

Assignment - Ensemble Techniques

Theoretical Questions and Answers

1. Can we use Bagging for regression problems?

Yes, Bagging can be used for regression. A common example is the BaggingRegressor, which combines predictions from multiple regressors by averaging their outputs to reduce variance and improve accuracy.

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

Single model training involves training one algorithm, while multiple model training (ensemble learning) trains several models and combines their outputs to improve performance, generalization, and stability.

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

In Random Forests, each tree considers a random subset of features for splitting at each node. This feature randomness increases diversity among trees and prevents overfitting.

4. What is OOB (Out-of-Bag) Score?

OOB Score is a performance estimate computed using data not included in each bootstrap sample. It provides an internal cross-validation score without needing a separate test set.

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

Feature importance is measured by evaluating how much a feature decreases the impurity (like Gini index) across all trees. The more a feature contributes to reducing impurity, the higher its importance score.

6. Explain the working principle of a Bagging Classifier.

A Bagging Classifier trains multiple models on different bootstrapped subsets of the training data and combines their outputs via majority voting (classification) or averaging (regression).

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

Performance can be evaluated using metrics like accuracy, precision, recall, F1-score, or AUC on validation/test sets. Out-of-Bag (OOB) score is also commonly used for evaluation.

8. How does a Bagging Regressor work?

It trains several regressors (usually decision trees) on bootstrapped samples and averages their predictions to reduce variance and improve model stability.

9. What is the main advantage of ensemble techniques?

The main advantage is better generalization and performance due to combining diverse models, which reduces variance, mitigates overfitting, and increases robustness.

10. What is the main challenge of ensemble methods?

Challenges include high computational cost, reduced model interpretability, and the need for tuning multiple parameters.

11. Explain the key idea behind ensemble techniques.

The key idea is to combine multiple weak or strong learners to produce a more accurate and robust model than any single one.

12. What is a Random Forest Classifier?

A Random Forest Classifier is an ensemble of decision trees, trained using bootstrap aggregation and random feature selection, and combined using majority voting.

13. What are the main types of ensemble techniques?

The main types are Bagging, Boosting, and Stacking. Each combines multiple models in different ways to improve performance.

14. What is ensemble learning in machine learning?

Ensemble learning is a technique that builds a set of models and combines their predictions to obtain better predictive performance than any single model alone.

15. When should we avoid using ensemble methods?

Avoid them when interpretability is crucial, computational resources are limited, or when the base models already perform well enough.

16. How does Bagging help in reducing overfitting?

Bagging reduces overfitting by averaging predictions from multiple models trained on different subsets, thus reducing variance.

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

Random Forests reduce overfitting and improve generalization by combining multiple decision trees, whereas a single decision tree can be sensitive to noise and data splits.

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

Bootstrap sampling creates diverse training sets by randomly sampling with replacement, allowing each model in the ensemble to learn different patterns.

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

Ensemble methods are used in fraud detection, medical diagnosis, recommendation systems, spam filtering, and stock market predictions.

20. What is the difference between Bagging and Boosting?

- **Bagging** builds models independently in parallel using bootstrap samples and averages them.
- **Boosting** builds models sequentially, focusing on correcting errors made by previous models, often resulting in lower bias.