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?**

- **R-squared** is generally considered a better measure of goodness of fit because it represents the proportion of the variance in the dependent variable that is predictable from the independent variables. RSS measures the total deviation of the response values from the fit, but it does not provide a normalized measure and can be misleading when comparing models with different numbers of predictors.

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.**

- **TSS (Total Sum of Squares)**: Total variance in the data.
- **ESS (Explained Sum of Squares)**: Portion of TSS that is explained by the regression model.
- **RSS (Residual Sum of Squares)**: Portion of TSS that remains unexplained.

  Equation: TSS=ESS+RSS\text{TSS} = \text{ESS} + \text{RSS}TSS=ESS+RSS

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

- Regularization is needed to prevent overfitting by penalizing large coefficients in the model. It helps to keep the model simpler and improves its generalization to unseen data.

4. **What is the Gini-impurity index?**

- The Gini-impurity index is a measure of the impurity or disorder in a dataset, used in decision tree algorithms. It quantifies how often a randomly chosen element would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the subset.

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

- Yes, unregularized decision trees are prone to overfitting because they can create overly complex trees that fit the noise in the training data instead of capturing the true underlying patterns.

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

- An ensemble technique combines multiple machine learning models to improve overall performance. The idea is that by combining diverse models, the ensemble can reduce errors and increase robustness compared to individual models.

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

- **Bagging (Bootstrap Aggregating)**: Trains multiple models independently on different subsets of the data and averages their predictions. It reduces variance and helps in improving stability.
- **Boosting**: Trains models sequentially, each new model focuses on correcting errors made by previous models. It reduces bias and can create strong predictive models.

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

- Out-of-bag (OOB) error is an estimate of the prediction error for a random forest model, calculated using the samples not included in the bootstrap sample for each tree. It provides an unbiased estimate of the model performance.

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

- K-fold cross-validation is a technique to evaluate the performance of a model by dividing the data into K subsets (folds). The model is trained on K-1 folds and tested on the remaining fold. This process is repeated K times, with each fold used exactly once as the test data. The results are averaged to estimate the model's performance.

10. **What is hyperparameter tuning in machine learning and why it is done?**

- Hyperparameter tuning is the process of finding the best set of hyperparameters for a machine learning model to improve its performance. It is done because the choice of hyperparameters can significantly affect the model's accuracy, learning speed, and overall performance.

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

- A large learning rate can cause the gradient descent algorithm to overshoot the minimum, leading to divergence instead of convergence. It can result in unstable training and poor performance.

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

- Logistic Regression by itself cannot classify non-linear data effectively because it creates a linear decision boundary. However, it can handle non-linear relationships if we use polynomial features or kernel methods.

13. **Differentiate between Adaboost and Gradient Boosting.**

- **Adaboost**: Focuses on adjusting the weights of misclassified instances and combines weak learners sequentially. It gives more weight to difficult cases.
- **Gradient Boosting**: Builds new models that predict the residuals or errors of prior models and combines them to minimize the overall error. It optimizes the loss function directly.

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

- The bias-variance trade-off is the balance between the error due to bias (error from incorrect assumptions in the model) and the error due to variance (error from sensitivity to small fluctuations in the training set). A good model balances these to minimize total error.

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

- **Linear Kernel**: Used for linearly separable data, it computes the dot product of two vectors.
- **RBF (Radial Basis Function) Kernel**: Used for non-linear data, it measures the distance between two points and can handle complex relationships.
- **Polynomial Kernel**: Represents the similarity of vectors in a feature space over polynomials of the original variables, allowing the creation of non-linear decision boundaries.