MACHINE LEARNING

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?**

Ans: Although they evaluate different aspects of the fit, R-squared and Residual Sum of Squares (RSS) are both indicators of a regression model's goodness of fit. The percentage of the dependent variable's variance that can be accounted for by the independent variables in the model is expressed as R-squared, or coefficient of determination. A higher number on the scale of 0 to 1 denotes a better fit. R-squared can be used to compare models or to calculate the percentage of the dependent variable's variability that the model explains. Conversely, after the model has been fitted, the residual sum of squares, or RSS, calculates the total amount of unexplained variation in the dependent variable. It is the total of the squared discrepancies between the dependent variable's actual and expected values. A better match is indicated by a lower RSS number. RSS is helpful in assessing how accurate the model's predictions are. Generally speaking, both metrics matter and ought to be taken into account when assessing a model's goodness of fit. Nonetheless, R-squared is sometimes regarded as a more accurate goodness of fit metric than RSS since it offers a single figure that more easily compares between models and more easily interprets the percentage of the dependent variable's variance that the model accounts for.

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.**

Ans: Equation: TSS=ESS+RSS
The variance in the modelled values is known as the explained sum of squares (ESS), the variation in the observed data is known as the total sum of squares (TSS), and the variation in the error between the modelled values and the observed data is known as the residual sum of squares (RSS) in regression. Stated differently, the variance in the target variable can be divided into two parts: the variation that the independent variables (ESS) explain and the variation that the model cannot explain (RSS). The best regression coefficients for the model, which can be used to forecast a new hire's salary based on their CGPA and IQ, can be found by minimising the RSS. In contrast, the ESS provides information that is crucial for assessing the model's performance: it indicates the proportion of the overall wage fluctuation that can be explained by the model.

3. **What is the need of regularization in machine learning?**

Ans: A machine learning model may quickly become overfit or underfit during training. To prevent this, we appropriately fit a model onto our test set using regularisation in machine learning. Regularisation approaches assist in obtaining an ideal model by lowering the likelihood of overfitting. In machine learning, regularisation is a technique used to improve a model's

generalisation performance by preventing overfitting. Regularisation is the process of modifying the magnitude and shrinkage of the coefficients of input parameters. The purpose of this is to uphold generalisation. Another useful tool for managing a linear regression model's complexity is regularisation. It accomplishes this by deducting coefficients from the forecast that are not significant. This can lower the variance of the model and increase its resistance to noise and anomalies. The capacity of a neural network to generalise can also be enhanced by regularisation approaches. By reducing superfluous complexity and exposing the network to a wider variety of data, they achieve this.

4. What is Gini–impurity index?

Ans: The likelihood that a randomly selected element in the dataset would be erroneously classified if its label were assigned at random in accordance with the dataset's class distribution is known as the Gini Impurity. A statistical metric known as the Gini index, or Gini impurity, determines the likelihood that a variable chosen at random would be incorrectly classified. The range of the Gini index is 0 to 1. A decreased probability of misclassification is indicated by a lower Gini index. To compute the Gini index, take each class's squared probability and subtract them all from one. Larger partitions are encouraged, and implementation is simple. When every case in the node fits into a single goal category, the Gini index hits its minimum, or zero. In terms of computational efficiency, the Gini index is superior to entropy.

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

Ans: Unregularized decision trees do indeed have an overfitting problem.
When a model fits the training set so closely that it is unable to understand the underlying behaviour of the data through generalisation, this is known as overfitting. Strong machine learning algorithms that can handle classification and regression issues are called decision trees. They are vulnerable to instability and overfitting, though. Although they are easier to understand and use, decision trees are prone to overfitting. The goal of regularisation approaches is to keep the tree from getting too complicated by making it simpler. Its capacity to generalise to fresh data is enhanced by this. If a decision tree is allowed to reach its maximum depth, it will always overfit the training set.

6. What is an ensemble technique in machine learning?

Ans: In machine learning, an ensemble approach is a multimodal system that combines many classifiers and methodologies in a deliberate manner to create a predictive model (classified as sequential model, parallel model, homogeneous and heterogeneous methods etc.). The ensemble technique also aids in minimising bias in the predictive model, reducing variation in the forecasted data, and improving the accuracy of statistical classification and prediction from complicated situations.

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

Ans: Bagging: Bagging is a technique that improves machine learning algorithms' performance and precision. The simplest method for grouping predictions of the same kind is bagging. Every model is treated equally when it comes to bagging. Every model in bagging is constructed independently. Overfitting is an issue that can be solved with the use of bagging. When the classifier is unstable, bagging is used.

Boosting: Boosting is a method that methodically modifies an observation's weight according to its prior classification. Combining many forms of forecasts is known as "boosting." The models' performance determines how important they are in boosting. In boosting, the way previously constructed models perform has an impact on how future models are constructed. Boosting helps to lessen prejudice. When the classifier reaches stability, it is boosted.

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

Ans: A technique for calculating the prediction error of machine learning models that employ bootstrap aggregating is called out-of-bag (OOB) error. Another name for it is the out-of-bag estimate. OOB error is the average mistake for every training sample in a random forest. Predictions from the trees that don't have the sample in their bootstrap sample are used to calculate it. The average of every training sample's prediction error is measured objectively by OOB error. It serves as a cross-validation within the company. The random forest classifier or regressor's oob_score_property in scikit-learn can be used to determine the OOB fault.

9. What is K-fold cross-validation?

Ans: In order to determine how well your machine learning model can predict the outcome of unknown data, cross validation is an evaluation technique used in machine learning. It is a well-liked approach since it is simple to understand, effective with small data sets, and provides a less biassed assessment. The term "K-fold Cross Validation" comes from the division of the data sample into "k" smaller samples. Terms like "fourfold cross validation" or "tenfold cross validation" may also be encountered; these refer to the process of dividing the sample data into four or ten smaller samples, respectively.

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

Ans: The process of choosing the ideal collection of hyperparameters for a machine learning model is known as hyperparameter tuning. Hyperparameters are configurations, such the number of neurons in a neural network or the learning rate, that regulate how the model learns. One of the most important stages in the model-development process is hyperparameter tweaking. The model's performance can be greatly affected by the selection of hyperparameters. By minimising a predetermined loss function, hyperparameter tuning aims to maximise the model's performance. This results in fewer errors and better outcomes. Similar to fine-tuning your car's settings to get optimal performance and fuel efficiency, hyperparameter tweaking

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

Ans: Problems with gradient descent may arise from an excessively high learning rate: Divergence: The gradients explode as a result of the weights increasing exponentially. Issues like instabilities and excessively high loss values may result from this. Overshooting the minimum: It's possible for the algorithm to exceed the minimum. Too rapid convergence could lead the model to a less-than-ideal outcome. The gradient descent algorithm's performance can be strongly impacted by the learning rate selection. There is no set value for the learning rate. For instance, you may create a rule that states that when training epochs rise, the learning rate would fall.

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

Ans: A linear classifier that generates a linear decision boundary is called a logistic regression. For categorical outcome variables, like death, it is employed. Since logistic regression implies a linear relationship between the input features and the output, it cannot be applied to non-linear issues. This indicates that the intricacy and non-linearity of the data are not captured by it. Trying different nonlinear functional forms for the logit function may yield better results if you suspect that the decision boundary is nonlinear. Regressors, which approximate a function between continuous X and Y values, are said by some to be a better method. Additionally, you can change your data using Linear Regression to approximate a function with a higher dimension than linear.

13. Differentiate between Adaboost and Gradient Boosting.

Ans: Both ensemble approaches, AdaBoost and Gradient Boosting, construct a series of models to fix mistakes in the prior model. They vary, nonetheless, in a few respects:
Flexibility: Compared to AdaBoost, Gradient Boosting is a more adaptable technique. Loss function: Gradient Boosting is a general technique, but AdaBoost is a special instance with a particular loss function. Iterations: Less iterations result in AdaBoost becoming stiffer. Underfitted values: Gradient Boosting attempts to fit the new predictor to the residual errors of the prior predictor, while AdaBoost modifies instance weights at each interaction. Voting weights: Gradient Boosting is primarily about "adding gradient optimisation," but AdaBoost is more about "voting weights." Overfitting: Since AdaBoost focuses more on "organising people to vote" than on actual voting, it does not overfit. AdaBoost is intended for binary classification problems and works best with weak learners. Decision trees can work better when this loss function associated with classification errors is minimised.
The differentiable loss function problem is resolved by the application of gradient boosting. To determine the outcome, it adds up the outcomes of each decision tree along the process.

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

Ans: A key idea in statistics and machine learning is the bias-variance trade-off. It explains the correlation between a model's complexity, prediction accuracy, and prediction performance on data that hasn't been seen before. The balance between bias and variance, two potential sources

of inaccuracy in a predictive model, is known as the bias-variance trade-off. Bias is the inaccuracy that results from the learning algorithm's excessively basic assumptions. The conflict a model experiences between its variance (ability to generalise successfully to new, unseen cases) and bias (ability to decrease errors on the training set) is known as the bias-variance trade off.

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

Ans: Kernel Linear
One kind of kernel function utilised in machine learning, particularly in Support Vector Machines (SVMs), is the linear kernel. This kernel function, which determines the dot product between the input vectors in the original feature space, is the most basic and widely used one.

Kernel of Polynomials
A polynomial kernel is a specific type of kernel function used in machine learning, such as in support vector machines, or SVMs. It is a nonlinear kernel function that moves the input data into a higher-dimensional feature space using polynomial functions.

RBF Gaussian Kernel
The radial basis function (RBF) kernel, sometimes referred to as the Gaussian kernel, is a widely used kernel function in machine learning, especially in support vector machines (SVMs). It is a nonlinear kernel function that uses a Gaussian function to transfer the input data into a higher-dimensional feature space**.**