

Maschinellen Lernen

Aufgabenblatt 07

Prof. Dr. David Spieler
Hochschule München

14. Oktober 2019

Aufgabe 1 (Entscheidungsbäume) *Zur Klassifikation, ob Sie wandern gehen sollten (**ja**, **nein**), haben Sie sich entschlossen einen Entscheidungsbaum zu trainieren. Hierfür haben Sie die Daten in Tabelle 1 erhoben.*

| Datenpunkt | $\mathbf{x}_1^{(i)}$ (Himmel) | $\mathbf{x}_2^{(i)}$ (Temperatur) | $\mathbf{x}_3^{(i)}$ (Luftfeuchtigkeit) | $y^{(i)}$ |
|--------------------|-------------------------------|-----------------------------------|---|-----------|
| $\mathbf{x}^{(1)}$ | frei | mild | normal | ja |
| $\mathbf{x}^{(2)}$ | frei | mild | hoch | ja |
| $\mathbf{x}^{(3)}$ | frei | heiß | normal | nein |
| $\mathbf{x}^{(4)}$ | bewölkt | mild | hoch | nein |
| $\mathbf{x}^{(5)}$ | bewölkt | heiß | normal | ja |
| $\mathbf{x}^{(6)}$ | bewölkt | heiß | hoch | nein |

Tabelle 1: Trainingsdaten für den Entscheidungsbaum.

1. Erstellen Sie einen Entscheidungsbaum nach dem Algorithmus in der Vorlesung mit Hilfe der Entropie als Unreinheitsmaß.
2. Würden Sie dem Entscheidungsbaum nach an einem heißen Tag mit hoher Luftfeuchtigkeit und freiem Himmel wandern gehen?
3. Würden Sie dem Entscheidungsbaum nach an einem milden Tag mit normaler Luftfeuchtigkeit und bewölktem Himmel wandern gehen?

Aufgabe 2 (Praxisbeispiel Entscheidungsbäume) *Wir wollen nun einen Entscheidungsbaum trainieren, welcher von den Features im `Auto.csv` Datensatz nach der Anzahl der Zylinder klassifiziert.*

1. Laden Sie dazu den `Auto.csv` Datensatz in einen `DataFrame` und korrigieren Sie ggf. fehlende bzw. fehlerhafte Daten.
2. Laden Sie in den Zielvektor \mathbf{y} die Zylinderzahlen und wählen Sie als Features \mathbf{X} alle vorhandenen Features außer `name` und `cylinders`.

3. Trainieren Sie einen `sklearn.tree.DecisionTreeClassifier` mit Hilfe 3-facher Kreuzvalidierung.
4. Welche Genauigkeit können Sie erreichen?