Assignment-Ensemble Learning

1. **Can we use Bagging for regression problems?**
    a. Yes, Bagging can be used for regression problems. A Bagging Regressor, like BaggingRegressor in scikit-learn, aggregates predictions from multiple base regressors (e.g., Decision Trees) by averaging their outputs.
2. **What is the difference between multiple model training and single model training?**
    a. Single model training uses one model to learn patterns from data, while multiple model training (ensemble learning) combines predictions from multiple models to improve accuracy and reduce overfitting.
3. **Explain the concept of feature randomness in Random Forest.**
    a. Random Forest introduces feature randomness by selecting a random subset of features at each split in a decision tree, preventing trees from becoming too similar and improving generalization.
4. **What is OOB (Out-of-Bag) Score?**
    a. OOB score is a validation metric in Random Forest that evaluates model performance using the bootstrap samples that were not included in training, providing an unbiased estimate of accuracy.
5. **How can you measure the importance of features in a Random Forest model?**
    a. Feature importance is measured using metrics like Gini Importance (for classification) or Mean Decrease in Impurity (for regression), which show how much each feature contributes to reducing the prediction error.
6. **Explain the working principle of a Bagging Classifier.**
    a. A Bagging Classifier trains multiple base models on different bootstrap samples of the dataset and aggregates their predictions (majority vote for classification, averaging for regression).
7. **How do you evaluate a Bagging Classifier's performance?**
    a. A Bagging Classifier can be evaluated using metrics like accuracy, precision, recall, F1-score, confusion matrix, ROC-AUC, and cross-validation.
8. **How does a Bagging Regressor work?**
    a. A Bagging Regressor trains multiple regressors on bootstrapped datasets and averages their predictions to reduce variance and improve stability.
9. **What is the main advantage of ensemble techniques?**
    a. The main advantage is improved accuracy and robustness by reducing overfitting and variance through model combination.
10. **What is the main challenge of ensemble methods?**
    a. The main challenge is increased computational complexity and difficulty in interpretability compared to single models.
11. **Explain the key idea behind ensemble techniques.**
    a. Ensemble techniques combine multiple models to achieve higher accuracy and robustness than a single model by reducing bias and variance.
12. **What is a Random Forest Classifier?**
    a. A Random Forest Classifier is an ensemble method that builds multiple decision trees and uses majority voting for classification.

13. What are the main types of ensemble techniques?

The main types are:

a. i. Bagging (Bootstrap Aggregating)
b. ii. Boosting (e.g., AdaBoost, Gradient Boosting)
c. iii. Stacking (Stacked Generalization)
d. iv. Voting & Averaging

14. What is ensemble learning in machine learning?

a. Ensemble learning is a technique that combines multiple models to improve prediction accuracy and robustness.

15. When should we avoid using ensemble methods?

a. When interpretability is crucial, when the dataset is too small, or when computational resources are limited.

16. How does Bagging help in reducing overfitting?

a. Bagging reduces overfitting by training multiple models on different bootstrap samples and averaging their predictions, preventing reliance on any single model.

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

a. Random Forest reduces variance, prevents overfitting, and provides better generalization compared to a single Decision Tree.

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

a. Bootstrap sampling creates multiple training subsets by randomly sampling with replacement, allowing diverse models to be trained.

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

a. Fraud detection, medical diagnosis, stock market prediction, customer churn prediction, recommendation systems, and image classification.

20. What is the difference between Bagging and Boosting?

a. Bagging trains models independently and combines their outputs (reduces variance).
b. Boosting trains models sequentially, with each new model correcting the errors of the previous ones (reduces bias).