COSC6260 - Assign #1

Henri Medeiros Dos Reis

September 19, 2023

1. Consider the following valid code segments. Write the worst- case recurrence relation and solve the Big-Oh time complexity.

a)

```
test(n)
{
    if(n<=0)
        return 1;
    return 3*test(n-1)+5
}
```

The recurrence for this problem is

$$t(n) = \begin{cases} t(n) + O(1), & \text{if } n > 0 \\ 1, & \text{if } n = 0 \end{cases}$$

Since both the multiplication by 3 and the addition of 5 take constant time.
Then using Master's Theorem we get
$t(n) = at(n-b) + f(n) = t(n-1) + O(1)$, where $a = 1, b = 1, f(n) = O(1), k = 0$
$\Rightarrow$ case 1 $\Rightarrow O(n^{k+1}) = O(n)$

b)

```
g(n)
{
    if(n<=1)
        return 1;
    else
        return g(n/2)+3*g(n/2)+n
}
```

The recurrence for this problem is

$$t(n) = \begin{cases} t(\frac{n}{2}) + t(\frac{n}{2}) + O(1), & \text{if } n > 1 \\ 1, & \text{if } n = 1 \end{cases}$$

Since the multiplication of the output of $t(\frac{n}{2})$ by 3 takes constant time, and adding n also takes constant time.

Then using Master's Theorem we get

$t(n) = at(\frac{n}{b}) + f(n) = 2t(\frac{n}{2}) + O(1)$, where $a = 2, b = 2, f(n) = O(1), log_b a = log_2 2 = 1, k = 0, p = 0$

$\Rightarrow$ case 1 $\Rightarrow O(n^{log_b a}) = O(n)$

2. Give the Big-Oh time complexity of the following recurrence relations.

a) $t(n) = 2t(\frac{n}{2}) + n^4$

Using Master's Theorem we get

$t(n) = at(\frac{n}{b}) + f(n) = 2t(\frac{n}{2}) + n^4$, where $a = 2, b = 2, f(n) = O(n^k log^p n) = O(n^4 log^0 n) = O(n^4 log^0 n), log_b a = log_2 2 = 1, k = 4, p = 0$

$\Rightarrow$ case 3, $p = 0 \Rightarrow O(n^k log^p n) = O(n^4 log^0 n) = O(n^4)$

b) $t(n) = 16t(\frac{n}{4}) + n^2$

Using Master's Theorem we get

$t(n) = at(\frac{n}{b}) + f(n) = 16t(\frac{n}{4}) + n^2$, where $a = 16, b = 4, f(n) = O(n^k log^p n) = O(n^2 log^0 n) = O(n^2), log_b a = log_4 16 = 2, k = 2, p = 0$

$\Rightarrow$ case 2, $p > -1 \Rightarrow O(n^k log^{p+1} n) = O(n^2 logn)$

c) $t(n) = 2t(\frac{n}{4}) + \sqrt{n}$

Using Master's Theorem we get

$t(n) = at(\frac{n}{b}) + f(n) = 2t(\frac{n}{4}) + n^{\frac{1}{2}}$, where $a = 2, b = 4, f(n) = O(n^k log^p n) = O(n^{\frac{1}{2}} log^0 n) = O(n^{\frac{1}{2}}), log_b a = log_4 2 = \frac{1}{2}, k = \frac{1}{2}, p = 0$

$\Rightarrow$ case 2, $p > -1 \Rightarrow O(n^k log^{p+1} n) = O(n^{\frac{1}{2}} logn)$

d) $t(n) = t(n-2) + n^2$

Using Master's Theorem we get

$t(n) = at(n-b) + f(n) = 1t(n-2) + n^2$, where $a = 1, b = 2, f(n) = O(n^2), k = 2$

$\Rightarrow$ case 1 $\Rightarrow O(n^{k+1}) = O(n^3)$

3. Give an example of a divide and conquer algorithm in machine learning space. Write a short report on the algorithm steps, usage and time complexity. Be sure to cite your sources.

1 Introduction

The chosen divide and conquer algorithm is Random Forests. Which is a powerful ensemble learning technique used in machine learning for both classification and regression. It is a divide and conquer algorithm that combines the predictions of multiple decision trees to improve predictive accuracy, reduce overfitting, and feature selection. Random Forest is widely used in various applications, including image classification, fraud detection, and recommendation systems.

2 Algorithm Steps

i **Random Sampling**: The first step in Random Forest is to randomly select a subset of the training data with replacement. This creates a lot of training data sets with the same size as the original data set but with some variations.

ii **Decision Tree Construction**: For each of the subset of data sets, a decision tree is constructed. Decision trees are created by recursively splitting the data based on features to maximize the information gain or minimize the mean squared error (MSE).

iii **Random Feature Selection**: During the construction of each decision tree, only a random subset of features is considered at each split. This helps to decorrelate the trees and reduce overfitting.

iv **Voting or Averaging**: After all the decision trees are built, they are used to make predictions on new data points. For classification, the majority vote among the trees is taken as the final prediction, and for regression, the average of the predictions from individual trees is computed.

3 Usage

Random Forest is versatile and can be used in a wide range of applications, including:

i **Classification**: It can be used for tasks like spam email detection, image classification, and sentiment analysis.

ii **Regression**: It can also be used for predicting continuous values such as stock prices, housing prices, or weather forecasting.

iii **Anomaly Detection**: Random Forest can identify unusual patterns in data, making it valuable for fraud detection and quality control.

iv **Feature Importance**: It can also be used to measure the importance of features in a data set, helping with feature selection and understanding the data.

4 Time Complexity

The time complexity of Random Forest primarily depends on the number of trees in the forest (n_{trees}), the number of data points (n), and the number of features (m). Each decision tree's construction has a time complexity of $O(nm\log(n))$, and since we construct multiple trees, the overall complexity is $O(n_{\text{trees}}nm\log(n))$.

The prediction time complexity for a single data point is $O(n_{\text{trees}}m)$, as each tree in the forest is evaluated independently.

Random Forest's training time can be parallelized, making it suitable for large datasets and high-dimensional feature spaces. It is known for its efficiency and ability to handle complex problems efficiently.

In conclusion, Random Forest is a powerful divide and conquer algorithm in the field of machine learning, known for its robustness, accuracy, and ability to handle various types of data and tasks. Its ensemble approach and randomization techniques make it a popular choice for both beginners and experienced practitioners.

5 References

https://www.geeksforgeeks.org/random-forest-regression-in-python/

https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

https://medium.com/analytics-vidhya/time-complexity-of-ml-models-4ec39fad2770