Universität Kassel
Fachbereich Mathematik/Informatik
Fachgebiet Wissensverarbeitung
Hertie-Stiftungslehrstuhl
Wilhelmshöher Allee 73
34121 Kassel

Email: hotho@cs.uni-kassel.de
Tel.: ++49 561 804-6252

Dr. Andreas Hotho
Prof. Gerd Stumme

29.06.04

Knowledge Discovery

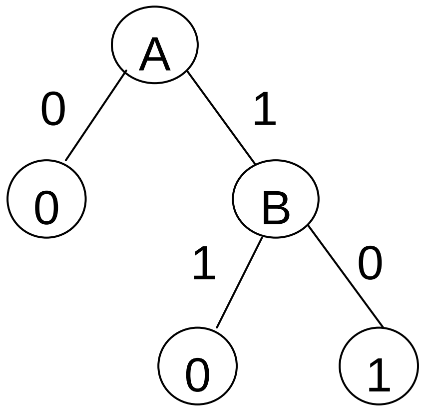
Lösungsblatt 3

Sommersemester 2004

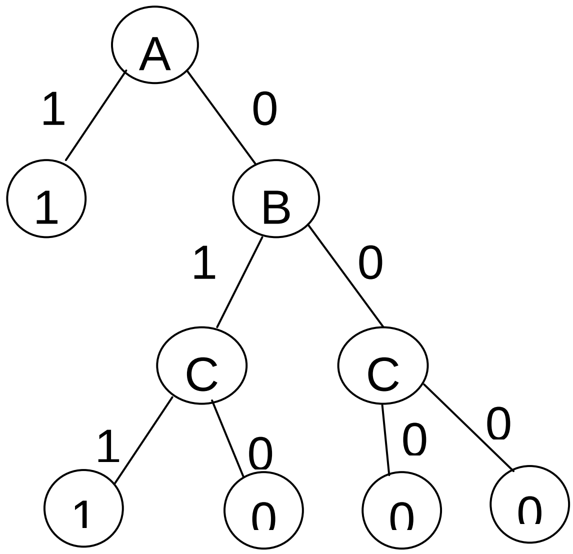
Aufgabe 1:

Geben Sie für die folgenden booleschen Funktionen einen Entscheidungsbaum an:

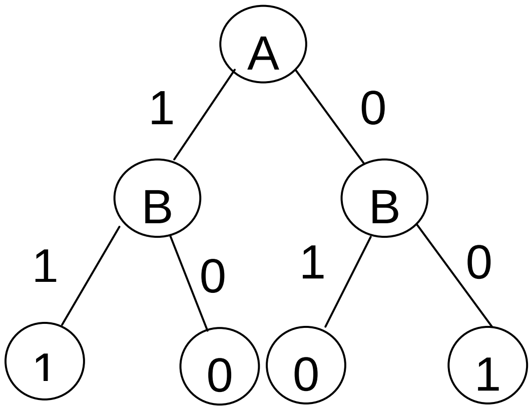
a) $A \wedge \neg B$



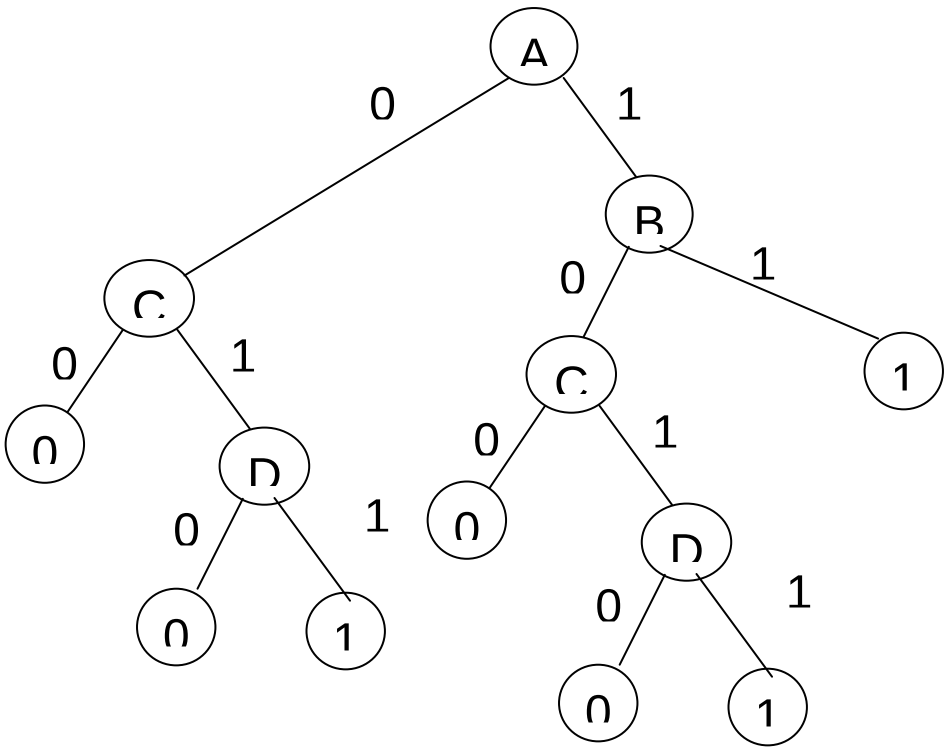
b) $A \vee [B \wedge C]$



c) A XOR B



d) $[A \wedge B] \vee [C \wedge D]$



Existiert Ihrer Meinung nach eine ähnliche (oder komplexere) Funktion für Aufgabe 2 des letzten Blattes.

Ja für die Trainingsdaten. Das entspricht dem Ziel der Entscheidungsbauverfahren, dem Erlernen solcher Funktionen aus Daten. Hinterfragt man aber die Existenz einer solchen Funktion im allgemeinen, so kann die Antwort auch nein sein. Vielleicht gibt es ja keine Regelmäßigkeiten, die wirklich eine Vorhersage erlauben. Um Entscheidungsbäume sinnvoll einsetzen zu können, muß man die Annahmne treffen, daß es so eine Funktion gibt. Ob die Annahme gerechtfertigt ist, muß man also vorher ja nach Kontext entscheiden.

Aufgabe 2:

a) Geben Sie ein Prinzip an, mit dem man eine aussagenlogische Beschreibung einer (im folgenden als positiv bezeichneten) Klasse aus dem Entscheidungsbaum erzeugen kann.

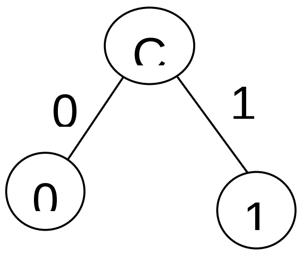
- Für alle Blattknoten der positiven Klasse:
  Verbinde Test der einzelnen Knoten des Baumes von der Wurzel bis zum Blatt mit dem logischen UND
- Alle Pfade werden mit dem logischen ODER verknüpft.

(If $(test_1 \wedge test_2 \wedge \ldots \wedge test_n) \vee (test_x \wedge test_y \wedge \ldots \wedge test_z)$ THEN class= positiv)

b) Machen Sie das für den Entscheidungsbaum des letzten Übungsblattes für die gesunden Patienten.

IF (HR = regular) $\wedge$ (BP = normal) then (Klasse = healthy)

c) Ein Entscheidungsbaum E2 entsteht durch einen zusätzlichen Split aus dem Entscheidungsbaum E1. Ist dann die zu E1 gehörende Formel allgemeiner als die zu E2 gehörende? (Beweis oder Gegenbeispiel)



C	D	Class
0	0	0
0	1	0
1	0	1
1	1	1

Gegenbeispiel für Baum E1: If (C = 1) then (class = 1), fügt man einen weiteren Knoten ein, so muss die Anzahl der Beispiele, die mit „1“ klassifiziert werden kleiner werden. Bei der Formel if (C = 1) $\vee$ (D = 1) wird aber die Anzahl der Einsen größer. Damit ist der Baum durch einem zusätzlichen Split allgemeiner geworden.

Aufgabe 3:

a) Berechnen sie für folgenden Datensatz die in der Vorlesung im Detail vorgestellten informationstheoretischen Maße und interpretieren Sie kurz das Ergebnis!

ID	Attribut I	Attribut II	Klasse
1	5	xxx	0
2	42	xxx	1
3	5	zzz	0
4	5	xxx	0
5	42	xxx	1
6	42	zzz	0
7	42	xxx	1
8	42	xxx	1

Im Detail wurden folgende Maße vorgestellt:

- Attributentropie H(B):

$$H(B) = -\sum_{i=1}^{K} p_i \log_2(p_i)$$

- Klassenentropie H(C): siehe H(B)

- Equivalent Number of Attributes

$$EN.attr = \frac{H(c)}{\bar{I}_{gain}(c,A)} \qquad \text{with}$$

$$\bar{I}_{gain}(c,A) = \frac{1}{s}\sum_{i=1}^{s} I_{gain}(c,A_i)$$

Für die Joint Entropy und den Gini Index wurde keine Formel angegeben.

Entropie für ID: $-8\left(\frac{1}{8}\log_2(\frac{1}{8})\right) = -\log_2(\frac{1}{8}) = 3$

Attr 1: $-\left(\frac{3}{8}\log(\frac{3}{8}) + \frac{5}{8}\log(\frac{5}{8})\right) = -\left(-0.531 + (-0.424)\right) = 0.955$

Attr 2: $-\left(\frac{2}{8}\log(\frac{2}{8}) + \frac{6}{8}\log(\frac{6}{8})\right) = -\left(-0.5 + (-0.311)\right) = 0.811$

Klasse: $-\left(\frac{4}{8}\log(\frac{4}{8}) + \frac{4}{8}\log(\frac{4}{8})\right) = -\left(-0.5 + (-0.5)\right) = 1$

Equivalent Number of Attributes:

H(C) = 1

und

$$H(c \mid A_i) = -\sum_{l=1}^{n} \frac{|T_l|}{|T|}\left(\frac{|c \cap T_l|}{|T_l|}\log_2\frac{|c \cap T_l|}{|T_{li}|}\right)$$

$$\mathrm{En.attr(c)} = \frac{\mathrm{H(c)}}{\left(\mathrm{H(c)} - \frac{1}{\mathrm{s}}\sum_{i=1}^{n} H(c \mid A_i)\right)}$$

c = 0 ➔

$$\mathrm{h(c)} = -\frac{4}{8}\log_2(\frac{4}{8}) = -\frac{1}{2}(1-2) = \frac{1}{2}$$

$H(c \mid A_{AttrI}) = \frac{5}{8}\left(\frac{1}{5}\log_2\frac{1}{5}\right) + \frac{3}{8}\left(\frac{3}{3}\log_2\frac{3}{3}\right) = \frac{5}{8}\frac{4}{5} + \frac{3}{8}0 = \frac{1}{2}$ (bei Nutzung von $\log_2 = 0.29$)

$H(c \mid A_{AttrII}) = \frac{6}{8}\frac{2}{3} + \frac{2}{8}0 = \frac{1}{2}$ (bei Nutzung von $\log_2 = 0.39$)

$\mathrm{En.attr(0)} = \frac{\frac{1}{2}}{\left(\frac{1}{2} - \frac{1}{2}1\right)}$ (bei Nutzung von $\log_2$ ca. 3)

b) Wäre es möglich, den Mittelwert für Attribut I zu berechnen? Wie sinnvoll ist dies?

Ja, es ist natürlich möglich, den Mittelwert zu berechnen. Er ist aber vermutlich nicht sinnvoll, da die Entstehung der Zahlen und deren Bedeutung aus der Tabelle nicht ersichtlich wird. Die Verteilung der Zahlen in der Tabelle lässt eher den Schluß zu, dass die Zahlen ein kategorisches Attribut kodieren. Das wäre aber noch mit einem Domänenexperten zu klären.

Aufgabe 4: Begriffe aus dem Bereich des Maschinellen Lernens

a) Maschinelle Lernverfahren zeichnen sich durch die Eigenschaft des inkrementellen bzw. des nicht-inkrementelles Lernen aus. Was charakterisiert Verfahren mit dieser Eigenschaft?

Inkrementelle vs. nicht-inkrementelle Lernverfahren:

inkrementell: Das Erlernen erfolgt sukzessive auf immer mehr/neuen Instanzen des Gesamtdatenbestandes.
Lernen und Anwendung können überlappen, d.h. Konzeptbeschreibung wird kontinuierlich verbessert. Problem: Ergebnis oft von der Reihenfolge der Verarbeitung abhängig
nicht-inkrementell: Operation wird auf gesamter gegebener Datenbasis ausgeführt, alle Beispiele müssen zu Beginn vorliegen.

b) Was unterscheidet induktives Schließen von deduktivem Schließen?

Induktives vs. deduktives Schließen

Induktives Schließen: Def. Induktion: Prozess des plausiblen Schließens vom Speziellen zum Allgemeinen.
Grund für die Plausibilität ist i.a. eine große Anzahl zutreffender Fälle.
Bsp.: Die meisten Raben sind schwarz => alle Raben sind schwarz

Deduktives Schließen: Deduktion = “Ableitung“ - Nutzen vorhandenen Wissens
- Neuformulierung vorhandenen Wissens

Bsp.: Alle Meschen sind sterblich. Sokrates ist ein Mensch => Sokrates ist sterblich

c) Induktive Hypothesen werden durch die Begriffe "Vollständigkeit" und "Konsistenz" gekennzeichnet. Beschreiben Sie diese beiden Begriffe graphisch.

Induktive Hypothese:
Lernen eines Begriffs: Finde eine Hypothese H, die konsistent und vollständig in Bezug auf alle Trainingsbeispiele ist. Konsistent heißt dabei, dass kein negatives Träningsbeispiel falsch eingeteilt wird und vollständig, dass keine positives Trainingsbeispiel ausgelassen wird.