

Maschinelles Lernen: Symbolische Ansätze



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Wintersemester 2009/2010
Musterlösung für das 4. Übungsblatt

Aufgabe 1

Gegeben sei das Golf-Spiel Datenset aus der Vorlesung.

```
@relation weather.symbolic
@attribute outlook {sunny, overcast, rainy}
@attribute temperature {hot, mild, cool}
@attribute humidity {high, normal}
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}
@data
sunny,hot,high,FALSE,no
sunny,hot,high,TRUE,no
overcast,hot,high,FALSE,yes
rainy,mild,high,FALSE,yes
rainy,cool,normal,FALSE,yes
rainy,cool,normal,TRUE,no
overcast,cool,normal,TRUE,yes
sunny,mild,high,FALSE,no
sunny,cool,normal,FALSE,yes
rainy,mild,normal,FALSE,yes
sunny,mild,normal,TRUE,yes
overcast,mild,high,TRUE,yes
overcast,hot,normal,FALSE,yes
rainy,mild,high,TRUE,no
```

Die positive Klasse sei die Klasse yes.

- a) Führen Sie eine Iteration des BATCH-FINDG Algorithmus aus der Vorlesung durch. Woran erkennen Sie, daß dieses Problem nicht mit diesem Algorithmus lösbar ist?

Lösung: Wie man an der nachfolgenden Tabelle sieht, kann der Algorithmus keine Bedingung auswählen, durch die alle positiven Beispiele abgedeckt werden. Der Algorithmus führt zu keinem Ergebnis, da er die while-Schleife nicht verlassen kann.

Attribut	Wert	+	-	Precision
Outlook	overcast	4	0	1.00
	rain	3	2	0.60
	sunny	2	3	0.40
Temperature	cool	3	1	0.75
	hot	2	2	0.50
	mild	4	2	0.67
Humidity	high	3	4	0.43
	normal	6	1	0.86
Windy	FALSE	6	2	0.75
	TRUE	3	3	0.50

b) Wenden Sie den Separate-And-Conquer-Algorithmus (siehe Folie 9) auf die Beispiele an. Konstruieren Sie die einzelnen Regeln mittels Top-Down Hill-Climbing (Siehe Lernen von Regel-Mengen, Folie 8 bzw. 14):

- mit dem Maß Precision

Lösung:

- Wir bestimmen die ersten Regel (Conquer-Schritt)

Attribut	Wert	+	–	Precision
Outlook	overcast	4	0	1.00
	rain	3	2	0.60
	sunny	2	3	0.40
Temperature	cool	3	1	0.75
	hot	2	2	0.50
	mild	4	2	0.67
Humidity	high	3	4	0.43
	normal	6	1	0.86
Windy	FALSE	6	2	0.75
	TRUE	3	3	0.50

Wir wählen Outlook = overcast (Precision = 1.0 → Regel fertig, da kein negatives Beispiel abgedeckt wird - siehe Top-Down Hill-Climbing)

Outlook = overcast → yes

- Update: Entfernen der abgedeckten Beispiele (Separate-Schritt)

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
rainy	mild	high	TRUE	no

- Bestimmen der zweiten Regel

Attribut	Wert	+	–	Precision
Outlook	rainy	3	2	0.60
	sunny	2	3	0.40
Temperature	cool	2	1	0.67
	hot	0	2	0.00
	mild	3	2	0.60
Humidity	high	1	4	0.20
	normal	4	1	0.80
Windy	FALSE	4	2	0.67
	TRUE	1	3	0.25

Wir wählen Humidity = normal und bestimmen abgedeckte Beispiele. Da diese Regel noch negative Beispiele abdeckt, müssen wir weiter verfeinern.

Outlook	Temperature	Humidity	Windy	Class
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes

Attribut	Wert	+	–	Precision
Outlook	rainy	2	1	0.67
	sunny	2	0	1.00
Temperature	cool	2	1	0.67
	mild	2	0	1.00
Windy	FALSE	3	0	1.00
	TRUE	1	1	0.50

Wir wählen Windy = FALSE (Precision = 1.0).

Humidity = normal \wedge Windy = FALSE \rightarrow yes

– Update: Entfernen der abgedeckten Beispiele

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
rainy	mild	high	FALSE	yes
rainy	cool	normal	TRUE	no
sunny	mild	high	FALSE	no
sunny	mild	normal	TRUE	yes
rainy	mild	high	TRUE	no

– Bestimmen der dritten Regel

Attribut	Wert	+	–	Precision
Outlook	rainy	1	2	0.33
	sunny	1	3	0.25
Temperature	cool	0	1	0.00
	hot	0	2	0.00
	mild	2	2	0.50
Humidity	high	1	4	0.20
	normal	1	1	0.50
Windy	FALSE	1	2	0.33
	TRUE	1	3	0.25

Wir wählen Temperature = mild und bestimmen die abgedeckten Beispiele.

Outlook	Temperature	Humidity	Windy	Class
rainy	mild	high	FALSE	yes
sunny	mild	high	FALSE	no
sunny	mild	normal	TRUE	yes
rainy	mild	high	TRUE	no

Attribut	Wert	+	–	Precision
Outlook	rainy	1	1	0.50
	sunny	1	1	0.50
Humidity	high	1	2	0.50
	normal	1	0	1.00
Windy	FALSE	1	1	0.50
	TRUE	1	1	0.50

Wir wählen Humidity = normal (Precision = 1.0).

Humidity = normal \wedge Temperature = mild \rightarrow yes

– Update: Entfernen der abgedeckten Beispiele

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
rainy	mild	high	FALSE	yes
rainy	cool	normal	TRUE	no
sunny	mild	high	FALSE	no
rainy	mild	high	TRUE	no

– Bestimmen der vierten Regel

Attribut	Wert	+	–	Precision
Outlook	rainy	1	2	0.33
	sunny	0	3	0.00
Temperature	cool	0	1	0.00
	hot	0	2	0.00
	mild	1	2	0.33
Humidity	high	1	4	0.20
	normal	0	1	0.00
Windy	FALSE	1	2	0.3
	TRUE	0	3	0.00

Wir wählen Outlook = rainy und bestimmen die abgedeckten Beispiele.

Outlook	Temperature	Humidity	Windy	Class
rainy	mild	high	FALSE	yes
rainy	cool	normal	TRUE	no
rainy	mild	high	TRUE	no

Attribut	Wert	+	-	Precision
Temperature	cool	0	1	0.00
	mild	1	1	0.50
Humidity	high	1	1	0.50
	normal	0	1	0.00
Windy	FALSE	1	0	1.00
	TRUE	0	2	0.00

Wir wählen Windy = FALSE (Precision = 1.0).

Outlook = rainy \wedge Windy = FALSE \rightarrow yes

– Update: Entfernen der abgedeckten Beispiele

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
rainy	cool	normal	TRUE	no
sunny	mild	high	FALSE	no
rainy	mild	high	TRUE	no

– Alle positiven Beispiele sind abgedeckt.

– Regelmenge:

- * Outlook = overcast \rightarrow yes
- * Humidity = normal \wedge Windy = FALSE \rightarrow yes
- * Humidity = normal \wedge Temperature = mild \rightarrow yes
- * Outlook = rainy \wedge Windy = FALSE \rightarrow yes

- mit dem Maß Accuracy, wobei die Regel solange verfeinert wird, bis keine negativen Beispiele mehr abgedeckt werden. Anschließend wählen Sie aus der hierbei entstehenden Menge von Verfeinerungen diejenige Regel mit der höchsten Bewertung aus.

Lösung:

– Bestimmen der ersten Regel

Attribut	Wert	+	-	Accuracy
Outlook	overcast	4	0	4
	rainy	3	2	1
	sunny	2	3	-1
Temperature	cool	3	1	2
	hot	2	2	0
	mild	4	2	2
Humidity	high	3	4	-1
	normal	6	1	5
Windy	FALSE	6	2	4
	TRUE	3	3	0

Wir wählen Humidity = normal und bestimmen die abgedeckten Beispiele.

Outlook	Temperature	Humidity	Windy	Class
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
overcast	cool	normal	TRUE	yes
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	hot	normal	FALSE	yes

Attribut	Wert	+	-	Accuracy
Outlook	overcast	2	0	2
	rainy	2	1	1
	sunny	2	0	2
Temperature	cool	3	1	2
	hot	1	0	1
	mild	2	0	2
Windy	FALSE	4	0	4
	TRUE	2	1	1

Wir wählen Windy = FALSE (keine negativen Beispiele abgedeckt).

Mögliche Regeln:

- * Humidity = normal → yes (Accuracy = 5)
- * Humidity = normal ∧ Windy = FALSE → yes (Accuracy = 4)

Die erste Regel wird ausgewählt.

- Update: Entfernen der abgedeckten Beispiele

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
overcast	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
sunny	mild	high	FALSE	no
overcast	mild	high	TRUE	yes
rainy	mild	high	TRUE	no

- Bestimmen der zweiten Regel

Attribut	Wert	+	-	Accuracy
Outlook	overcast	2	0	2
	rainy	1	1	0
	sunny	0	3	-3
Temperature	hot	1	2	-1
	mild	2	2	0
Humidity	high	3	4	-1
Windy	FALSE	2	2	0
	TRUE	1	2	-1

Wir wählen Outlook = overcast (keine negativen Beispiele abgedeckt).

Outlook = overcast → yes

- Update: Entfernen der abgedeckten Beispiele

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
rainy	mild	high	FALSE	yes
sunny	mild	high	FALSE	no
rainy	mild	high	TRUE	no

- Bestimmen der dritten Regel

Attribut	Wert	+	-	Accuracy
Outlook	rainy	1	1	0
	sunny	0	3	-3
Temperature	hot	0	2	-1
	mild	1	2	-1
Humidity	high	1	4	-3
Windy	FALSE	1	2	-1
	TRUE	0	2	-2

Wir wählen Outlook = rainy und bestimmen die abgedeckten Beispiele.

Outlook	Temperature	Humidity	Windy	Class
rainy	mild	high	FALSE	yes
rainy	mild	high	TRUE	no

Attribut	Wert	+	-	Accuracy
Temperature	mild	1	1	0.50
Humidity	high	1	1	0.50
Windy	FALSE	1	0	1.00
	TRUE	0	1	0.00

Wir wählen Windy = FALSE (keine negativen Beispiele abgedeckt).

Mögliche Regeln:

- * Outlook = rainy → yes (Accuracy = 0)
- * **Outlook = rainy ∧ Windy = FALSE → yes** (Accuracy = 1)

Die zweite Regel wird ausgewählt.

– Alle positiven Beispiele sind abgedeckt. Damit erhalten wir folgende Regelmenge:

- * Humidity = normal → yes
- * Outlook = overcast → yes
- * Outlook = rainy ∧ Windy = FALSE → yes

Diskutieren Sie die Ergebnisse. Welche Regelmenge sieht am besten aus?

Das Maß Precision hat mehr Regeln (4) gefunden als Accuracy (3) und diese haben mehr Bedingungen (insgesamt 7 bei Precision und 4 bei Accuracy). Daher tendiert Precision eher dazu speziellere Regeln (also Regeln die weniger Beispiele abdecken) zu finden, wobei diese dann auch genauer sind (also eher an die Trainingsmenge angepasst). Mit dem Maß Accuracy werden auch negative Beispiele (in der Aufgabe 1 Beispiel) abgedeckt. Daher findet dieses Maß eher generelle Regeln (also welche deren Abdeckung höher ist), die allerdings auch nicht so genau sind (also weniger an die Trainingsmenge angepasst sind). Für eine weiterführende Erklärung siehe Aufgabe 3.2.

c) Wiederholen Sie 1.2, indem sie die Rolle der Klassen vertauschen (also die positive Klasse sei jetzt no).

Lösung: Die Berechnung erfolgt analog zur Aufgabe 1.2. Die resultierenden Regelmengen sehen sowohl für Precision als auch für Accuracy wie folgt aus (wobei die einzelnen Bedingungen in einer anderen Reihenfolge gefunden werden):

- Outlook = sunny ∧ Humidity = high → yes
- Outlook = rainy ∧ Windy = TRUE → yes

d) Eine Bottom-Up (also Specific-To-General) Lern-Strategie zur Batch-Induktion einzelner Regeln könnte so aussehen, daß ein positives Beispiel zufällig ausgewählt wird, und dann sukzessive generalisiert wird. Simulieren Sie diese Strategie an diesen Trainings-Beispielen, wobei aus Gründen der Vergleichbarkeit bitte als erstes "zufällig" ausgewähltes Beispiel das fünfte Beispiel verwenden. In allen weiteren Iterationen wählen Sie bitte das erste positive Beispiel der Trainingsmenge.

Lösung:

- Precision
 - Wir wählen das fünfte Beispiel (rainy,cool,normal,FALSE,yes) aus und testen alle Möglichkeiten einen Vergleich (Attribut = Attributwert) zu entfernen bzw. die Regel (Beispiel 5) zu generalisieren.

Outlook	Temperature	Humidity	Windy	p	n	Precision
?	cool	normal	FALSE	2	0	1.00
rainy	?	normal	FALSE	2	0	1.00
rainy	cool	?	FALSE	1	0	1.00
rainy	cool	normal	?	1	1	0.50

Wir entfernen die Bedingung Outlook = rainy und generalisieren weiter (wir wählen bei gleicher Precision die oberste Regel, sofern beide Regeln gleich viele positive Beispiele abdecken. Ansonsten wird die Regel, die mehr positive Beispiele abdeckt ausgewählt).

Outlook	Temperature	Humidity	Windy	p	n	Precision
?	?	normal	FALSE	4	0	1.00
?	cool	?	FALSE	2	0	1.00
?	cool	normal	?	3	1	0.75

Wir entfernen die Bedingung Temperature = cool und generalisieren weiter.

Outlook	Temperature	Humidity	Windy	p	n	Precision
?	?	?	FALSE	6	2	0.75
?	?	normal	?	6	1	0.86

Kein Vergleich kann entfernt werden, da nun negative Beispiele abgedeckt werden. Also erhalten wir die folgende Regel:

Humidity = normal \wedge Windy = FALSE \rightarrow yes

- Update: Entfernen der abgedeckten Beispiele

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
overcast	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
rainy	cool	normal	TRUE	no
overcast	cool	normal	TRUE	yes
sunny	mild	high	FALSE	no
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
rainy	mild	high	TRUE	no

- Wir wählen das dritte Beispiel (overcast,hot,high,FALSE,yes) aus:

Outlook	Temperature	Humidity	Windy	p	n	Precision
?	hot	high	FALSE	1	1	0.50
overcast	?	high	FALSE	1	0	1.00
overcast	hot	?	FALSE	1	0	1.00
overcast	hot	high	?	1	0	1.00

Wir entfernen die Bedingung Temperature = hot und generalisieren weiter.

Outlook	Temperature	Humidity	Windy	p	n	Precision
?	?	high	FALSE	2	2	0.50
overcast	?	?	FALSE	1	0	1.00
overcast	?	high	?	2	0	1.00

Wir entfernen die Bedingung Windy = FALSE und generalisieren weiter.

Outlook	Temperature	Humidity	Windy	p	n	Precision
?	?	high	?	3	4	0.43
overcast	?	?	?	3	0	1.00

Wir entfernen die letzte entfernbare Bedingung Humidity = high.

Outlook = overcast \rightarrow yes

- Update: Entfernen der abgedeckten Beispiele

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
rainy	mild	high	FALSE	yes
rainy	cool	normal	TRUE	no
sunny	mild	high	FALSE	no
sunny	mild	normal	TRUE	yes
rainy	mild	high	TRUE	no

- Wir wählen das dritte Beispiel (rainy,mild,high,FALSE,yes) aus:

Outlook	Temperature	Humidity	Windy	p	n	Precision
?	mild	high	FALSE	1	1	0.50
rainy	?	high	FALSE	1	0	1.00
rainy	mild	?	FALSE	1	0	1.00
rainy	mild	high	?	1	0	1.00

Wir entfernen die Bedingung Temperature = mild und generalisieren weiter.

Outlook	Temperature	Humidity	Windy	p	n	Precision
?	?	high	FALSE	1	2	0.33
rainy	?	?	FALSE	1	0	1.00
rainy	?	high	?	1	1	0.50

Wir entfernen die Bedingung Humidity = high und generalisieren weiter.

Outlook	Temperature	Humidity	Windy	p	n	Precision
?	?	?	FALSE	1	2	0.33
rainy	?	?	?	1	2	0.33

Kein Vergleich kann entfernt werden. Also erhalten wir die folgende Regel:

Outlook = rainy \wedge Windy = FALSE \rightarrow yes

– Update: Entfernen der abgedeckten Beispiele

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
rainy	cool	normal	TRUE	no
sunny	mild	high	FALSE	no
sunny	mild	normal	TRUE	yes
rainy	mild	high	TRUE	no

– Wir wählen das fünfte Beispiel (sunny,mild,normal,TRUE,yes) aus:

Outlook	Temperature	Humidity	Windy	p	n	Precision
?	mild	normal	TRUE	1	0	1.00
sunny	?	normal	TRUE	1	0	1.00
sunny	mild	?	TRUE	1	0	1.00
sunny	mild	normal	?	1	0	1.00

Wir entfernen die Bedingung Outlook = sunny und generalisieren weiter.

Outlook	Temperature	Humidity	Windy	p	n	Precision
?	?	normal	TRUE	1	1	0.50
?	mild	?	TRUE	1	1	0.50
?	mild	normal	?	1	0	1.00

Wir entfernen die Bedingung Windy = TRUE und generalisieren weiter.

Outlook	Temperature	Humidity	Windy	p	n	Precision
?	?	normal	?	1	1	0.50
?	mild	?	?	1	2	0.33

Kein Vergleich kann entfernt werden. Also erhalten wir die folgende Regel:

Temperature = mild \wedge Humidity = normal \rightarrow yes

– Alle positiven Beispiele sind abgedeckt. Wir erhalten also die folgende Regelmenge:

- * Humidity = normal \wedge Windy = FALSE \rightarrow yes
- * Outlook = overcast \rightarrow yes
- * Outlook = rainy \wedge Windy = FALSE \rightarrow yes
- * Temperature = mild \wedge Humidity = normal \rightarrow yes

• Accuracy

– Wir beginnen wieder mit dem fünften Beispiel:

(rainy,cool,normal,FALSE,yes) Abdeckung: p=1, n=0, Accuracy = 1

Outlook	Temperature	Humidity	Windy	p	n	Accuracy
?	cool	normal	FALSE	2	0	2
rainy	?	normal	FALSE	2	0	2
rainy	cool	?	FALSE	1	0	1
rainy	cool	normal	?	1	1	0

Wir entfernen die Bedingung Outlook = rainy und generalisieren weiter.

Outlook	Temperature	Humidity	Windy	p	n	Accuracy
?	?	normal	FALSE	4	0	4
?	cool	?	FALSE	2	0	2
?	cool	normal	?	3	1	2

Wir entfernen die Bedingung Temperature = cool und generalisieren weiter.

Outlook	Temperature	Humidity	Windy	p	n	Accuracy
?	?	?	FALSE	6	2	4
?	?	normal	?	6	1	5

Wir entfernen die Bedingung Windy = FALSE. Kein weiterer Vergleich kann entfernt werden, da die generellste Regel nicht betrachtet wird. Also erhalten wir die folgenden, möglichen Regeln:

- * Temperature = cool \wedge Humidity = normal \wedge Windy = FALSE \rightarrow yes (Accuracy = 2)
- * Humidity = normal \wedge Windy = FALSE \rightarrow yes (Accuracy = 4)
- * **Humidity = normal \rightarrow yes (Accuracy = 5)**

Die dritte Regel wird ausgewählt.

- Update: Entfernen der abgedeckten Beispiele

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
overcast	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
sunny	mild	high	FALSE	no
overcast	mild	high	TRUE	yes
rainy	mild	high	TRUE	no

- Wir wählen das dritte Beispiel (overcast,hot,high,FALSE,yes) aus:

Outlook	Temperature	Humidity	Windy	p	n	Accuracy
?	hot	high	FALSE	1	1	2
overcast	?	high	FALSE	1	0	1
overcast	hot	?	FALSE	1	0	1
overcast	hot	high	?	1	0	1

Wir entfernen die Bedingung Temperature = hot und generalisieren weiter.

Outlook	Temperature	Humidity	Windy	p	n	Accuracy
?	?	high	FALSE	2	2	0
overcast	?	?	FALSE	1	0	1
overcast	?	high	?	2	0	2

Wir entfernen die Bedingung Windy = FALSE und generalisieren weiter.

Outlook	Temperature	Humidity	Windy	p	n	Accuracy
overcast	?	high	?	3	4	-1
overcast	?	?	?	2	0	2

Wir entfernen die Bedingung Humidity = high. Kein weiterer Vergleich kann entfernt werden. Also erhalten wir die folgenden, möglichen Regeln:

- * Outlook = overcast \wedge Humidity = high \wedge Windy = FALSE \rightarrow yes (Accuracy = 1)
- * Outlook = overcast \wedge Humidity = high \rightarrow yes (Accuracy = 2)
- * **Outlook = overcast \rightarrow yes (Accuracy = 2)**

Die dritte Regel wird ausgewählt.

- Update: Entfernen der abgedeckten Beispiele

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
rainy	mild	high	FALSE	yes
sunny	mild	high	FALSE	no
rainy	mild	high	TRUE	no

- Wir wählen das dritte Beispiel (rainy,mild,high,FALSE,yes) aus:

Outlook	Temperature	Humidity	Windy	p	n	Accuracy
?	mild	high	FALSE	1	1	0
rainy	?	high	FALSE	1	0	1
rainy	mild	?	FALSE	1	0	1
rainy	mild	high	?	1	1	0

Wir entfernen die Bedingung Temperature = mild und generalisieren weiter.

Outlook	Temperature	Humidity	Windy	p	n	Accuracy
?	?	high	FALSE	1	2	-1
rainy	?	?	FALSE	1	0	1
rainy	?	high	?	1	1	0

Wir entfernen die Bedingung Humidity = high und generalisieren weiter.

Outlook	Temperature	Humidity	Windy	p	n	Accuracy
?	?	?	FALSE	1	2	-1
rainy	?	?	?	1	0	1

Wir entfernen die Bedingung Windy = FALSE. Kein weiterer Vergleich kann entfernt werden. Also erhalten wir die folgenden, möglichen Regeln:

- * Outlook = rainy \wedge Humidity = high \wedge Windy = FALSE \rightarrow yes (Accuracy = 1)
- * Outlook = rainy \wedge Windy = FALSE \rightarrow yes (Accuracy = 1)
- * Outlook = rainy \rightarrow yes (Accuracy = 0)

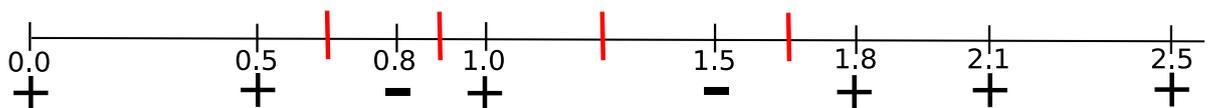
Wir wählen die zweite Regel aus.

- Alle positiven Beispiele sind abgedeckt damit erhalten wir die folgende Regelmenge:
 - * Humidity = normal \rightarrow yes
 - * Outlook = overcast \rightarrow yes
 - * Outlook = rainy \wedge Windy = FALSE \rightarrow yes (Accuracy = 1)

- e) Eine alternative Strategie wäre, alle Beispiele in Regeln zu verwandeln, zwei beliebige Regeln auszuwählen, das lgg dieser Beispiele zu finden, und dann die beiden alten Regeln durch diese neue zu ersetzen. Wieso wird diese Strategie i.A. nicht funktionieren? Wie könnte man sie verbessern (z.B. durch Auswahl der Regeln, Abbruchbedingungen, etc.)?

Lösung: Das lgg (die least general generalization, also die am wenigsten generelle Generalisierung) von zwei Regeln, die jeweils ein Beispiel repräsentieren, wird gebildet, indem man alle unterschiedlichen Attributwerte durch Fragezeichen ersetzt. Wählt man nun zufällig 2 Beispiele und bildet das lgg, so kann es passieren, dass man die allgemeinste Regel (?, ?, ?, ?) erhält (bei Verwendung von Beispiel 6 und 8). Da diese aber alle negativen Beispiele abdeckt, ist die Strategie i.a. nicht gut. Besser wäre, für alle möglichen Regeln (nicht nur für zufällig gewählte) das lgg zu bilden (Verbesserung der Auswahl) und sich dann die herauszusuchen, die ein gewisses Qualitätskriterium erfüllen (also einen hohen Heuristikwert erhält), was der Abbruchbedingung entspricht.

- f) Überlegen Sie sich, wie dieser Algorithmus mit numerischen bzw. hierarchischen Attributen umgehen könnte.



Lösung: Um numerische Daten verarbeiten zu können, haben wir zwei Vergleichsmöglichkeiten. Entweder vergleichen wir den numerischen Attributwert mit einem konstanten, beim Lernen bestimmten Wert (Attributwert $<$, \leq Konstante bzw. Attributwert $>$, \geq Konstante) oder wir teilen die entsprechende Zahlenmenge (natürliche/reelle/etc. Zahlen) in mehrere Intervalle auf und testen dann, ob ein Attributwert in diesem Intervall enthalten ist (Attributwert \in Intervall). Die Konstante bzw. die beiden Grenzen eines Intervalls liegen, wie wir noch im Laufe der Vorlesung sehen werden, im allgemeinen an einem Übergang (rote Balken in Skizze) von + nach - bzw. umgekehrt.

Bei hierarchischen Daten würde zusätzlich zu den "normalen" Attributen als Spezialisierung die Möglichkeit hinzugenommen werden in der Hierarchie abwärts zu gehen. Als Generalisierung würde man die Möglichkeit erhalten in der Hierarchie nach oben zu wandern.

Aufgabe 2

Gegeben sei der folgende Datensatz.

```
@relation x
@attribute a1 {0,1}
@attribute a2 {0,1}
@attribute a3 {0,1}
@attribute a4 {0,1}
@attribute x {yes, no}
@data
1,0,0,0,yes
1,1,0,1,yes
0,0,1,1,no
1,0,0,1,no
1,1,1,0,no
0,0,1,0,yes
0,0,0,1,no
1,1,0,0,no
0,1,1,1,yes
1,0,1,0,yes
0,1,0,1,yes
0,1,1,0,no
```

- a) Versuchen Sie eine möglichst einfache Regelmenge zu finden oder zu lernen, die diesen Datensatz erklärt.

Lösung:

R_1 :

* $a_2 = 0 \wedge a_4 = 0 \rightarrow \text{yes}$

* $a_2 = 1 \wedge a_4 = 1 \rightarrow \text{yes}$

R_2 :

* $a_2 = 0 \wedge a_4 = 1 \rightarrow \text{no}$

* $a_2 = 1 \wedge a_4 = 0 \rightarrow \text{no}$

- b) Warum hat der Separate-and-Conguer Algorithmus (unabhängig von der eingesetzten Heuristik) Probleme beim Lernen dieses Datensatz?

Lösung:

Alle möglichen Bedingungen haben exakt die gleiche Abdeckung von (3,3). Aus diesem Grund würde der SE-Co-Algorithmus eine der Bedingungen zufällig wählen. Wählt er die Richtige, so funktioniert der Algorithmus. Wird hingegen die Falsche ausgewählt, versagt der Algorithmus, da SE-Co-Algorithmen nicht in der Lage sind eine im vorherigen Schritt getroffene Entscheidung rückgängig zu machen. Daher gibt es Trainingsmengen, die mit SE-Co-Algorithmen nicht gelernt werden können (außer man wählt der Reihe nach jede Bedingung aus, was im ursprünglichen Algorithmus nicht vorgesehen ist).

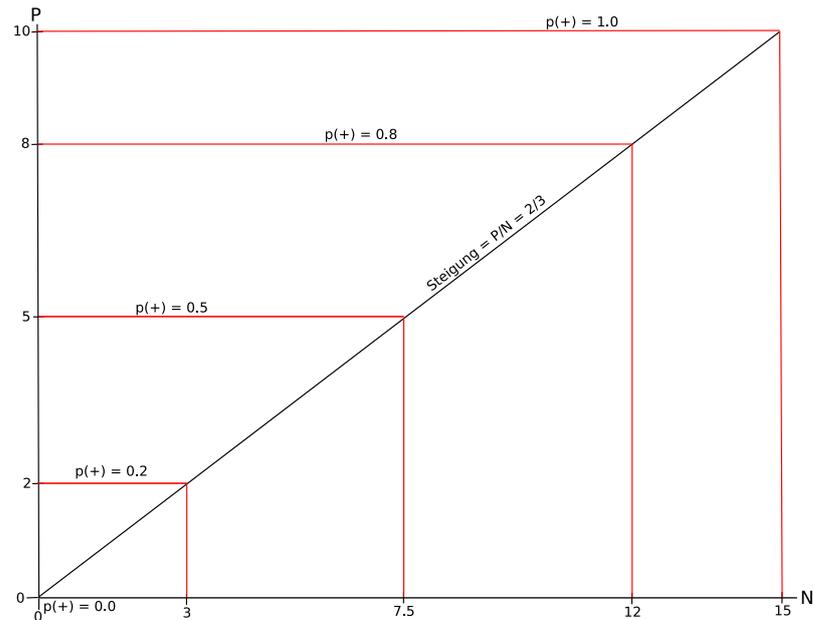
Aufgabe 3

- a) Gegeben seien Klassifizierer, die mit der Wahrscheinlichkeit p_+ für ein Beispiel unabhängig von seinen konkreten Attribut-Werten die Klasse + vorhersagt. Entsprechend wird mit der Wahrscheinlichkeit $1 - p_+$ für ein Beispiel die Klasse - vorhergesagt. Wo im Coverage Space liegen diese Klassifizierer für verschiedene Wahrscheinlichkeiten von p_+ (z.B. 0,2, 0,5, 0,8).

Lösung: Insgesamt gibt es $E = P + N$ Beispiele. Man kann nun die Wahrscheinlichkeit p_+ frei wählen. Da mit p_+ die positive Klasse vorhergesagt wird, sagt der Klassifizierer $p_+ \cdot E$ Beispiele als positiv und $p_- \cdot E$ Beispiele als negativ vorher, wobei $p_- = 1 - p_+$ gilt. Schaut man sich nun unter den als positiv vorhergesagten Beispielen diejenigen an, die tatsächlich positiv sind, so erhält man $p_+ \cdot P$ positive und $p_+ \cdot N$ negative.

Beispiel: Wähle $p_+ = 0.8$ und nehme an, dass $P = 10$ und $N = 15$ ist. $\Rightarrow E = 10 + 15 = 25$

Nun gilt: von der Theorie abgedeckte Beispiele (Beispiele, die als positiv klassifiziert sind): $p_+ \cdot E = 0.8 \cdot 25 = 20$; davon tatsächlich positiv: $p_+ \cdot P = 0.8 \cdot 10 = 8$; davon negativ: $p_+ \cdot N = 0.8 \cdot 15 = 12$. In der Grafik sind für unterschiedliche Werte von p_+ jeweils die resultierenden Linien im Coverage Space gezeichnet worden.



Der wählbare Parameter p_+ gibt also an, wie weit man sich vom Ursprung entfernt auf der Diagonalen befindet. Es ist zu beachten, dass auf alle Klassifizierer, die auf der Diagonalen liegen, noch nichts gelernt haben. Erst wenn der Klassifizierer eine von der Apriori-Verteilung $\frac{p}{p+N}$ unterschiedliche Verteilung der Beispiele realisiert hat, ist etwas gelernt worden.

- b) Overfitting aufgrund von fehlerhaften Trainings-Beispielen äußert sich oft, indem Regeln mit geringer Coverage gelernt werden. Identifizieren Sie den für Overfitting ausschlaggebenden Bereich im Coverage Space und überlegen Sie sich die Eigenschaften der in der Vorlesung besprochenen Maße bezüglich Overfitting. Z.B. welches Maß neigt eher zu Overfitting, Precision oder Accuracy?

Lösung: Der für Overfitting ausschlaggebende Bereich im Coverage Space ist eben genau der Bereich, wo eine geringe Abdeckung vorliegt. In der obigen Grafik wäre das in etwa das untere rote Rechteck. Regeln mit niedriger Coverage decken nur wenige positive Beispiele ab und kommen häufig dann vor, wenn die Trainingsbeispiele so einzigartig sind, dass sie nur von einzelnen Regeln abgedeckt werden können.

Da man in einer Trainingsmenge jedes einzelne Beispiel mit genau einer Regel abdecken kann und mit dem Maß Precision jeweils für die Regeln die höchste Bewertung erhält (da es bei Precision = $\frac{p}{p+n}$ egal ist, wie viele positive Beispiele abgedeckt sind, Hauptsache es ist kein negatives Beispiel abgedeckt), neigt dieses Maß stark zu Overfitting. Anders verhält es sich hingegen bei dem Maß Accuracy. Hier wird darauf fokussiert möglichst viele positive Beispiele abzudecken, auch wenn man gleichzeitig ein paar wenige negative mit abdeckt.

Beispiel: Man hat zwei Regeln R_1 deckt 100 positive und 1 negatives Beispiel ab und R_2 deckt 1 positives und 0 negative ab.

R_1

Precision: $\frac{p}{p+n} = \frac{100}{101} \approx 0.99$

Accuracy: $p - n = 100 - 1 = 99$

R_2

Precision: $\frac{1}{1} = 1$

Accuracy: $1 - 0 = 1$

Da die erste Regel 100 positive und nur ein negatives Beispiel abdeckt, ist diese viel besser und auch viel genereller und würde also nicht für Overfitting verantwortlich sein.