Decision Tree Algorithm in Python From Scratch

January 8, 2023

Abstract

Coding the popular algorithm using just NumPy and Pandas in Python and explaining what's under the hood

1 Introduction

The aim of this article is to make all the parts of a decision tree classifier clear by walking through the code that implements the algorithm. The code uses only NumPy, Pandas and the standard python libraries.

The full code can be accessed here.

As of now, the code creates a decision tree when the target variable is binary and the features are numeric. This is completely sufficient to understand the algorithm.

The golden standard of building decision trees in python is the scikit-learn implementation: 1.10. Decision Trees - scikit-learn 0.24.1 documentation