

# 12th April Assignment

April 25, 2023

## 1 Assignment 67

**Q1.** How does bagging reduce overfitting in decision trees?

**Ans.** Bagging reduces overfitting in decision trees by creating multiple subsets of the training data set, training a decision tree model on each subset, and then combining the predictions made by each model to give a final prediction. This technique helps to reduce the variance of the predictions made by the individual models and improves the overall accuracy and generalizability of the model.

**Q2.** What are the advantages and disadvantages of using different types of base learners in bagging?

**Ans.** Bagging is a technique that can be applied to a wide range of base learners, including decision trees, neural networks, support vector machines, and more. Each type of base learner has its own advantages and disadvantages when used in bagging.

**Decision Trees:**

**Advantages:**

- Fast to train
- Can handle both numerical and categorical
- Easily interpretable

**Disadvantage:**

- Prone to overfitting
- Limited predictive power for complex relationships in the data
- Not as effective on high-dimensional data

**Neural Networks:**

**Advantages:**

- Good at handling complex relationships in data
- Can model nonlinear relationships
- Can handle high-dimensional data

**Disadvantages:**

- Can be slow to train
- Can be prone to overfitting
- Not easily interpretable

**SVM****Advantages:**

- Effective for high-dimensional data
- Good at handling both numerical and categorical data
- Can model nonlinear relationships

**Disadvantages:**

- Can be slow to train
- May require a large amount of memory
- Difficult to interpret

**Q3. How does the choice of base learner affect the bias-variance tradeoff in bagging?**

**Ans. The choice of base learner can affect the bias-variance tradeoff in the following ways:**

1. High-bias base learners, such as decision trees, may benefit from bagging by reducing the variance of the predictions. This can lead to improved generalization performance without increasing bias.
2. High-variance base learners, such as neural networks, may benefit from bagging by reducing the variance of the predictions. This can lead to improved generalization performance without sacrificing too much bias.

In some cases, the choice of base learner may not have a significant impact on the bias-variance tradeoff in bagging. For example, if the base learner is already a low-bias, low-variance model, bagging may not provide significant improvements.

**Q4. Can bagging be used for both classification and regression tasks? How does it differ in each case?**

**Ans.**Yes, bagging can be used for both classification and regression tasks.

In classification tasks, bagging involves creating multiple subsets of the training data using bootstrapping, training a classifier on each subset, and then combining the predictions of these individual classifiers by taking a majority vote to produce a final prediction. This technique is known as bagging or bootstrap aggregating.

In regression tasks, bagging involves creating multiple subsets of the training data using bootstrapping, training a regression model on each subset, and then combining the predictions of these individual models by taking the average to produce a final prediction.

**Q5.** What is the role of ensemble size in bagging? How many models should be included in the ensemble?

**Ans.**The ensemble size, or the number of models included in the bagging ensemble, plays an important role in determining the performance of the bagging technique.

Increasing the ensemble size can improve the performance of bagging by reducing the variance of the individual models and increasing the stability of the ensemble. This is because the more models are included in the ensemble, the more likely it is that the errors and biases of the individual models will cancel out, leading to better generalization performance.

**Q6.** Can you provide an example of a real-world application of bagging in machine learning?

**Ans.**One example of a real-world application of bagging in machine learning is in the field of credit risk assessment for lending institutions.

Lending institutions such as banks, credit card companies, and loan providers need to assess the credit risk of applicants before deciding whether to approve a loan or credit application. This involves analyzing various factors such as income, employment history, credit history, and other personal information to determine the likelihood of the applicant defaulting on the loan.

Bagging can be used in this context to improve the accuracy of the credit risk assessment model. Multiple subsets of the training data can be created using bootstrapping, and a decision tree classifier can be trained on each subset. The predictions of the individual decision trees can then be combined by taking a majority vote to produce a final prediction.

This approach helps to reduce the variance of the individual decision trees and improve the stability of the ensemble, leading to better generalization performance and more accurate credit risk assessments. Bagging can also help to identify important features that contribute to credit risk and provide insights into the decision-making process.

Overall, bagging is a powerful technique that can be used in a wide range of machine learning applications, including credit risk assessment, fraud detection, image classification, and natural language processing, among others.