

Arbeitsblatt 1: Daten mithilfe von Entscheidungsbäumen klassifizieren

**Aufgabe**

Björn ist ein begnadeter Skifahrer. Er ist beruflich flexibel und kann es daher sowohl unter der Woche als auch am Wochenende einrichten, einen Tag Skifahren zu gehen. Bei der Entscheidung, ob er Skifahren geht, spielen folgenden Fragen eine Rolle:

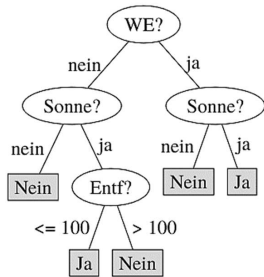


- Ist es *sonnig*?  
mögliche Antworten: ja, nein
- Handelt es sich um einen Tag am *Wochenende*?  
mögliche Antworten: ja, nein
- Wie weit ist das nächstgelegene Skigebiet mit guten Schneeverhältnissen *entfernt*?  
mögliche Antworten:  $\leq 100$  km,  $> 100$  km

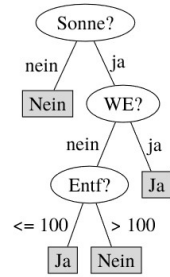
InformatikerInnen sprechen in diesem Zusammenhang nicht von *Frage* und *Antwort*, sondern von **Merkmal** und **Ausprägung**.

Björn möchte seine Entscheidung nicht aus dem Bauch heraus treffen, sondern einen **Entscheidungsbaum** nutzen. Nachfolgend sind zwei unterschiedliche Entscheidungsäume dargestellt, die jeweils die Frage „Lohnt es sich, heute Skifahren zu gehen?“ beantworten.

Entscheidungsbaum e1



Entscheidungsbaum e2



a) Bestimme Björns Entscheidung in folgende Situationen, in Abhängigkeit von dem von ihm gewählten Entscheidungsbaum.

ID	Merkmale			Entscheidung	
	Sonne?	WE?	Entf?	Baum e1	Baum e2
1	nein	ja	$\leq 100$		
2	ja	ja	$\leq 100$		
3	ja	nein	$> 100$		
4	nein	nein	$\leq 100$		
5	ja	ja	$> 100$		

b) (*Zusatzaufgabe*) Erstelle einen weiteren Entscheidungsbaum, der in *allen* Situationen die gleiche Antwort liefert, aber in der Wurzel das Merkmal *Entfernung* abfragt.

Arbeitsblatt 2: Entscheidungsbäume (intuitiv) erstellen

**Aufgabe**

Gegeben sind mehrere Listen von Datenelementen mit den zugeordneten Entscheidungen (rechte Spalte).

Situation 1

ID	Merkmale		Entscheidung
	Sonne?	Wind?	Sport
1	ja	ja	Ja
2	ja	nein	Nein
3	nein	ja	Nein
4	nein	nein	Ja

Situation 2

ID	Merkmale		Entscheidung
	Terrain?	Straße?	Geschwindigkeit
1	flach	eben	hoch
2	flach	holprig	niedrig
3	steil	eben	niedrig
4	steil	holprig	niedrig

Situation 3

ID	Merkmale			Entscheidung
	Größe?	Farbe?	Form?	Klasse
1	groß	bunt	rund	A
2	klein	bunt	rund	A
3	groß	bunt	eckig	B
4	groß	sw	eckig	A
5	klein	bunt	eckig	B
6	klein	sw	eckig	A
7	groß	sw	rund	A

- a) Welche Merkmale werden in den verschiedenen Situationen betrachtet? Welche Entscheidung soll jeweils getroffen werden?
- b) Erstelle zu jeder Situation einen Entscheidungsbaum, der die gegebenen Daten korrekt klassifiziert.

## Unterrichtsmodul: „Lernende Algorithmen“

unplugged-Aufgaben

### Arbeitsblatt 3: Entscheidungsbäume vergleichen

#### Aufgabe

(Zur Bearbeitung in Vierergruppen) 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:

	gleicher Bezugskurs?	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

- Erstellt (arbeitsteilig) vier Entscheidungsbäume, die zu genau diesen Entscheidungen führen. Verwendet hierbei jeweils unterschiedliche Startmerkmale für die Abfrage in der Wurzel.
- Vergleicht die in eurer Gruppe entstandenen Bäume hinsichtlich Höhe und Knotenanzahl.
- Gibt es noch weitere Entscheidungsbäume, die alle Daten korrekt klassifizieren? Begründet eure Antwort.

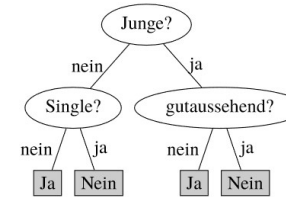
## Unterrichtsmodul: „Lernende Algorithmen“

unplugged-Aufgaben

### Arbeitsblatt 4: Entscheidungsbaum analysieren

#### Aufgabe

Tatsächlich hat Peter den folgenden Entscheidungsbaum genutzt:



- Zeige exemplarisch anhand von 2-3 Beispielen, dass dieser Baum die Datenelemente von Aufgabenblatt 3 korrekt klassifiziert.
- Vergleiche diesen Baum mit den zuvor erstellten Entscheidungsbäumen (vgl. Arbeitsblatt 3). Welche Unterschiede fallen dir auf? Welche(n) Vorteil(e) bietet dieser Entscheidungsbaum gegenüber den anderen Bäumen?
- Interpretiere den Baum als *Regelwerk*: Formuliere eine Bedingung, die eine Person erfüllen muss, um zu Peters Party eingeladen zu werden (Formulierungshilfe: „Wenn ..., dann ...“).
- Leite aus dem Entscheidungsbaum eine begründete Vermutung über Peters Motiv bei der Auswahl seiner Gäste her.

## Unterrichtsmodul: „Lernende Algorithmen“

unplugged-Aufgaben

### Arbeitsblatt 5: Merkmale geschickt auswählen (Teil 1)

#### Aufgabe

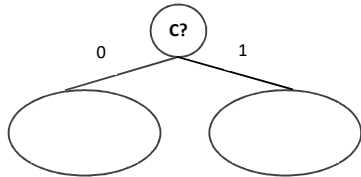
Bei der Konstruktion eines Entscheidungsbaumes müssen wiederholt Merkmale zur Durchführung von Tests ausgewählt werden. Da diese Auswahl Einfluss auf die Größe des Entscheidungsbaumes hat, sollte sie möglichst geschickt erfolgen. Wir untersuchen dies beispielhaft für die erste Testauswahl (Wurzel) anhand folgender Datenliste:

ID	A	B	C	D	Klasse
d1	0	0	0	0	Ja
d2	0	1	0	0	Ja
d3	0	0	0	1	Nein
d4	0	1	0	1	Nein
d5	1	0	0	0	Ja
d6	1	0	0	0	Ja
d7	1	1	0	1	Nein
d8	1	1	0	1	Nein

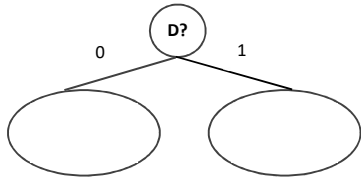
Stellt man die Klassenzuordnung farblich dar, so ergibt sich folgende grafische Darstellung für die Gesamtheit der Datenelemente:

d1, d2, d3, d4  
d5, d6, d7, d8

- a) Stelle die Situation für einen initialen Test anhand von Merkmal C dar und begründe, dass sich dieses Merkmal nicht eignet.



- b) Stelle die Situation für einen initialen Test anhand von Merkmal D dar und begründe, dass es sich besonders eignet.



- c) Während die Splits anhand von Merkmal C und D gewissermaßen Extreme darstellen, zeigt folgende Abbildung (Splits nach A bzw. B) Situationen, die in dieser oder ähnlicher Form häufig vorkommen.



Nimm an, du müsstest dich für eine dieser beiden Varianten entscheiden: Welche wählst du? Begründe deine Antwort!

## Unterrichtsmodul: „Lernende Algorithmen“

unplugged-Aufgaben

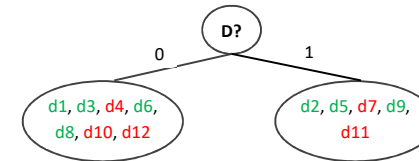
### Arbeitsblatt 6: Merkmale geschickt auswählen (Teil 2)

#### Aufgabe 1

Gegeben ist eine Liste von Datenelementen, zu der ein möglichst kleiner Entscheidungsbaum konstruiert werden soll. Wir versuchen, das beste Merkmal für den ersten Test (Wurzel) ausfindig zu machen.

	A	B	C	D	Klasse
d1	0	0	0	0	ja
d2	0	0	1	1	ja
d3	0	1	0	0	ja
d4	0	1	1	0	nein
d5	0	1	1	1	ja
d6	1	0	0	0	ja
d7	1	0	0	1	nein
d8	1	0	1	0	ja
d9	1	0	1	1	ja
d10	1	1	0	0	nein
d11	1	1	0	1	nein
d12	1	1	1	0	nein

- a) Die folgende Abbildung zeigt die Situation für einen Split nach Merkmal D. Stelle jeweils die Situation dar, wenn für den ersten Split eines der anderen Merkmale verwendet wird.



- b) Welches dieser vier „Anfangsstücke“ würdest du zur Erstellung eines Entscheidungsbaumes nutzen? Begründe deine Entscheidung!

#### Aufgabe 2

- a) Berechne den *Gini-Index* der Datenliste aus Aufgabe 1.  
b) Bestimme für die beiden Kinderknoten, die in Aufgabe 1 bei einem Split nach Merkmal D bzw. Merkmal B entstehen, jeweils den *Gini-Index* der enthaltenen Daten.

Arbeitsblatt 7: Gini-Index nutzen

**Aufgabe**

Gegeben ist eine Datenliste mit 100 Elementen, von denen jeweils 50 zur Klasse „ja“ bzw. „nein“ gehören. Hierzu soll ein möglichst kompakter Entscheidungsbaum erzeugt werden. In der Abbildung ist die Situation für den initialen Split nach Merkmal A bzw. B dargestellt.

Julia und Peter, die in ihrem Informatikunterricht den *Gini-Index* kennengelernt haben, sind sich uneins, welches der beiden Merkmale für den ersten Split gewählt werden sollte: Während Peter für die Wahl des Merkmals A plädiert, findet Julia einen Split nach Merkmal B sinnvoller.



- Berechne den *Gini-Index* für jeden der vier dargestellten Kinderknoten.
- Wie könnte man Julias bzw. Peters Standpunkt unter Verwendung des *Gini-Index* begründen? Liefere für jeden der beiden (mindestens) ein Argument.
- Wie würdest du den *Gini-Index* verwenden, um ein Merkmal unter mehreren Alternativen auszuwählen?

Arbeitsblatt 8: Induktive Baumerzeugung

**Aufgabe**

In dieser Aufgabe erstellen wir von Hand einen vollständigen Entscheidungsbaum. Als Auswahlkriterium für die Tests dient der *gewichtete Gini-Index*. Wir nutzen erneut die Datenliste von Arbeitsblatt 6:

	A	B	C	D	Klasse
d1	0	0	0	0	ja
d2	0	0	1	1	ja
d3	0	1	0	0	ja
d4	0	1	1	0	nein
d5	0	1	1	1	ja
d6	1	0	0	0	ja
d7	1	0	0	1	nein
d8	1	0	1	0	ja
d9	1	0	1	1	ja
d10	1	1	0	0	nein
d11	1	1	0	1	nein
d12	1	1	1	0	nein

- Zeige rechnerisch, dass Merkmal B für den ersten Test herangezogen werden sollte.
- Setze das Verfahren fort und entwickle damit den vollständigen Entscheidungsbaum.