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: R-squared (R^2) and Residual Sum of Squares (RSS) are both commonly used measures to assess the goodness of fit of a regression model, but they capture different aspects of model performance and the choice between them depends on the context and what you want to evaluate.

R-squared(R2) is often used when you want to understand the proportion of variance explained by the model, especially in the context of comparing different models. It can provide insights into how well the predictors collectively contribute to explaining the variation in the dependent variable.

RSS is useful when you want to evaluate the absolute goodness of fit, focusing on the magnitude of the prediction errors. If minimizing prediction errors is a primary concern, RSS is a more appropriate choice.

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:

TSS(Total Sum Of Squares): The total sum of squares (TSS) is the sum of squared differences between the observed dependent variables and the overall mean. It can be explained as the dispersion of the observed variables around the mean similar to the variance in descriptive statistics.

In general terms, the sum of squares is a statistical technique used in regression analysis to determine the dispersion of data points. In a regression analysis, the goal is to determine how well a data series can be fitted to a function that might help to explain how the data series was generated. The sum of squares is used as a mathematical way to find the function that best fits (varies least) from the data.

ESS(Explained Sum of Squares): The sum of squares due to regression (SSR) or explained sum of squares (ESS) is the sum of the differences between the predicted value and the mean of the dependent variable. In other words, it describes how well our line fits the data.

RSS(Residual Sum of Squares): The residual sum of squares (RSS) is a statistical technique used to measure the amount of variance in a data set that is not explained by a regression model itself. Instead, it estimates the variance in the residuals, or error term.

The RSS measures the amount of error remaining between the regression function and the data set after the model has been run. A smaller RSS figure represents a regression function that is well-fit to the data. It essentially determines how well a regression model explains or represents the data in the model.

Relation between TSS, ESS and RSS:

$$TSS = ESS + RSS$$

where TSS is Total Sum of Squares, ESS is Explained Sum of Squares and RSS is Residual Sum of Squares. The aim of Regression Analysis is to explain the variation of dependent variable Y.

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

Ans: While training a machine learning model, the model can easily be overfitted or under fitted. To avoid this, we use regularization in machine learning to properly fit a model onto our test set. Regularization techniques help reduce the chance of overfitting and help us get an optimal model.

4. **What is Gini–impurity index?**

Ans: The Gini impurity measure is one of the methods used in decision tree algorithms to decide the optimal split from a root node, and subsequent splits.
Gini Impurity tells us what is the probability of misclassifying an observation. Note that the lower the Gini the better the split. In other words, the lower the likelihood of misclassification.

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

Ans: Yes, unregularised decision tree have a tendency to overfit to the training set because they can keep growing deeper and more complex until they perfectly classify the training data. This can lead to the tree capturing noise in the data, rather than the underlying relationship, and thus performing poorly on new, unseen data.

6. **What is an ensemble technique in machine learning?**

Ans: Ensemble learning is a machine learning technique that enhances accuracy and resilience in forecasting by merging predictions from multiple models. It aims to mitigate errors or biases that may exist in individual models by leveraging the collective intelligence of the ensemble.

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

Ans: Bagging is the simplest way of combining predictions that belong to the same type while Boosting is a way of combining predictions that belong to the different types. Bagging aims to decrease variance, not bias while Boosting aims to decrease bias, not variance.

1. **Bagging**: It is a homogeneous weak learners' model that learns from each other independently in parallel and combines them for determining the model average.
2. **Boosting**: It is also a homogeneous weak learners' model but works differently from Bagging. In this model, learners learn sequentially and adaptively to improve model predictions of a learning algorithm.

8. **What is K-fold cross-validation?**

Ans: K-fold cross-validation is a technique for evaluating predictive models. The dataset is divided into k subsets or folds. The model is trained and evaluated k times, using a different fold as the validation set each time. Performance metrics from each fold are averaged to estimate the model's generalization performance.

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

Ans: When we are training machine learning models, each dataset and model needs a different set of hyper parameters, which are a kind of variable. The only way to determine these is through multiple experiments, where you pick a set of hyper parameters and run them through your model. This is called hyper parameter tuning.

Hyper parameter tuning allows data scientists to tweak model performance for optimal results. This process is an essential part of machine learning, and choosing appropriate hyper parameter values is crucial for success.

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

Ans: In machine learning, as we try to minimize one component of the error (e.g., bias), the other component (e.g., variance) tends to increase, and vice versa. Finding the right balance of bias and variance is key to creating an effective and accurate model. This is called the bias-variance tradeoff..

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

Ans:

Linear Kernel: A linear kernel is a type of kernel function used in machine learning, including in SVMs (Support Vector Machines). It is the simplest and most commonly used kernel function, and it defines the dot product between the input vectors in the original feature space. Where x and y are the input feature vectors.

RBF kernel: RBF SVM works by mapping the input data into a higher-dimensional feature space, where the classes can be separated by a hyperplane. The algorithm uses a kernel function, such as the Radial Basis Function, to measure the similarity between pairs of data points in the feature

Polynomial kernel: The polynomial kernel 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.

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

Ans: Choosing a proper learning rate can be difficult. A learning rate that is too small leads to painfully slow convergence, while a learning rate that is too large can hinder convergence and cause the loss function to fluctuate around the minimum or even to diverge.

If the learning rate is too high, the algorithm may overshoot the minimum, and if it is too low, the algorithm may take too long to converge. Overfitting Gradient descent can overfit the training data if the model is too complex or the learning rate is too high.