

Aufgaben

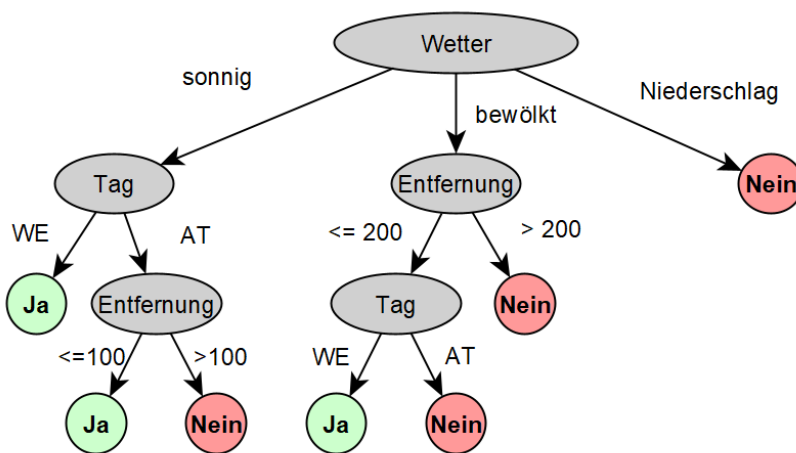
Unterrichtsmodul Einführung maschinelles Lernen mit Entscheidungsbäumen

Pascal Schmidt, Stefan Strobel



Aufgabe 1

Björn ist ein begnadeter Skifahrer. Durch die Auswertung einiger Daten aus seinem persönlichen Kalender ist folgender *Entscheidungsbaum* entstanden:



Legende

- *Entfernung*: Wegstrecke (in km) bis zum nächstgelegenen Skigebiet mit guten Schneeverhältnissen
- *Tag*: Wochenende (WE) bzw. Arbeitstag (AT)

- a) Treffe mithilfe dieses Entscheidungsbaums eine Vorhersage, ob Björn an den folgenden Tagen Skifahren gehen wird.

	Wetter	Tag	Entfernung
1	Niederschlag	WE	78
2	sonnig	AT	210
3	bewölkt	AT	145
4	bewölkt	WE	145

- b) Beschreibe die Struktur (den Aufbau) eines Entscheidungsbaumes. Verwende dazu möglichst die von Baumstrukturen bekannten Fachbegriffe.

Aufgabe 2

In dieser Aufgabe geht es um den *Prozess der Baumerzeugung*. Wir betrachten daher ein stark vereinfachtes Szenario:

Frau Dr. Müller ist Landärztin in Garmisch-Patenkirchen. Aufgrund dieser Tätigkeit muss sie trotz vollen Terminkalenders regelmäßig Patienten zu Hause besuchen. Wie schnell sie dabei mit ihrem Jeep fahren kann (*hohe* vs. *niedrige* Geschwindigkeit), hängt vom Zustand der Straße (*gut*, *schlecht*) und ihrer Steigung (*flach*, *geneigt*, *steil*) ab.

- a) Die folgenden Daten sind während Frau Müllers Fahrten erhoben worden. Erstelle einen Entscheidungsbaum, der diese Daten korrekt klassifiziert.

	Straße	Steigung	Geschwindigkeit
1	gut	flach	hoch
2	schlecht	steil	niedrig
3	gut	geneigt	hoch
4	gut	flach	hoch
5	schlecht	flach	hoch
6	gut	steil	niedrig
7	schlecht	geneigt	niedrig
8	gut	geneigt	hoch
9	schlecht	steil	niedrig

- b) Klassifiziere das folgende, neu hinzukommende Datenelement mithilfe deines Entscheidungsbaumes. Beschreibe deine Beobachtung.


	Straße	Steigung	Geschwindigkeit
10	gut	flach	niedrig

Hättest du einen anderen Entscheidungsbaum erzeugt, wenn dir dieses Datenelement von Anfang an bekannt gewesen wäre? Begründe kurz!

Aufgabe 3

Inzwischen hast du einen ersten Entscheidungsbaum von Hand erstellt, *Duplikate* und *widersprüchliche Daten* sollten keine Fremdwörter mehr sein. Die folgende Situation ist bereits deutlich komplexer. Erstelle einen Entscheidungsbaum, der möglichst viele Datenelemente korrekt klassifiziert.

ID	Größe?	Farbe?	Form?	Klasse
1	groß	sw	rund	A
2	klein	bunt	rund	A
3	klein	sw	eckig	B
4	groß	bunt	eckig	B
5	klein	bunt	eckig	B
6	klein	sw	eckig	B
7	klein	sw	rund	A
8	klein	bunt	eckig	A
9	klein	bunt	eckig	A
10	groß	bunt	rund	A
11	groß	sw	eckig	B
12	klein	bunt	eckig	A
13	groß	sw	eckig	B

Hilfsmittel: Datenelemente zum Ausschneiden 

groß 1
sw
rund

klein 2
bunt
rund

klein 3
sw
eckig

groß 4
bunt
eckig

klein 5
bunt
eckig

klein 6
sw
eckig

klein 7
sw
rund

klein 8
bunt
eckig

klein 9
bunt
eckig

groß 10
bunt
rund

groß 11
sw
eckig

klein 12
bunt
eckig

groß 13
sw
eckig

Aufgabe 4

In dieser Aufgabe lernst du die *systematische* Erzeugung eines *möglichst kompakten* Entscheidungsbaums kennen. Dies ist ein wichtiger Baustein des Lernalgorithmus, der regelbasiert und automatisch aus Daten einen Entscheidungsbaum konstruiert.

Wir verwenden dazu den Datensatz aus Aufgabe 3 (Merkmale: *Größe, Farbe, Form*; Klassen: *A, B*).

- Bei der Konstruktion eines Baumes müssen wir zuerst das zu testende Merkmal für den Wurzelknoten auswählen. Nimm an, der Baum dürfte nur diesen einen Knoten mit zwei Blättern enthalten. Stelle die möglichen Situationen dar und beschrifte jeweils die beiden Blätter begründet mit den Klassenbezeichnern *A* bzw. *B*.
- Welchen der Entscheidungsbäume aus Teil a) würdest du verwenden, wenn du dich für einen entscheiden müsstest? Begründe deine Entscheidung kurz.
- Verbessere den ausgewählten Entscheidungsbaum, indem du die beiden Blätter durch innere Knoten ersetzt. Wie gehst du dabei vor? Orientiere dich an dem Auswahlverfahren aus Teil a).

Aufgabe 5

Peter plant zu seinem 18. Geburtstag eine Party, zu der er MitschülerInnen seiner Klassenstufe einladen möchte. Von zwölf MitschülerInnen sind einige persönliche Informationen und Peters Entscheidung bezüglich ihrer Einladung bekannt.

- Wende den *Lernalgorithmus* an, um aus den folgenden Daten einen Entscheidungsbaum zu konstruieren.

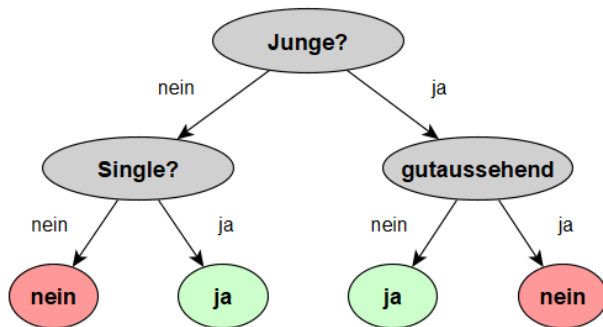
	gleicher Kurs?	Junge?	Single?	gutaussehend?	Einladung
1	nein	nein	nein	nein	Nein
2	nein	nein	nein	ja	Nein
3	nein	nein	ja	nein	Ja
4	nein	nein	ja	ja	Ja
5	nein	ja	nein	ja	Nein
6	nein	ja	ja	nein	Ja
7	ja	nein	nein	nein	Nein
8	ja	nein	ja	nein	Ja
9	ja	nein	ja	ja	Ja
10	ja	ja	nein	nein	Ja
11	ja	ja	nein	ja	Nein
12	ja	ja	ja	ja	Nein



Hilfsmittel: Datenelemente zum Ausschneiden

K: nein J: nein S: nein g: nein	K: nein 2 J: nein S: nein g: ja	K: nein 3 J: nein S: ja g: nein	K: nein 4 J: nein S: ja g: ja	K: nein 5 J: ja S: nein g: ja	K: nein 6 J: ja S: ja g: nein
K: ja 7 J: nein S: nein g: nein	K: ja 8 J: nein S: ja g: nein	K: ja 9 J: nein S: ja g: ja	K: ja 10 J: ja S: nein g: nein	K: ja 11 J: ja S: nein g: ja	K: ja 12 J: ja S: ja g: ja

b) Der folgende Entscheidungsbaum klassifiziert ebenfalls alle Datenelemente korrekt:



Interpretiere diesen Baum! Nach welchen Regeln scheint Peter bei seinen Entscheidungen vorgegangen zu sein?

c) Begründe, weshalb der Lernalgorithmus diesen optimalen Baum nicht erzeugt.

Aufgabe 6

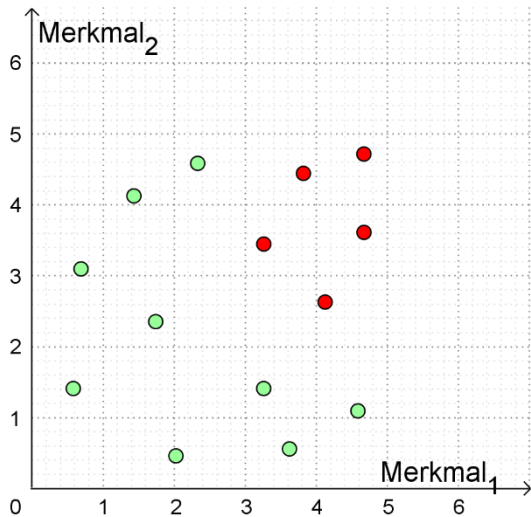
Gegeben ist die folgende Liste von Datenelementen:

	Merkmal_1	Merkmal_2	Klasse
1	4,12	2,63	B
2	4,59	1,09	A
3	3,82	4,45	B
4	1,43	4,13	A
5	2,33	4,59	A
6	3,62	0,56	A
7	0,69	3,1	A
8	3,26	3,45	B
9	4,67	4,72	B
10	3,26	1,41	A
11	1,74	2,35	A
12	4,67	3,61	B
13	0,58	1,41	A
14	2,02	0,46	A

a) Erstelle mit dem bisher erarbeiteten Lernalgorithmus einen zu den Daten konsistenten Entscheidungsbaum. Welches Problem ergibt sich bei der anschließenden Klassifikation des Datenelements (1,31; 0,94)?

b) Im folgenden Koordinatensystem, einem sogenannten *Scatterplot*, sind die Daten grafisch veranschaulicht. Separiere die Datenelemente unterschiedlicher Klassen, indem du (möglichst wenige) horizontale und vertikale Linien in das Koordinatensystem einzeichnest.

Zeichne anschließend einen Entscheidungsbaum, der die Aufteilung entsprechend dieser Linien nachbildet.



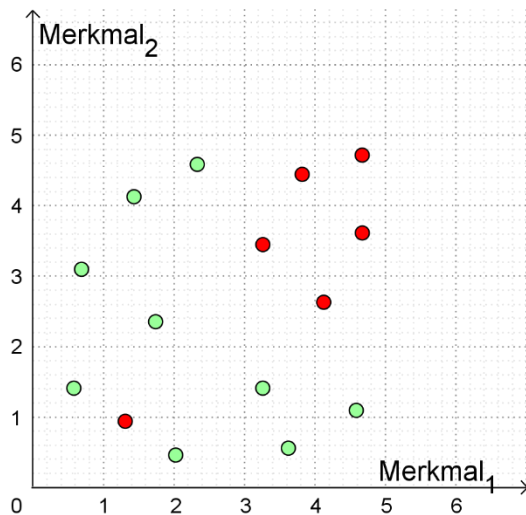
c) Wie klassifiziert der neue Entscheidungsbaum das Datenelement (1,31; 0,94)?

Aufgabe 7

Wir ergänzen den Datensatz aus Aufgabe 6 um folgendes Datenelement:

	Merkmal_1	Merkmal_2	Klasse
15	1,31	0,94	B

Grafisch wird diese Situation durch folgenden Scatterplot dargestellt:



- Zeichne (möglichst wenige) horizontale und vertikale Linien ein, sodass die Datenelemente entsprechend ihrer Klassenzugehörigkeit separiert werden. Zeichne anschließend einen dieser Aufteilung entsprechenden Entscheidungsbaum.
- Begründe, dass es trotz der offensichtlichen Fehlklassifikation eines Datenelements auch in dieser Situation sinnvoll sein könnte, einen der Entscheidungsbäume aus Aufgabenteil 6b) zu verwenden.