

Machine Learning

Übungsblatt 11

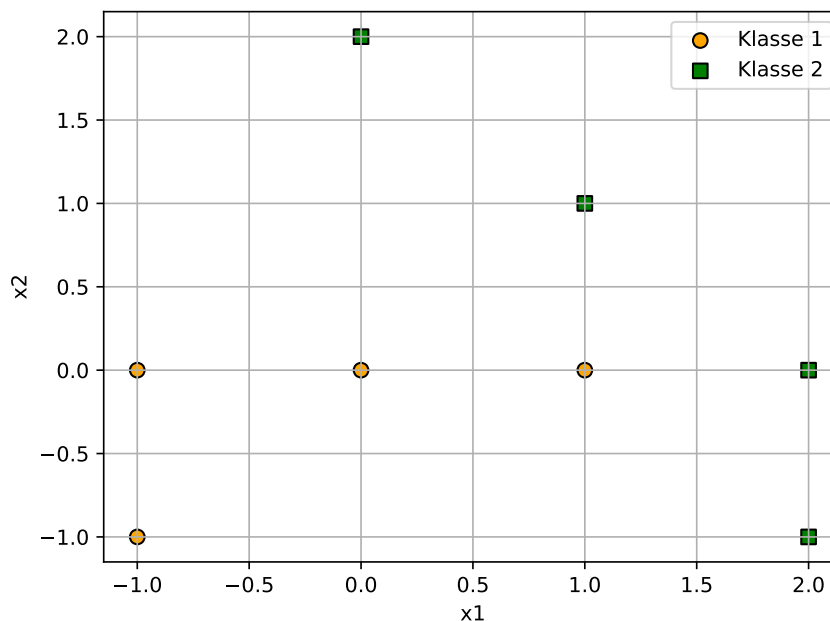
25 Punkte

Hinweis zu Schwellwerten. In [1, Abschnitt 18.1.2], ist die Menge der zu betrachteten Schwellwerten gleich der Menge der Koordinaten der Beobachtungen. Betrachten Sie stattdessen die Mittelwerte der Koordinaten zweier "benachbarter" Beobachtungen.

Aufgabe 1. *Entscheidungsbäume*

18 P.

Die folgende Abbildung zeigt Daten, auf die ein Entscheidungsbaum Modell gefittet werden soll.



- (a) Geben Sie einen Entscheidungsbaum minimaler Tiefe an, der die Trainingsdaten korrekt klassifiziert, und zeichnen Sie die Entscheidungsgrenze ein; vgl. [1, Abbildung 18.3 (a) & (b)].
- (b) Bestimmen Sie einen Entscheidungsbaum der Tiefe 1. Nutzen Sie dazu den Algorithmus aus [1, Abschnitt 18.1.2].

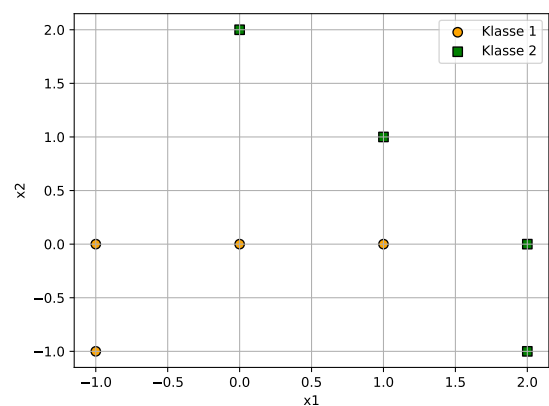
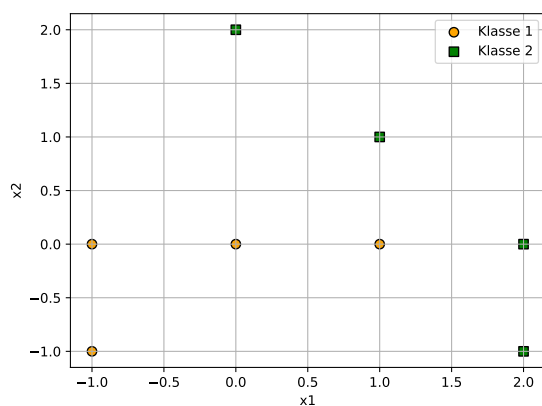
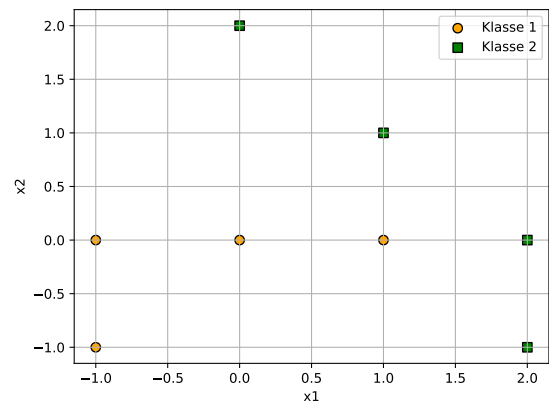
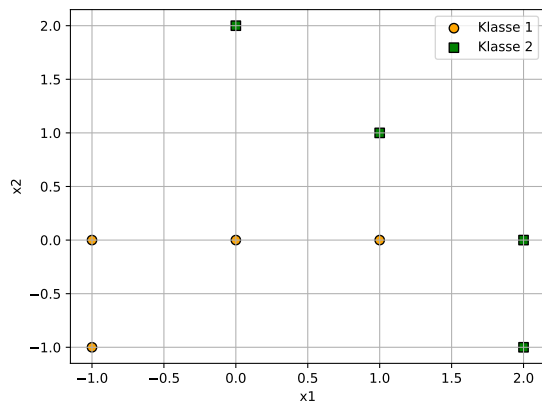
- (i) Geben Sie zunächst die Menge aller zu betrachteten Schwellwerte an, sowie die Anzahl der Daten pro Klasse in den resultierenden Zerlegungen.
- (ii) Geben Sie an, welche Schwellwerte zum gleichen Wert des Verzweigungskriterium $f(j, t)$ führen, wobei

$$f(j, t) = \frac{|\mathcal{D}^L(j, t)|}{|\mathcal{D}|} c(\mathcal{D}^L(j, t)) + \frac{|\mathcal{D}^R(j, t)|}{|\mathcal{D}|} c(\mathcal{D}^R(j, t)) , \quad \text{Vergleiche [1, Gleichung (18.5)]}.$$

Begründen Sie außerdem, für welche Schwellwerte das Verzweigungskriterium nicht berechnet werden muss, da dieses definitiv größer ist, als das eines anderen Schwellwerts.

- (iii) Berechnen Sie für die restlichen Schwellwerte die Gini Koeffizienten und f .
- (iv) Geben Sie ein Modell mit minimalem f an. Ist das Modell eindeutig?
- (c) Bestimmen Sie nun einen Entscheidungsbaum der Tiefe 2, indem Sie einen Entscheidungsbaum der Tiefe 1 auf die Zerlegung aus Aufgabenteil (b) fitten. Sollte die Zerlegung aus (b) nicht eindeutig sein, so ist für jede Zerlegung ein Entscheidungsbaum zu bestimmen. Geben Sie die resultierenden Entscheidungsbäume an und zeichnen Sie die Entscheidungsgrenzen in folgende Abbildungen ein.

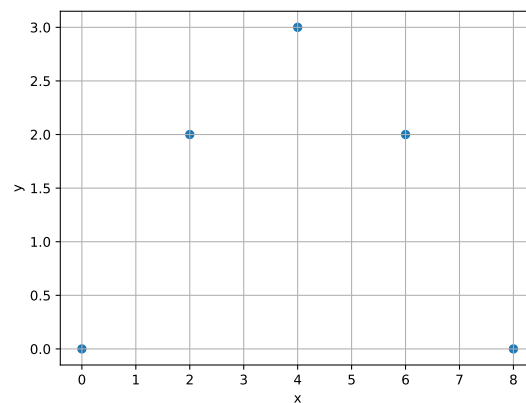
Auch in diesem Aufgabenteil empfiehlt es sich, zuerst zu überlegen (und zu begründen), welche Schwellwerte in Frage kommen. Sollte dies nur ein Schwellwert sein, so müssen Sie die Gini Koeffizienten nicht berechnen.



Aufgabe 2. Regressionsbäume

7 P.

Gegeben sind die folgenden Daten, wobei x die Input- und u die Zielvariable ist.



Bestimmen Sie einen Regressionsbaum $f(x; \theta) = \sum_j w_j \mathbb{1}_{x \in R_j}$ der Tiefe 2.

- Bestimmen Sie zunächst die Regionen R_j indem Sie die zugehörigen Schwellwerte mithilfe des Algorithmus aus [1, Abschnitt 18.1.2] bestimmen. Gehen Sie dazu wie in Aufgabe 1 (b) vor. Beachten Sie außerdem den [Hinweis zu den Schwellwerten](#).
- Bestimmen Sie nun die Gewichte w_j .
- Zeichnen Sie die Funktion f in obige Abbildung ein.

Literatur

[1] Kevin P. Murphy, Probabilistic Machine Learning - An Introduction, MIT Press, 2022.