1. R-squared. This is because it is not dependent on the scale of the data, the number observation and easy to compare models across distinguished data set.
2. a. Total Sum of Squares (TSS): This is a measure of total variation in y about the sample mean

   $$\sum (y_i - \bar{y})^2$$

   mean.

   b. Explained Sum of Squares(ESS): This total variation in y, about the sample mean, that is explained by, or due to regression.

   $$\sum (\hat{y}_i - \bar{y})^2$$

   c. Residual Sum of Squares (RSS): This is the part of the total variation in y about its mean that is not explained by the regression.

   $$\sum \hat{e}_i^2$$

   TSS = ESS +RSS
3. What is Regularization: This is the process of searching for a hypothesis that directly minimizes the weighted sum of empirical loss and the complexity of the hypothesis. This process directly addresses overfitting by explicitly penalizes complex hypothesis.
4. What is Gini–impurity index?

   This is a metric used to measure the impurity or disorder of a dataset, especially as it relates to decision trees in machine learning. It is used to learn how of frequent a randomly chosen element from the dataset would be improperly labelled if it were randomly labelled according to the distribution of labels in the dataset. The Gini impurity is specifically useful for binary and multiclass classification problems.

   Definition

   For a dataset with n classes, the Gini impurity GGG is calculated as follows:

   $$G = 1 - \sum_{i=1}^{n} p_i^2$$

   where p_i is the probability (or proportion) of selecting a class i from the dataset.

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

   Answer: Yes, because they are susceptible to producing complex models that fit the training data set completely.
6. What is an ensemble technique?

This is a learning method that select a collection, or ensemble, of hypotheses, $h_1, h_2 \ldots, h_n$, and combine their predictions by averaging, voting, or by another level of machine learning to reduce variance and bias. These are combinations of the outcomes produced by two or more analytics models into a compound output. Ensembles are primarily used for prediction modelling where the scores of two or more models are combined to produce a better prediction.

7. Bagging vs Boosting techniques

   Bagging is the simplest and most common type of ensemble method; it builds multiple prediction models (e.g., decision trees) from bootstrapped/resampled data and combines the predicted values through averaging or voting while Boosting is an ensemble method where a series of prediction models are built progressively to improve the predictive performance of the cases/samples incorrectly predicted by the previous ones.

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

   This is an important concept in the context of random forests, which are an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) or the mean prediction (regression) of the individual trees. By using the OOB error, you obtain a reliable estimate of the random forest's generalization error without needing a separate test set.

9. k-fold cross-validation: This is a popular accuracy assessment technique for prediction models where the complete data set is randomly split into k mutually exclusive subsets of approximately equal size. The classification model is trained and tested k times. Each time it is trained on all but one fold and then tested on the remaining single fold. The cross-validation estimate of the overall accuracy of a model is calculated by simply averaging the k individual accuracy measures.

10. Hyperparameter Tuning: This is the process of selecting the best set of hyperparameters for a learning algorithm

    Reasons: For model generalization, training efficiency in terms of time and resource consumption and improve model accuracy and precision.

11. (a) Overshooting the Minimum (optimal point) and miss it entirely

    (b). loss function value increases instead of decreasing.

(C) Failure of the model to Converge into optimal values.

12. Yes, by using polynomial features, interaction terms and kernel method,
13. Adaboost adjusts weights, while Gradient Boosting adjusts the model itself based on gradient descent principles.
14. The bias-variance trade-off in machine learning refers to the delicate balance between a model's ability to capture the complexity of the underlying data (low bias) and its sensitivity to noise and fluctuations in the training data (low variance). Increasing model complexity typically reduces bias but increases variance, and vice versa. Achieving optimal performance involves managing this trade-off to generalize well to new, unseen data while capturing important patterns from the training data.

15(a) Linear Kernel: The simplest kernel, computes the dot product of the input features. It is Suitable for linearly separable data or when the number of features is large.

(b) RBF (Radial Basis Function) Kernel: It uses a Gaussian-like function to measure the similarity between two samples. It Effective for non-linear data by mapping features into a higher-dimensional space.

(c) Polynomial Kernel: It computes the similarity between samples using a polynomial function. It can capture complex non-linear relationships between features.