

Informationssysteme II

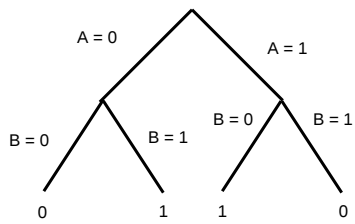
Übungsblatt 6

Christine Pries, Mareike Wagner, Felix Oppermann (Gruppe 18)

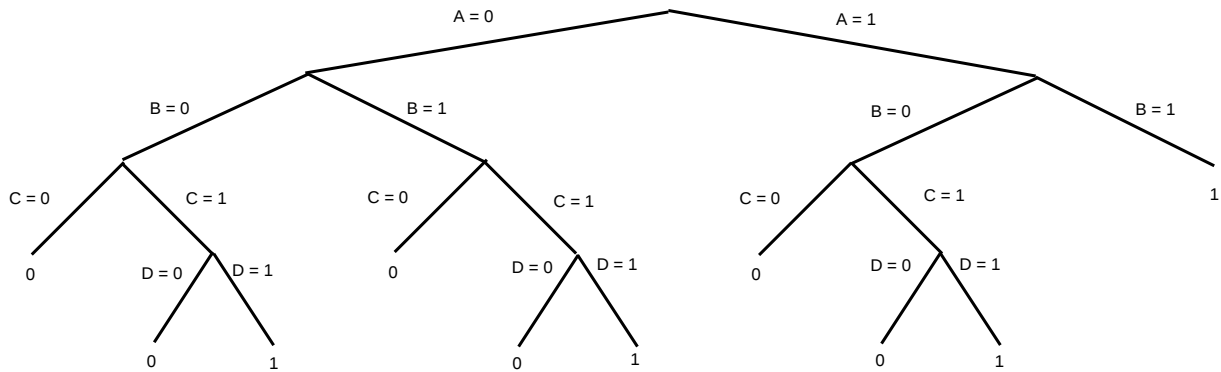
30. Juni 2005

Aufgabe 1 - Entscheidungsbaum (28 Punkte)

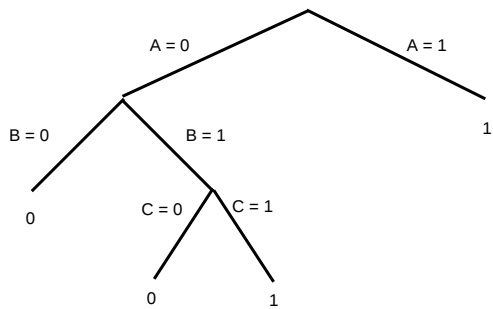
(a) $A \text{ XOR } B$



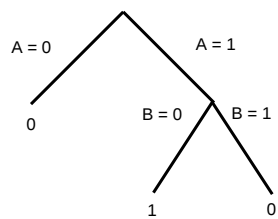
(b) $[A \wedge B] \vee [C \wedge D]$



(c) $A \vee [B \wedge C]$



(c) $A \wedge \text{not } B$



Aufgabe 2 - C4.5 Algorithmus (50 Punkte)

Eingabe

Beispielmenge $A =$

$\{(Irregular, Normal, Ill),$
 $(Regular, Normal, Healthy),$
 $(Irregular, Abnormal, Ill),$
 $(Irregular, Normal, Ill),$
 $(Regular, Normal, Healthy),$
 $(Regular, Abnormal, Ill),$
 $(Regular, Normal, Healthy),$
 $(Regular, Normal, Healthy)\}$

Schritt 1

Wurzelknoten des Entscheidungsbaums B ist A

Schritt 2

Nicht alle Beispiele aus A gehören zur gleichen Klasse und $|A| \neq 0$

Schritt 3

a) – **Information vor dem Split**

$$info([4, 4]) = entropy(\frac{4}{8}, \frac{4}{8}) = -\frac{1}{2} * \log_2(\frac{1}{2}) - \frac{1}{2} * \log_2(\frac{1}{2}) = 1 \text{ bits}$$

– **Informationsgewinn für Heart Rate**

Heart Rate = irregular

$$info([0, 3]) = entropy(\frac{0}{3}, \frac{3}{3}) = -0 * \log_2(0) - 1 * \log_2(1) = 0 \text{ bits}$$

Heart Rate = regular

$$info([4, 1]) = entropy(\frac{4}{5}, \frac{1}{5}) = -\frac{4}{5} * \log_2(\frac{4}{5}) - \frac{1}{5} * \log_2(\frac{1}{5}) = 1 \text{ bits}$$

Informationsgewinn

$$info([0, 3], [4, 1]) = \frac{3}{8} * 0 + \frac{5}{8} * 1 = \frac{5}{8} \text{ bits}$$

$$informationgain(Heart Rate) = info([4, 4]) - info([0, 3], [4, 1]) = 1 - \frac{5}{8} = \frac{3}{8} \text{ bits}$$

– **Informationsgewinn für Blood Pressure**

Blood Pressure = abnormal

$$info([0, 2]) = entropy(\frac{0}{2}, \frac{2}{2}) = -0 * \log_2(0) - 1 * \log_2(1) = 0 \text{ bits}$$

Blood Pressure = normal

$$info([4, 2]) = entropy(\frac{4}{6}, \frac{2}{6}) = -\frac{4}{6} * \log_2(\frac{4}{6}) - \frac{2}{6} * \log_2(\frac{2}{6}) = 1 \text{ bits}$$

Informationsgewinn

$$info([0, 2], [4, 2]) = \frac{2}{8} * 0 + \frac{6}{8} * 1 = \frac{3}{4} \text{ bits}$$

$$informationgain(Blood Pressure) = info([4, 4]) - info([0, 2], [4, 2]) = 1 - \frac{3}{4} = \frac{1}{4} \text{ bits}$$

b) Der Informationsgewinn ist für Heart Rate maximal

c) $A_1 =$

$\{(Irregular, Normal, Ill),$
 $(Irregular, Abnormal, Ill),$
 $(Irregular, Normal, Ill)\}$

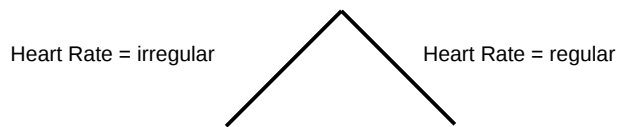
$A_2 =$

$\{(Regular, Normal, Healthy),$
 $(Regular, Normal, Healthy),$
 $(Regular, Abnormal, Ill),$
 $(Regular, Normal, Healthy),$
 $(Regular, Normal, Healthy)\}$

mit $A = A_1 \cup A_2$

Schritt 4

a)

b) Wir setzen fort mit A_1

Schritt 1

Wurzelknoten des Entscheidungsbaums B ist A_1

Schritt 2

Alle Beispiele aus A_1 gehören zur gleichen Klasse. Wir können den Algorithmus an dieser Stelle stoppen und mit A_2 fortfahren.

Schritt 1

Wurzelknoten des Entscheidungsbaums B ist A_1

Schritt 2

Nicht alle Beispiele aus A_2 gehören zur gleichen Klasse und $|A_2| \neq 0$

Schritt 3

a) – **Information vor dem Split**

$$info([4, 1]) = entropy(\frac{4}{5}, \frac{1}{5}) = -\frac{4}{5} * \log_2(\frac{4}{5}) - \frac{1}{5} * \log_2(\frac{1}{5}) = 1 \text{ bits}$$

– **Informationsgewinn für Heart Rate**

Heart Rate = irregular

$$info([0, 0]) = entropy(\frac{0}{0}, \frac{0}{0}) = -0 * \log_2(0) - 0 * \log_2(0) = 0 \text{ bits}$$

Heart Rate = regular

$$info([4, 1]) = entropy(\frac{4}{5}, \frac{1}{5}) = -\frac{4}{5} * \log_2(\frac{4}{5}) - \frac{1}{5} * \log_2(\frac{1}{5}) = 1 \text{ bits}$$

Informationsgewinn

$$info([0, 0], [4, 1]) = \frac{0}{5} * 0 + \frac{5}{5} * 1 = 1 \text{ bits}$$

$$informationgain(Heart Rate) = info([4, 1]) - info([0, 0], [4, 1]) = 1 - 1 = 0 \text{ bits}$$

– **Informationsgewinn für Blood Pressure**

Blood Pressure = abnormal

$$info([0, 1]) = entropy(\frac{0}{1}, \frac{1}{1}) = -0 * \log_2(0) - 1 * \log_2(1) = 0 \text{ bits}$$

Blood Pressure = normal

$$info([4, 0]) = entropy(\frac{4}{4}, \frac{0}{4}) = -1 * \log_2(1) - 0 * \log_2(0) = 0 \text{ bits}$$

Informationsgewinn

$$info([0, 1], [4, 0]) = \frac{1}{5} * 0 + \frac{4}{5} * 0 = 0 \text{ bits}$$

$$informationgain(Blood Pressure) = info([4, 1]) - info([0, 1], [4, 0]) = 1 - 0 = 1 \text{ bits}$$

b) Der Informationsgewinn ist für Blood Pressure maximal

c) $A_3 =$

$$\{(Regular, Abnormal, Ill)\}$$

 $A_4 =$

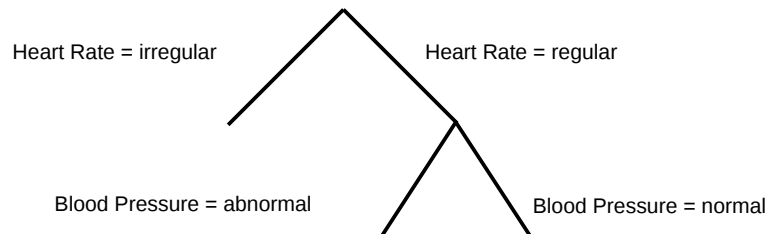
$$\{(Regular, Normal, Healthy),$$

$(Regular, Normal, Healthy),$
 $(Regular, Normal, Healthy),$
 $(Regular, Normal, Healthy)\}$

mit $A_2 = A_3 \cup A_4$

Schritt 4

a)



b) Wir setzen fort mit A_3

Schritt 1

Wurzelknoten des Entscheidungsbaums B ist A_3

Schritt 2

Alle Beispiele aus A_3 gehören zur gleichen Klasse. Wir können den Algorithmus an dieser Stelle stoppen und mit A_4 fortfahren.

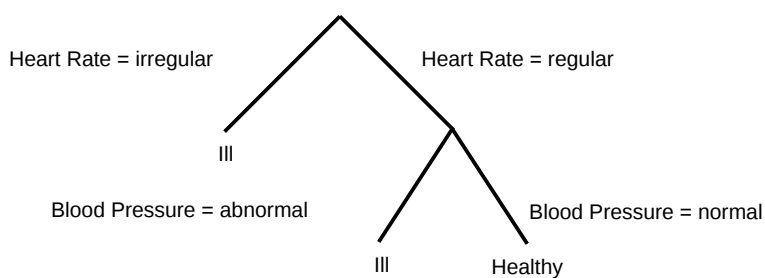
Schritt 1

Wurzelknoten des Entscheidungsbaums B ist A_4

Schritt 2

Alle Beispiele aus A_4 gehören zur gleichen Klasse. Wir können den Algorithmus an dieser Stelle stoppen.

Ausgabe



Aufgabe 3 - Toolgestützte Entscheidungsbauminduktion (22 Punkte)

(a)

```
@relation uebung06_3a
```

```
@attribute Heart_Rate {Irregular, Regular}
@attribute Blood_Pressure {Abnormal, Normal}
@attribute Klasse {Ill, Healthy}
```

```
@data
Irregular,Normal,Ill
Regular,Normal,Healthy
Irregular,Abnormal,Ill
Irregular,Normal,Ill
Regular,Normal,Healthy
Regular,Abnormal,Ill
Regular,Normal,Healthy
Regular,Normal,Healthy
```

(b)

```
=== Run information ===
```

```
Scheme:      weka.classifiers.trees.J48 -L -U -M 1
Relation:    uebung06_3a
Instances:   8
Attributes:  3
              Heart_Rate
              Blood_Pressure
              Klasse
Test mode:   evaluate on training data
```

```
=== Classifier model (full training set) ===
```

```
J48 unpruned tree
-----
```

```
Heart_Rate = Irregular: Ill (3.0)
Heart_Rate = Regular
|   Blood_Pressure = Abnormal: Ill (1.0)
|   Blood_Pressure = Normal: Healthy (4.0)
```

```
Number of Leaves   :    3
```

```
Size of the tree   :    5
```

```
Time taken to build model: 0 seconds
```

```
=== Evaluation on training set ===
=== Summary ===
```

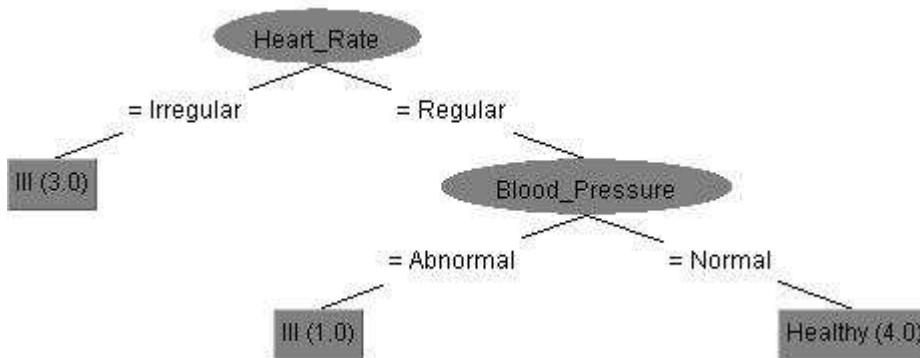
Correctly Classified Instances	8	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Relative absolute error	0	%	
Root relative squared error	0	%	
Total Number of Instances	8		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Ill
1	0	1	1	1	Healthy

=== Confusion Matrix ===

```
a b  <-- classified as
4 0 | a = Ill
0 4 | b = Healthy
```



Aufgabe 4 - Klausurvorbereitung (0 Punkte)

Erledigt!