in the output, and hence the model is called overfitted. This problem can be deal with the help of a regularization technique.

Answer 4. The Gini index is used **to indicate the inequality of a nation**. Greater the value of the index, higher would be the inequality. The index is used to determine the differences in the possession of the people. The Gini Coefficient is a measure of inequality.

It is also known as Gini Impurity and it is **calculated by subtracting the sum of the squared probabilities of each class from one**. It favours mostly the larger partitions and is very simple to implement. In simple terms, it calculates the probability of a certain randomly selected feature that was classified incorrectly.

Answer 5. 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.

Answer 6. Ensemble **techniques 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.

Bagging and Boosting are two of the most used techniques in machine learning.

Answer 7. Difference between Boosting and Bagging:-

S.NO	Bagging	Boosting
1.	The simplest way of combining predictions that belong to the same type.	A way of combining predictions that belong to the different types.
2.	Aim to decrease variance, not bias.	Aim to decrease bias, not variance.
3.	Each model receives equal weight.	Models are weighted according to their performance.
4.	Each model is built independently.	New models are influenced by the performance of previously built models.
5.	Different training data subsets are selected using row sampling with replacement and random sampling methods from the entire training dataset.	Every new subset contains the elements that were misclassified by previous models.
6.	Bagging tries to solve the over-fitting problem.	Boosting tries to reduce bias.
7.	If the classifier is unstable (high variance), then apply bagging.	If the classifier is stable and simple (high bias) the apply boosting.
8.	In this base classifiers are trained parallelly.	In this base classifiers are trained sequentially.
9	Example: The Random forest model uses Bagging.	Example: The AdaBoost uses Boosting techniques

Answer 8. 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.

Answer 9. K-fold Cross-Validation is **when the dataset is split into a K number of folds and is used to evaluate the model's ability when given new data**. K refers to the number of groups the data sample is split into. For example, if you see that the k-value is 5, we can call this a 5-fold cross-validation.

Answer 10. 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.

It takes advantage of the processing infrastructure of Google Cloud **to test different hyperparameter configurations when training your model**. It can give you optimized values for hyperparameters, which maximizes your model's predictive accuracy.

Answer 11. 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.

Answer 12. Logistic regression is considered a generalized linear model because the outcome **always** depends on the **sum** of the inputs and parameters. Or in other words, the output cannot depend on the product (or quotient, etc.) of its parameters!

Answer 13. The Comparison

Loss Function:

The technique of Boosting uses various loss functions. In case of Adaptive Boosting or AdaBoost, it minimises the exponential loss function that can make the algorithm sensitive to the outliers. With Gradient Boosting, any differentiable loss function can be utilised. Gradient Boosting algorithm is more robust to outliers than AdaBoost.

Flexibility

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.

Benefits

AdaBoost minimises loss function related to any classification error and is best used with weak learners. The method was mainly designed for binary classification problems and can be utilised to boost the performance of decision trees. Gradient Boosting is used to solve the differentiable loss function problem. The technique can be used for both classification and regression problems.

Answer 14. 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.

So we can say, 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**.

Answer 15. SVM- 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 – It is **a popular kernel function used in various kernelized learning algorithms**. In particular, it is commonly used in support vector machine classification.

Polynomial kernel – It is a kernel function commonly used with support vector machines (SVMs) and other kernelized models, that represents the similarity of vectors (training samples) in a feature space over polynomials of the original variables, allowing learning of non-linear models.