

Übungsblatt: Nichtlineare Verfahren 2/2

CART

In dieser Übung sollen Sie die fehlenden Bestandteile eines vereinfachten CART (Classification And Regression Tree) Algorithmus implementieren.

Tipp: Schreiben Sie sich eigene Testfälle, um die Funktionsweise der verschiedenen Funktionen zu verifizieren bevor Sie das Gesamtergebnis testen.

Aufgabe 1

Implementieren Sie die Berechnung des Gini-Index, wie er in der Vorlesung behandelt wurde, in der Methode *calc_gini_idx*. Mithilfe dieser Methode wird die Unreinheit eines Splits des Baums gemessen.

Aufgabe 2

Implementieren Sie die Methode *evaluate_split*. In dieser Methode soll ein Split durchgeführt werden, sodass das übergebene Datenset durch die Anwendung einer Binär-Entscheidung in zwei Teilmengen geteilt wird.

Aufgabe 3

Implementieren Sie die Methode *find_best_split*. Diese prüft die Qualität jedes möglichen Splits und liefert den Feature-Index, Schwellwert sowie die beiden Teilmengen des übergebenen Datensets (nach dem Split) zurück.