Question Bank on Random Forest and Regression Trees

Multiple Choice Questions

1. What is the purpose of bagging in Random Forest?

   a) To reduce variance and improve model generalization
   b) To increase bias and reduce overfitting
   c) To create a single strong decision tree
   d) To ensure all trees are identical

2. In Random Forest, how is the final prediction made?

   a) By selecting the prediction of the deepest tree
   b) By averaging (for regression) or majority voting (for classification)
   c) By taking the maximum predicted value
   d) By summing the predictions of all trees

3. What does the term 'ensemble method' refer to in Random Forest?

   a) A single large decision tree
   b) A collection of models that work together to improve performance
   c) A process of pruning decision trees
   d) A technique that removes outliers from the dataset

4. What percentage of the training data is typically used in each bag in Random Forest?

   a) 50%
   b) 67%
   c) 80%
   d) 100%

5. Which of the following is a hyperparameter in `RandomForestRegressor`?

   a) Learning rate
   b) Number of hidden layers
   c) Number of estimators (`n_estimators`)
   d) Dropout rate

6. What is the key benefit of using subsets of features in Random Forest?

   a) It prevents all trees from looking the same
   b) It increases overfitting
   c) It ensures trees are deeper
   d) It eliminates the need for bagging

7. What does 'OOB' stand for in Random Forest?

   a) Out-of-Bag
   b) Overfitting Optimization Bias
   c) Ordered Observations Batch
   d) Optimal Output Boost

8. In the `RandomForestRegressor` algorithm, what does the parameter `max_depth` control?

   a) The number of trees in the forest
   b) The number of features used in each split
   c) The maximum depth of each decision tree
   d) The minimum number of samples required to split a node

9. How are the samples for bagging selected in Random Forest?

   a) Randomly with replacement
   b) Randomly without replacement
   c) Based on entropy measures
   d) Using stratified sampling

10. What is a potential limitation of regression trees that Random Forest addresses?

    a) They are too fast to compute
    b) They are highly interpretable
    c) They are prone to high variance and overfitting
    d) They do not require hyperparameter tuning

Short Answer Questions

1. What is the main purpose of the 'ensemble method' in Random Forest?

2. How does bagging reduce overfitting in regression trees?

3. Define the term 'OOB' and its role in Random Forest.

4. Why are subsets of features used in Random Forest for splitting?

5. Explain the significance of hyperparameters in RandomForestRegressor.

6. What role does bootstrap sampling play in Random Forest?

7. How does Random Forest handle missing data during training?

8. What is the importance of the `n_estimators` parameter in RandomForestRegressor?

9. Why is Random Forest considered more robust than a single regression tree?

10. How can you evaluate the performance of a Random Forest model?

True/False Questions

1. Random Forest always uses the entire training dataset for each tree.

2. Bagging in Random Forest ensures that every sample is used exactly once in training.

3. Out-of-bag (OOB) samples are used to evaluate model performance.

4. Hyperparameters in RandomForestRegressor must always be tuned manually.

5. The final prediction in Random Forest is made by taking a weighted average of individual tree predictions.

6. Random Forest can handle both regression and classification tasks.

7. The `max_features` parameter controls the maximum number of trees in the forest.

8. Bagging reduces variance by averaging predictions across multiple models.

9. Random Forest models are inherently interpretable.

10. Increasing the number of trees in a Random Forest always improves performance.

Answers:

MCQ

1. a) To reduce variance and improve model generalization

2. b) By averaging (for regression) or majority voting (for classification)

3. b) A collection of models that work together to improve performance

4. b) 67%

5. c) Number of estimators (`n_estimators`)

6. a) It prevents all trees from looking the same

7. a) Out-of-Bag

8. c) The maximum depth of each decision tree

9. a) Randomly with replacement

10. c) They are prone to high variance and overfitting

Short Answer Questions

1. To combine predictions from multiple models to improve accuracy.

2. By averaging predictions from multiple randomly sampled datasets.

3. Out-of-Bag (OOB) samples are used to evaluate model performance without additional validation datasets.

4. To reduce correlation between trees and enhance model generalization.

5. They control the behavior of the model, such as the number of trees, depth, and features used.

6. It creates diverse datasets by sampling with replacement from the original data.

7. It imputes missing values based on proximity to other data points.

8. It sets the number of trees in the forest, impacting model performance and stability.

9. It reduces overfitting and variance by averaging predictions across multiple trees.

10. Using metrics such as OOB error, cross-validation, and performance on unseen test data.

True/False Questions

1. False

2. False

3. True

4. False

5. False

6. True

7. False

8. True

9. False

10. False