MACHINE LEARNING ASSIGNMENT - 5

Q1 to Q15 are subjective answer type questions, Answer them briefly.

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 generally considered a better measure of the goodness of fit of a regression model compared to Residual Sum of Squares (RSS) because R-squared is a standardized measure that ranges from 0 to 1, where 1 indicates a perfect fit. It provides a straightforward interpretation of the proportion of the variability in the dependent variable that is explained by the independent variables. In contrast, RSS is an absolute measure and doesn't provide a clear sense of the model's overall explanatory power.

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. In regression analysis, Total Sum of Squares (TSS), Explained Sum of Squares (ESS), and Residual Sum of Squares (RSS) are important metrics used to assess the goodness of fit of the model and the equation relating these three matrics with each other are TSS=ESS+RSS

3. What is the need of regularization in machine learning?
Ans. Regularization in machine learning is a technique used to prevent overfitting and improve the generalization performance of a model. Overfitting occurs when a model captures noise or random fluctuations in the training data, leading to poor performance on new, unseen data. Regularization is introduced to address this issue and achieve better model performance and some of the key reasons for the need of regularization:

a. Preventing Overfitting
b. Handling Collinearity
c. Feature Selection
d. Improving Numerical Stability
e. Enhancing Model Interpretability

4. What is Gini–impurity index?
Ans. The Gini impurity index is a measure of how often a randomly chosen element from the set would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the set. In the context of decision trees and machine learning, the Gini impurity is commonly used as a criterion for splitting nodes in a decision tree

5. Are unregularized decision-trees prone to overfitting? If yes, why?
Ans. Yes, unregularized decision trees are prone to overfitting
Reason for overfitting are:

a. High Variance
b. Memorization of Training Data
c. Overfitting to Outliers
d. Failure to Generalize

6. What is an ensemble technique in machine learning?
Ans. An ensemble technique in machine learning involves combining the predictions of multiple individual models to create a stronger, more robust model. The idea behind ensemble methods is to leverage the diversity of multiple models to improve overall performance, reduce overfitting, and enhance generalization to new, unseen data. Ensemble methods are particularly powerful when individual models have different strengths and weaknesses

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

1. Training Process:

- Bagging involves training multiple instances of the same learning algorithm on different subsets of the training data.
- Each subset is typically obtained through bootstrapping, which means sampling with replacement from the original training data.
- The base learners are trained independently and in parallel.

2. Diversity of Models:

- The goal in bagging is to create diverse base models by exposing them to different subsets of the data.
- The diversity is achieved through the randomness introduced by bootstrapping.

3. Combining Predictions:

- The predictions of individual models are combined, often by averaging (for regression problems) or voting (for classification problems).
- The aggregation of predictions helps reduce variance and improve generalization.

4. Example:

- Random Forest is a popular example of a bagging ensemble method, where the base learners are decision trees.

B. Boosting:

1. Training Process:

- Boosting involves training base learners sequentially, where each subsequent learner focuses on correcting the errors made by the previous ones.

- The training process assigns higher weights to the instances that were misclassified by earlier models, emphasizing the challenging cases.
2. Adaptive Learning:
- Boosting is adaptive and assigns more importance to instances that are difficult to classify, effectively giving more attention to the mistakes made by the previous models.
3. Combining Predictions:
- The predictions of individual models are combined by giving different weights to each model's prediction.
- The final model is a weighted sum of the base learners.
4. Example:
- AdaBoost (Adaptive Boosting) and Gradient Boosting Machines (GBM) are examples of boosting ensemble methods

8. What is out-of-bag error in random forests?
Ans. The out-of-bag (OOB) error is a concept associated with ensemble methods, particularly with Random Forests. In a Random Forest, each tree in the ensemble is trained on a bootstrapped sample of the original dataset, which means that some data points are not included in the training set for each tree. The OOB error is a way to estimate the performance of the Random Forest on unseen data points

9. What is K-fold cross-validation?
Ans. K-fold cross-validation is a technique used in machine learning for assessing the performance and generalization ability of a model. It involves splitting the dataset into K subsets, or "folds," and then performing training and evaluation K times. In each iteration, one of the folds is used as the test set, while the remaining K-1 folds are used for training

10. What is hyper parameter tuning in machine learning and why it is done?
Ans. Hyperparameter tuning, also known as hyperparameter optimization, is the process of selecting the optimal hyperparameters for a machine learning model. Hyperparameters are configuration settings external to the model that cannot be learned from the data. They are set before the training process and influence the model's performance
- Here's why hyperparameter tuning is done:

a. Optimizing Performance
b. Preventing Overfitting or Underfitting
c. Adapting to Data Characteristics
d. Enhancing Model Robustness
e. Exploring Model Complexity

11. What issues can occur if we have a large learning rate in Gradient Descent?
Ans. Here are some common problems associated with a large learning rate:

a. Divergence
b. Failure to Converge
c. Oscillations and Bouncing
d. Instability
e. Skipping Over the Minimum
f. Slow Convergence in the End

12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?
Ans. Logistic Regression is a linear classification algorithm, which means it assumes a linear relationship between the input features and the log-odds of the predicted output. While Logistic Regression is powerful for linearly separable datasets, it may not perform well on datasets with complex, non-linear decision boundaries
Reasons why we cant use logistic regression for non linear data:

a. Linear Decision Boundary
b. Underfitting
c. Limited Expressiveness
d. Feature Engineering

13. Differentiate between Adaboost and Gradient Boosting.
Ans.**A. Adaboost (Adaptive Boosting)**:

- Weighted Data Points:

Adaboost assigns weights to data points during training. Initially, all data points have equal weights, but the weights are adjusted based on the performance of the weak learners.

- Sequential Training:

Adaboost trains a series of weak learners sequentially. After each iteration, the weights of misclassified data points are increased, and the subsequent weak learner focuses more on these misclassified points.

- Combining Predictions:

The final prediction is obtained by combining the weighted predictions of all weak learners. The weight of each weak learner in the final prediction is determined by its accuracy.

- Adaptive Learning:

The algorithm adapts by giving higher importance to misclassified instances, allowing subsequent models to focus on the more challenging cases.

- Weak Learners:

Adaboost typically uses simple weak learners, such as shallow decision trees or stumps (trees with a single level).

B. Gradient Boosting:

- Residuals as Targets:

Gradient Boosting focuses on minimizing the residuals (errors) of the predictions made by the weak learners. It sequentially fits models to the residuals of the previous models.

- Sequential Training:

Similar to Adaboost, Gradient Boosting trains a series of weak learners sequentially. However, each learner is trained to correct the errors made by the ensemble so far.

- Learning Rate:

Gradient Boosting introduces a learning rate parameter that controls the contribution of each weak learner. A lower learning rate makes the model more robust but requires more iterations.

- Combining Predictions:

The final prediction is the sum of the predictions from all weak learners, each multiplied by its learning rate. The learning rate helps in preventing overfitting and controlling the step size during optimization.

- Weak Learners:

Gradient Boosting often uses decision trees as weak learners, and the trees can be deeper compared to those in Adaboost.

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

Ans. The bias-variance trade-off is a fundamental concept in machine learning that refers to the balance between bias and variance in the performance of a model. It helps explain the trade-off between the simplicity and complexity of a model and their impact on the model's ability to generalize to new, unseen data

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

Ans.

- Linear Kernel: The linear kernel is the simplest kernel and is used for linearly separable data. It computes the dot product between the input features, effectively implementing a linear decision boundary.
- RBF (Radial Basis Function) Kernel:The RBF kernel, also known as the Gaussian kernel, transforms data into an infinite-dimensional space. It is capable of capturing complex, non-linear relationships between features.
- Polynomial Kernel:The polynomial kernel transforms data into higher-dimensional spaces using polynomial functions. It allows SVMs to capture non-linear relationships between features.