

1. ID3

Implementieren Sie einen Entscheidungsbaum-Klassifikator nach dem ID3-Algorithmus. Wenden Sie Ihren Klassifikator auf den Car Evaluation Data Set¹ an. Der Datensatz ist im CSV-Format gegeben und besteht aus $N = 1728$ Fahrzeugen, deren Qualit  t anhand der Merkmale Kaufpreis, Wartungskosten, Anzahl der T  ren, Anzahl der Sitzpl  tze, Gr   e des Kofferraums, Sicherheit bestimmt wurde. Jede Zeile besteht aus diesen sechs Merkmalen gefolgt von einer Gesamtbewertung. Alle Attribute sind diskret. Der Entscheidungsbaum soll die Gesamtbewertung anhand der sechs Merkmale vorhersagen. Teilen Sie den Datensatz in 80% Trainingsdaten und 20% Testdaten. Erstellen Sie den Entscheidungsbaum aus den Trainingsdaten und messen Sie den Klassifikationsfehler mit den Testdaten.

2. Cross Validation, Jackknife, Bootstrap

Bewerten Sie die Klassifikationsg  te des Entscheidungsbaums mit folgenden Verfahren:

(a) **k-fold Cross Validation**

Teilen Sie den Datensatz zuf  llig in $k = 10$ gleich gro  e Segmente auf. Werten Sie f  r alle 10 Kombinationen den Klassifikationsfehler aus und berechnen Sie den Mittelwert und die Varianz.

(b) **Jackknife**

F  hren Sie Cross Validation mit $k = N$ durch.

(c) **Bootstrap**

Teilen Sie den Datensatz in 80% Trainingsdaten (N_1) und 20% Testdaten. W  hlen sie zuf  llig N_1 Datenpunkte aus (Ziehen mit Zur  cklegen). Erstellen Sie mit diesem Datensatz den Entscheidungsbaum und berechnen Sie den Klassifikationsfehler auf dem Testdatensatz. F  hren Sie dieses Verfahren 1000 Mal durch und berechnen Sie Mittelwert und Varianz des Klassifikationsfehlers.

Hinweis: Bitte bearbeiten Sie die Aufgaben in Zweier-Gruppen und laden Sie alle Ergebnisse (Quelltext + Dokument mit Plots, Tabellen und Erl  uterungen) auf der Vorlesungsseite im Whiteboard hoch². Geben Sie die Namen beider   bungspartner an. Die Bewertung erfolgt bin  r (bestanden/nicht bestanden).

¹ <https://archive.ics.uci.edu/ml/datasets/car+evaluation> → Data Folder

² <https://kvv.imp.fu-berlin.de/portal/site/30550a4f-be9f-4be0-8ec2-a35a294ab3a7>