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 is preferred because it gives a normalized measure of fit, making it easier to compare different models or datasets. It shows how well the model explains the variance in the dependent variable, which helps in assessing the model's goodness of fit.**

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 : **The relationship among TSS, ESS, and RSS is:**

TSS = ESS + RSS

This shows how the total variance (TSS) is split into the variance explained by the model (ESS) and the variance left unexplained (RSS).

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

Ans : **Regularization prevents overfitting in machine learning. Overfitting happens when a model learns noise in training data, hurting performance on new data. Regularization adds a penalty for complexity, encouraging simpler models that generalize better. This improves test data performance by balancing fit and complexity.**

4. What is Gini–impurity index?

Ans : **The Gini impurity index measures the impurity or diversity of a dataset and is used in decision tree algorithms to determine the best splits.**

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

Ans : **Unregularized decision trees often overfit by capturing noise in the training data. Regularization techniques like limiting depth, setting minimum samples for splits and leaf nodes, and pruning help reduce overfitting and improve generalization to new data.**

6. What is an ensemble technique in machine learning?

Ans. : **Ensemble techniques use multiple models to boost predictive performance. By combining individual models' strengths and reducing their weaknesses, they provide more accurate and reliable predictions.**

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

Ans : **Bagging is parallel and independent; Boosting is sequential and dependent.**
Bagging reduces variance; Boosting reduces bias.
Bagging uses average/majority vote; Boosting uses a weighted sum.

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

Ans. : **Out-of-Bag error in Random Forests uses samples not included in bootstrap samples for each tree to estimate model accuracy. It gives an internal performance measure without needing a separate validation set.**

9. What is K-fold cross-validation?

Ans. : **K-fold cross-validation assesses model performance by splitting the dataset into K folds, training and validating the model K times, and averaging the results for a reliable performance estimate.**

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

Ans. : **Hyperparameter tuning in machine learning optimizes a model's settings to boost performance. Unlike model parameters (like weights in neural networks) learned during training, hyperparameters are set before training and control the learning process.**

Hyperparameter tuning improves model performance by finding the best configuration for higher accuracy and other metrics. It optimizes learning efficiency, speeds up convergence, and enhances generalization. Proper tuning also helps avoid overfitting and ensures the model performs well on unseen data.

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

Ans. : **A large learning rate in Gradient Descent can cause divergence, oscillation, or poor convergence, leading to instability and ineffective optimization. A smaller, more appropriate learning rate can help achieve stable and efficient convergence.**

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

Ans. : **Logistic Regression is used for binary classification but assumes a linear relationship between input features and the log-odds of the target class. It's not great for modeling non-linear relationships directly. To handle non-linear data, we can use feature engineering, kernel methods, or more complex models.**

13. Differentiate between Adaboost and Gradient Boosting.

Ans. :

- **AdaBoost**: Focuses on correcting errors by adjusting the weights of misclassified instances. It combines weak learners, giving more weight to harder-to-classify examples and uses a weighted majority vote for final predictions.

- **Gradient Boosting**: Focuses on minimizing a loss function by sequentially adding models that correct residual errors of the existing ensemble. It uses gradient descent to improve predictions, often with more complex base learners.

Key Difference: AdaBoost adjusts weights based on classification errors, while Gradient Boosting improves predictions by fitting residuals and using gradient descent.

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

Ans. : **The bias-variance trade-off is a fundamental concept in machine learning The bias-variance trade-off balances a model's complexity to minimize both bias (error from too-simple models) and variance (error from too-complex models) for better performance and generalization.**

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

Ans. :

- **Linear Kernel**: Computes dot product xTyx^T yxTy. Simple and fast for linearly separable data.

- **RBF Kernel**: Measures similarity with exp⁡(−γ∥x−y∥2)\exp(-\gamma \|x - y\|^2)exp(−γ∥x−y∥2). Handles non-linear data by mapping to a high-dimensional space.

- **Polynomial Kernel**: Computes similarity with (xTy+c)d(x^T y + c)^d(xTy+c)d. Captures feature interactions with adjustable polynomial degree.