Part-III: Ensemble Learning – Bagging, Boosting, Random Forest

Task 7: Conceptual Questions

1. What is the difference between Bagging and Boosting?
Bagging (Bootstrap Aggregating) trains multiple models independently on random subsets of data and averages their predictions to reduce variance. Boosting builds models sequentially, where each new model focuses on correcting the errors of the previous one, reducing both bias and variance.

2. How does Random Forest reduce variance?
Random Forest reduces variance by combining predictions from many decision trees trained on different bootstrap samples and random subsets of features. This ensemble approach averages out individual tree errors, leading to more stable and generalized predictions.

3. What is the weakness of boosting-based methods?
Boosting methods are sensitive to noisy data and outliers because they give more weight to misclassified points in each iteration. They also require more computational resources and can overfit if the number of iterations is too high.