MACHINE LEARNING

---------Solutions----------

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?

Solution 1-Both R-squared and RSS are measures of goodness of fit, but with different perspectives:

R-squared: Represents the proportion of variance in the response variable explained by the model (higher is better).
RSS: Represents the sum of squared errors between predicted and actual values (lower is better).
R-squared is generally preferred as it's normalized between 0 and 1, allowing comparison across models with different scales. However, RSS can be useful for analyzing individual error patterns.

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.

Solution 2-
TSS (Total Sum of Squares): Measures the total variance in the response variable.
ESS (Explained Sum of Squares): Represents the variance explained by the model.
RSS (Residual Sum of Squares): Represents the variance not explained by the model (errors).
Equation: TSS = ESS + RSS

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

Solution 3-Regularization prevents overfitting by penalizing model complexity, leading to simpler models that may generalize better to unseen data.

4. What is Gini–impurity index?

Solution 4- Gini-impurity index measures the impurity of a node in a decision tree, indicating how well it separates different classes. Lower values signify purer nodes.

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

Solution 5- Yes, unregularized decision trees are prone to overfitting due to their high flexibility. They can fit the training data very closely, potentially capturing noise and not generalizing well to new data.

6. What is an ensemble technique in machine learning?

Solution 6- Ensemble methods combine multiple weak learners (models) to create a stronger learner. They leverage the strengths of individual models to improve overall performance.

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

Solution 7- Bagging (Bootstrap aggregating): Trains each model on a different random subset of training data with replacement. Reduces variance but may increase bias.
Boosting: Sequentially builds models, with each subsequent model focusing on correcting the errors made by the previous one. Reduces bias but may increase variance.

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

Solution 8-Out-of-bag error: Estimate of the test error for a random forest. Calculated by averaging the predictions of each tree for data points it was not trained on.

9. What is K-fold cross-validation?

Solution 9- K-fold cross-validation: Splits data into k folds, trains the model on k-1 folds, and tests it on the remaining fold. Repeats this process k times and averages the results. More robust than single validation set.

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

Solution 10- Hyperparameter tuning: Optimizing hyperparameters (model parameters not learned from data) to improve model performance. Done to find the best configuration for generalizability.

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

Solution 11- Large learning rate in Gradient Descent: Can lead to:
Overshooting minima: Skipping optimal solution and oscillating around it.
Divergence: Weights exploding and model training failing.

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

Solution 12- No, Logistic Regression is a linear model and can only handle linearly separable data. Non-linear data requires non-linear models like Support Vector Machines or Neural Networks.

13. Differentiate between Adaboost and Gradient Boosting.

Solution 13- Adaboost: Adaptively boosts weights of training examples based on their difficulty for previous models. Focuses on hard-to-learn instances.
Gradient Boosting: Sequentially builds models focusing on the gradient (error) of the previous model. Reduces errors iteratively.

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

Solution 14- Bias-variance trade-off: Balancing bias (model's tendency to underfit) and variance (model's tendency to overfit). Lower bias and variance lead to better generalizability.

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

Solution 15-Linear Kernel: Measures the linear similarity between data points. Suitable for linearly separable data.

RBF (Radial Basis Function) Kernel: Measures similarity based on Euclidean distance in a transformed space. Handles non-linear data well.

Polynomial Kernel: Raises the inner product of data points to a power to capture non-linear relationships. More complex but prone to overfitting.