Randomized Grid Search and Decision Trees Test Paper – 20 Questions (20 mins) with answers:

🔍 Randomized Grid Search Section

Q1. What is Randomized Grid Search, and why is it used?

Answer:

Randomized Grid Search is a hyperparameter tuning technique that samples a fixed number of parameter combinations from a specified distribution. It is used to efficiently search a large parameter space with fewer computations than Grid Search.

Q2. How does Randomized Grid Search differ from Grid Search?

Answer:

Grid Search tries every possible combination of parameters, while Randomized Grid Search selects a random subset of combinations, which reduces computation time.

Q3. What is the purpose of the n\_iter parameter in Randomized Grid Search?

Answer:

N\_iter specifies the number of parameter settings that will be sampled. A higher value increases search thoroughness but also computational cost.

Q4. What are the advantages of using Randomized Grid Search over Grid Search?

Answer:

* Faster for large parameter spaces
* Can find good solutions with fewer evaluations
* Allows searching across distributions rather than fixed values

Q5. When might Randomized Grid Search be preferable to Grid Search?

Answer:

When the hyperparameter space is large or expensive to search exhaustively, or when there is limited time or computational resources.

🌳 Decision Trees Section

Q6. What is a Decision Tree in machine learning?

Answer:

A Decision Tree is a supervised learning model used for classification and regression that splits data into subsets based on feature values, forming a tree structure of decisions.

Q7. Describe the structure of a Decision Tree.

Answer:

A Decision Tree is composed of nodes:

* Root Node: The topmost node that starts the split
* Internal Nodes: Represent decisions or tests on features
* Leaf Nodes: Represent output values or class labels

Q8. What is Gini impurity, and how is it used in Decision Trees?

Answer:

Gini impurity measures the likelihood of an incorrect classification of a randomly chosen element. It is used to determine how good a feature is at splitting the data.

Q9. How does entropy help Decision Trees make splits?

Answer:

Entropy measures the amount of disorder or uncertainty. Lower entropy after a split means better separation. The algorithm uses entropy to calculate information gain.

Q10. What is a leaf node in a Decision Tree?

Answer:

A leaf node represents a class label (in classification) or a value (in regression) and does not split further.

Q11. What does the max\_depth parameter do in a Decision Tree?

Answer:

Max\_depth limits the depth of the tree, preventing it from growing too deep and potentially overfitting the training data.

Q12. Define information gain and explain its role in Decision Trees.

Answer:

Information gain measures the reduction in entropy from a split. It is used to decide which feature to split on at each node.

Q13. What are some common ways to avoid overfitting in Decision Trees?

Answer:

* Pruning
* Limiting max\_depth
* Setting min\_samples\_split or min\_samples\_leaf
* Using cross-validation

Q14. What is pruning in the context of Decision Trees?

Answer:

Pruning is the process of removing branches or nodes that have little importance to improve generalization and reduce overfitting.

Q15. How does a Decision Tree handle continuous and categorical features differently?

Answer:

For continuous features, it finds optimal split points (e.g., feature < 5.3). For categorical features, it typically splits based on exact matches or groupings of categories.

Q16. Describe the role of the min\_samples\_split parameter in a Decision Tree.

Answer:

It sets the minimum number of samples required to split an internal node, preventing splits that are not statistically significant.

Q17. What are some advantages and disadvantages of using Decision Trees?

Answer:

Advantages:

* Easy to interpret
* Handles both numeric and categorical data
* No need for feature scaling

Disadvantages:

* Prone to overfitting
* Unstable to small data variations
* Greedy nature may miss optimal splits

Q18. Why are Decision Trees considered “greedy” algorithms?

Answer:

They make locally optimal decisions at each node (best immediate split) without considering future consequences or global optimality.

Q19. Explain the difference between classification and regression Decision Trees.

Answer:

Classification Trees predict class labels; Regression Trees predict continuous values. They use different criteria (e.g., Gini vs. MSE) to make splits.

Q20. What is the impact of the min\_samples\_leaf parameter in a Decision Tree?

Answer:

It specifies the minimum number of samples required in a leaf node. Increasing it helps reduce overfitting and creates a more generalized model.