

Theory

1. Can we use Bagging for regression problems?

Yes, Bagging can be used for regression problems by using base regressors such as Decision Trees, Random Forest, or SVR. Bagging Regressor helps to reduce variance and improve the stability of predictions.

2. What is the difference between multiple model training and single model training?

Single Model Training: Uses one model trained on the entire dataset, leading to potential overfitting or underfitting.

Multiple Model Training: Uses several models (ensemble) to improve generalization by combining multiple predictions.

3. Explain the concept of feature randomness in Random Forest.

Feature randomness in Random Forest ensures that at each decision node, a random subset of features is considered instead of all features. This increases model diversity and reduces correlation among trees.

4. What is OOB (Out-of-Bag) Score? OOB Score is an internal validation score in Bagging-based models like Random Forest. It is computed using only those data points that were not selected during bootstrapping, providing an unbiased accuracy estimate.

5. How can you measure the importance of features in a Random Forest model?

Feature importance in Random Forest is measured using:

1. **Gini Importance (Mean Decrease in Impurity)** – Measures how much a feature contributes to reducing impurity.

2. **Permutation Importance** – Assesses the drop in model accuracy after randomly shuffling a feature's values.

6. Explain the working principle of a Bagging Classifier.

A Bagging Classifier:

1. Creates multiple bootstrapped (randomly sampled with replacement) datasets.
2. Trains independent base models (e.g., Decision Trees) on each subset.
3. Aggregates predictions using majority voting (for classification).

7. How do you evaluate a Bagging Classifier's performance?

A Bagging Classifier can be evaluated using:

Accuracy

Precision, Recall, F1-score

ROC-AUC Score

Cross-validation

8. How does a Bagging Regressor work? A Bagging Regressor:

1. Trains multiple base regressors on bootstrapped subsets of data.
2. Averages their predictions to reduce variance and improve stability.

9. What is the main advantage of ensemble techniques?

The main advantage is improved **generalization** by reducing overfitting and bias through model diversity.

10. What is the main challenge of ensemble methods?

The main challenge is increased **computational complexity** due to multiple models being trained.

11. Explain the key idea behind ensemble techniques.

Ensemble methods improve predictions by **combining multiple weak models** to create a strong learner.

12. What is a Random Forest Classifier?

A Random Forest Classifier is an ensemble of Decision Trees trained on bootstrapped samples, using random feature selection to improve diversity and reduce overfitting.

13. What are the main types of ensemble techniques?

1. **Bagging (Bootstrap Aggregation)** – Reduces variance (e.g., Random Forest).
2. **Boosting** – Reduces bias by training weak models sequentially (e.g., AdaBoost, XGBoost).
3. **Stacking** – Combines multiple models using a meta-learner.

14. What is ensemble learning in machine learning?

Ensemble learning is a technique where multiple models are combined to improve performance.

15. When should we avoid using ensemble methods?

When the dataset is small and simple models already achieve good accuracy.

When interpretability is critical, as ensembles are complex.

16. How does Bagging help in reducing overfitting?

Bagging reduces overfitting by training models on different bootstrapped datasets and averaging predictions, reducing model variance.

17. Why is Random Forest better than a single Decision Tree?

Less prone to overfitting due to multiple trees.

More robust and generalizable.

Handles missing values and feature importance well.

18. What is the role of bootstrap sampling in Bagging?

Bootstrap sampling ensures each base model gets a **randomized** version of the training data, promoting model diversity.

19. What are some real-world applications of ensemble techniques?

Fraud detection

Stock market prediction

Medical diagnosis

Image and speech recognition

20. What is the difference between Bagging and Boosting?

Feature	Bagging	Boosting
Focus	Reduces variance	Reduces bias
Training	Parallel	Sequential
Model Complexity	Less complex	More complex
Example	Random Forest	AdaBoost, XGBoost