Maschinelles Lernen: Symbolische Ansätze

Musterlösung für das 7. Übungsblatt

Aufgabe 1

Gegeben sei folgende Beispielmenge:

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	?	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No
D15	Sunny	Mild	Normal	Weak	No

Erzeugen Sie einen Entscheidungsbaum mittels des Verfahrens ID3 (TDIDT mit Maß Gain). Berechnen Sie den Gain des Attributes Outlook, indem Sie Anteile des Beispiels in die Teilbäume propagieren (siehe letzter Punkt auf Folie 34).

Anmerkung: ? steht hier für einen unbekannten/fehlenden Attributwert.

Lösung: Das Beispiel *D1* wird entsprechend der relativen Häufigkeit der Attributwerte *Overcast*, *Rain* und *Sunny* zu den den absoluten Häufigkeiten dieser Attributwerte hinzugerechnet. *D1* zählt also für

- Overcast als ein $\frac{4}{14} = \frac{2}{7}$ Beispiel, für
- Rain als ein $\frac{5}{14}$ Beispiel und für

- Sunny als ein $\frac{5}{14}$ Beispiel.

Also erhalten wir die folgenden absoluten Häufigkeiten:

Test	Werte	p	n	Entropie
Outlook	Overcast	4	$^2/_7$	0,353
	Rain	3	2 $^5/_{14}$	0,990
	Sunny	2	3 $^5/_{14}$	0,953
Temperature	Cool	3	1	0,811
	Hot	2	2	1,000
	Mild	4	3	0,985
Humidity	High	3	4	0,985
	Normal	6	2	0,811
Wind	Strong	3	3	1,000
	Weak	6	3	0,918

Für die vier möglichen Tests erhalten wir die folgenden Gains:

- $Gain(S, Outlook) =$ **0,176**
- $Gain(S, Temperature) = 0,028$
- $Gain(S, Humidity) = 0,078$
- $Gain(S, Wind) = 0,02$

Wir entscheiden uns für den Test $Outlook$. Wir teilen die Beispiele entsprechend auf und erhalten drei Beispielmengen $S_{Overcast}$, S_{Rain} und S_{Sunny}, die das Beispiel zu $\frac{4}{14}$, $\frac{5}{14}$ bzw. $\frac{5}{14}$ beinhalten. Betrachten wir $S_{Overcast}$, erhalten wir die folgenden absoluten Häufigkeiten.

Test	Werte	p	n	Entropie
Temperature	Cool	1	0	0,000
	Hot	2	$^2/_7$	0,544
	Mild	1	0	0,000
Humidity	High	2	$^2/_7$	0,544
	Normal	2	0	0,000
Wind	Strong	2	0	0,000
	Weak	2	$^2/_7$	0,544

Damit ergeben sich die folgende Gains:

- $Gain(S_{Overcast}, Temperature) =$ **0,063**
- $Gain(S_{Overcast}, Humidity) = 0,063$
- $Gain(S_{Overcast}, Wind) = 0,063$

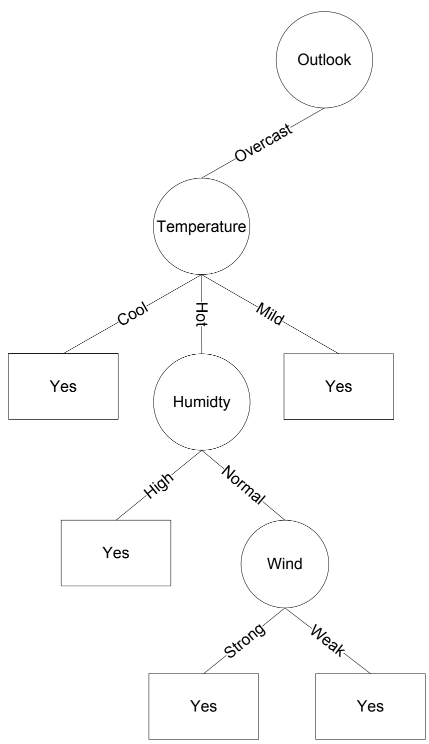
Wir entscheiden uns für den Test *Temperature* und müssen nur die Beispielmenge $S_{Overcast,Hot}$ betrachten.

Test	Werte	p	n	Entropie
Humidity	High	1	$^2/_7$	0,430
	Normal	1	0	0,000
Wind	Strong	0	0	1,000
	Weak	2	$^2/_7$	0,544

Mit diesen Werten ergeben sich die folgenden Gains:

- $Gain(S_{Overcast,Hot}, Humidity) =$ **0,114**
- $Gain(S_{Overcast,Hot}, Wind) = 0$

Wir entscheiden uns für den Test *Humidity* und betrachten die noch zu trennenden Beispiele *D1* (mit Outlook = Overcast) und *D3*. Diese können durch keinen Test getrennt werden, da sie sich in keinem Attribut unterscheiden. Da sie sich aber in ihrer Klasse unterscheiden, müssen wir uns für eine der beiden Klassen entscheiden. Wir wählen die Klasse von *D3*, da wir uns sicher sind, daß dieses Beispiel auch wirklich zu dieser Klasse gehört. Damit erhalten wir den folgenden Teilbaum (beim Test von "Humidity" sind "High" und "Normal" vertauscht):



Wie man sieht ist dieser Teilbaum semantisch äquivalent zu einem einzelnen Blatt, das *yes* vorhersagt. Aus diesem Grund ersetzen wir den Teilbaum durch ein derartiges Blatt.

Wenden wir uns der Beispielmenge S_{Rain} zu.

Test	Werte	p	n	Entropie
Temperature	Cool	1	1	1,000
	Hot	0	$^5/_{14}$	0,000
	Mild	2	1	0,918
Humidity	High	1	1 $^5/_{14}$	0,983
	Normal	2	1	0,918
Wind	Strong	0	2	0,000
	Weak	3	$^5/_{14}$	0,489

Wir erhalten folgende Gains:

- $Gain(S_{Rain}, Temperature) = 0,102$
- $Gain(S_{Rain}, Humidity) = 0,043$
- $Gain(S_{Rain}, Wind) =$ **0,683**

Wir wählen den Test *Wind* und müssen nur noch $S_{Rain,Wind}$ betrachten.

Test	Werte	p	n	Entropie
Temperature	Cool	1	0	0,000
	Hot	0	$^5/_{14}$	0,000
	Mild	2	0	0,000
Humidity	High	2	$^5/_{14}$	0,614
	Normal	1	0	0,000

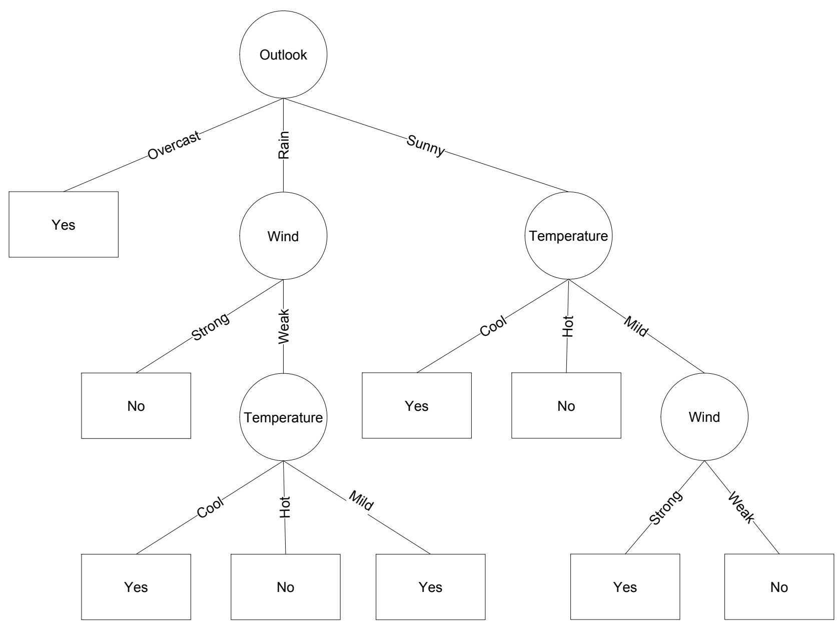
Damit erhalten wir die folgenden Gains:

- $Gain(S_{Rain,Wind}, Temperature) =$ **0,489**
- $Gain(S_{Rain,Wind}, Humidity) = 0,219$

Damit ist dieser Teilbaum (*Outlook* = *Rain*). Die Berechnungen für *Outlook* = *Sunny* erfolgen analog, deshalb geben wir nur noch die Gains (ohne Zwischenrechnung- bzw. werte) an.

- $Gain(S_{Sunny}, Temperature) =$ **0,439**
- $Gain(S_{Sunny}, Humidity) = 0,439$
- $Gain(S_{Sunny}, Wind) = 0,029$

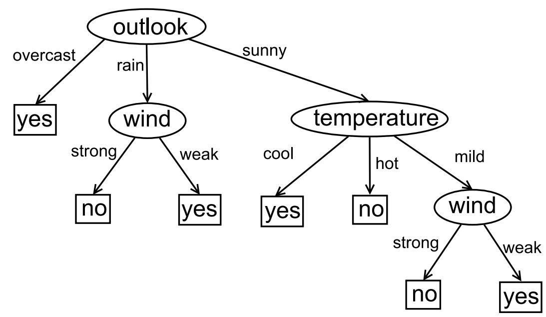
Wir wählen den Test *Temperature* und untersuchen nur noch $S_{Sunny,Mild}$, da diese Beispielmenge als einzige aus positiven und negativen Beispielen besteht. Das Beispiel *D1* befindet sich aber nicht in diesem Teilbaum, deshalb erfolgt die Berechnung analog zur Aufgabe 2a) der vorherigen Übung. Damit sieht der endgültige Entscheidungsbaum wie folgt aus:



Man sieht an diesem Entscheidungsbaum gut, daß unvollständige Daten zu einem komplexeren Baum führen können.

Aufgabe 2

Gegeben sei der Entscheidungsbaum aus der Übung vom 28.11.06



und die folgende Pruning-Menge (Validierungsmenge):

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D16	Sunny	Mild	High	Strong	No
D17	Rain	Hot	Normal	Weak	Yes
D18	Overcast	Cool	High	Strong	No
D19	Overcast	Mild	Normal	Strong	Yes
D20	Sunny	Cool	High	Strong	No

Wenden Sie Reduced-Error Pruning (Folie 28) auf den Entscheidungsbaum an. Benutzen Sie als Evaluierungsmaß die Anzahl der korrekt klassifizierten Beispiele der Pruning-Menge.

Lösung: Beim Reduced-Error Pruning ersetzt man sukzessive Knoten durch Blätter, die dann die Majority-Klasse anhand der Trainingsmenge im jeweiligen Knoten vorhersagen. Dieser Vorgang wiederholt sich so lange, bis keine Verbesserung mehr erreicht wird, wobei mit Verbesserung auch kein Genauigkeitsverlust gemeint ist ($\geq$), da sonst nicht die kleinste Version des Baumes erzeugt werden würde.

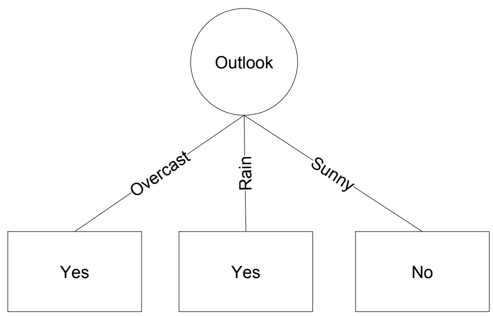
Die Genauigkeit des ursprünglichen Baumes liegt bei $^3/_5$. Wir beginnen mit dem Test des Wurzelknotens *outlook*: Die Majority-Klasse ist "yes" (9 mal "yes" und 6 mal "no"). Sagt man also immer "yes" vorher erreicht man auf der Pruning-Menge eine Genauigkeit von $^2/_5 < {}^3/_5$.

Als nächstes ersetzen wir den Knoten *wind*, bei welchem "yes" die Majority-Klasse ist. Mit dieser Ersetzung erreicht der Baum eine Genauigkeit von $^3/_5$.

Nun prunen wir den Knoten *temperature*, in welchem "no" die Majority-Klasse ist. Es ergibt sich eine Genauigkeit von $^4/_5 \geq {}^3/_5$.

Als letztes wird noch im rechten Teilbaum der Knoten *wind* ersetzt. Das Blatt sagt "no" vorher und der Baum erreicht eine Genauigkeit von $^4/_5$.

Nun suchen wir den kleinsten Baum, der mindestens eine Genauigkeit von $^4/_5$ aufweist. Daher prunen wir den Knoten *temperature* zu einem Blatt welches die Klasse “no” vorhersagt. Des weiteren können wir ohne einen Verlust an Genauigkeit den Knoten *wind* ebenfalls in ein Blatt verwandeln, das die Klasse “yes” vorhersagt. Der geprunte Baum sieht dann wie folgt aus:



Aufgabe 3

Wiederholen Sie die vorherige Aufgabe mit Regel-Post-Pruning (Folie 31). Gehen Sie wie folgt vor:

1. Konvertieren Sie den Baum in eine Regelmenge.

   Lösung: Um einen Entscheidungsbaum in eine Regelmenge zu konvertieren, muss man für jeden Pfad von der Wurzel bis zu einem Blatt eine Regel erstellen. In dem Beispiel entstehen daher 7 verschiedene Regeln:

   1. IF $outlook = overcast \rightarrow yes$

   2. IF $outlook = rain \wedge wind = strong \rightarrow no$

   3. IF $outlook = rain \wedge wind = weak \rightarrow yes$

   4. IF $outlook = sunny \wedge temperature = cool \rightarrow yes$

   5. IF $outlook = sunny \wedge temperature = hot \rightarrow no$

   6. IF $outlook = sunny \wedge temperature = mild \wedge wind = strong \rightarrow no$

   7. IF $outlook = sunny \wedge temperature = mild \wedge wind = weak \rightarrow yes$

   Ein wichtiger Unterschied dieser Art von Regeln zu den bisher beim Separate-and-conquer Lernen verwendeten ist, dass hier nun auch im Kopf einer Regel (der Teil hinter dem Pfeil) “no” stehen kann. Daher ist es nun nicht mehr so, dass alle Beispiele die von den Regeln abgedeckt werden positiv klassifiziert werden und alle nicht abgedeckten negativ, sondern die Regel, die das Beispiel abdeckt sagt genau voraus, ob es positiv (“yes”) oder negativ (“no”) ist. Wird ein Beispiel von keiner Regel abgedeckt, ist eine sog. *Default-Regel* nötig, die dann den Klassenwert vorhersagt. Solche in ihrer Semantik veränderten Regeln werden üblicherweise zur Klassifikation von Mehrklassenproblemen (mehr als 2 Klassen) verwendet.

2. Entfernen Sie eine Bedingung einer beliebigen Regel, so daß die Anzahl der korrekt klassifizierten Beispiele

   a) dieser Regel

   b) der Regelmenge

   maximiert wird.

   Lösung: Nachdem wir nun den Entscheidungsbaum in eine äquivalente Regelmenge konvertiert haben, soll diese geprunt werden. Im Aufgabenteil 2a) geschieht dies dadurch, dass jede Regel unabhängig von allen anderen jeweils um eine Bedingung geprunt wird und die Genauigkeit der so entstandenen Regel errechnet wird. Diese ist wie folgt definiert:

$$Genauigkeit = \frac{korrekt\ klassifizierte\ Beispiele}{korrekt\ klassifizierte\ Beispiele + falsch\ klassifizierte\ Beispiele} \quad (1)$$

   Als geprunte Regel wird die ausgewählt, die (unabhängig von ihrer vorherigen Genauigkeit) die höchste Genauigkeit aufweist:

   Streicht man aus der ersten Regel die einzige Bedingung heraus, so erhält man IF $FALSE \rightarrow yes$, wobei diese Regel kein Beispiel abdeckt. Da es aber in unserer Pruning-Menge 2 Beispiele gibt, deren Attributwert bei *outlook* “overcast” ist (Beispiel 3 und 4), werden diese nun durch keine Regel mehr abgedeckt. Daher müssten wir zur Bewertung der Regelmenge eine *Default-Regel* einführen. Diese könnte beispielsweise die Majority-Klasse (entweder die der Trainingsmenge oder die der Validierungsmenge) vorhersagen, womit die Genauigkeit der Regelmenge $^3/_5$ sein würde. Diese Default-Regel wird am Ende der Regelmenge eingefügt. Da die Beispiele 2,5 und 1 durch die vorhandenen Regeln (Regel 3,4 und 6) abgedeckt werden, werden nur die beiden Beispiele mit *outlook* = overcast von der neuen *Default-Regel* klassifiziert, was zur oben genannten Genauigkeit der gesamten Regelmnge führt. Da aber für die Bewertung in Aufgabe 2a) nur wichtig ist, wie viele Beispiele die aktuelle geprunte Regel richtig klassifiziert, erhalten wir eine Genauigkeit von 0.

   Nun erstellen wir alle möglichen geprunten Regeln jeweils durch das Wegstreichen von einer einzelnen Bedingung und schlagen die korrekt und nicht korrekt klassifizierten Beispiele nach:

   2. IF $outlook = rain \rightarrow no$ (0,1); IF $wind = strong \rightarrow no$ (3,1)

   3. IF $outlook = rain \rightarrow yes$ (1,0); IF $wind = weak \rightarrow yes$ (1,0)

   4. IF $outlook = sunny \rightarrow yes$ (0,2); IF $temperature = cool \rightarrow yes$ (0,2)

   5. IF $outlook = sunny \rightarrow no$ (2,0); IF $temperature = hot \rightarrow no$ (0,1)

6. IF $outlook = sunny \wedge temperature = mild \rightarrow no$ (1,0); IF $outlook = sunny \wedge wind = strong \rightarrow no$ (2,0); IF $temperature = mild \wedge wind = strong \rightarrow no$ (1,1)

7. IF $outlook = sunny \wedge temperature = mild \rightarrow yes$ (0,1); IF $outlook = sunny \wedge wind = weak \rightarrow yes$ (0,0); IF $temperature = mild \wedge wind = weak \rightarrow yes$ (0,0)

Die rot und grün markierten Regeln haben jeweils eine Genauigkeit von 1. Wir wählen die aus, die am meisten Beispiele korrekt klassifizieren, in diesem Beispiel also die grün markierten. In der ersten Iteration des Algorithmus wählen wir die geprunte 5. Regel aus. Wir erhalten somit eine Regelmenge die aus allen vorherigen Regeln (bis auf die 5.) und der geprunten 5. Regel besteht.

Im nächsten Iterationsschritt müssen wir uns die Regeln nicht nochmal betrachten, da diese jeweils nur einzeln bewertet werden. Allerdings muss die 5. Regel nochmals geprunt werden. Da diese aber nur eine Bedingung hat, gibt es eine Regel mit höherer Genauigkeit. Daher wird im nächsten Schritt die geprunte 6. Regel ausgewählt (grün markiert).

Es ergibt sich demnach folgende Regelmenge: IF $outlook = overcast \rightarrow yes$; IF $outlook = rain \wedge wind = strong \rightarrow no$; IF $outlook = rain \wedge wind = weak \rightarrow yes$; IF $outlook = sunny \wedge temperature = cool \rightarrow yes$; IF $outlook = sunny \rightarrow no$; IF $outlook = sunny \wedge wind = strong \rightarrow no$; IF $outlook = sunny \wedge temperature = mild \wedge wind = weak \rightarrow yes$

Wenn man davon ausgeht, dass die Regeln geordnet vorliegen und ein Beispiel von der ersten Regel klassifiziert wird, die dieses abdeckt, dann hat sich die Genauigkeit der Regelmenge durch das Prunen nicht verbessert. Die maximal erreichbare Genauigkeit beträgt $^4/_5$, da die erste Regel immer ein Beispiel falsch klassifiziert. Da aber die 4. Regel ebenfalls ein Beispiel falsch klassifiziert (welches allerdings durch die folgenden Regeln richtig klassifiziert werden würde), kommt es zu der Situation, dass sich die Genauigkeit der Regelmenge nicht verbessert.

Würde man nun die Regeln im Nachhinein sortieren (z.B. nach ihrer Güte und Länge) und Regeln wegstreichen, die kein Beispiel der Pruning-Menge abdecken, könnte man die Genauigkeit auf $^4/_5$ verbessern. Die resultierende Regelmenge wäre dann:

IF $outlook = overcast \rightarrow yes$; IF $outlook = rain \wedge wind = weak \rightarrow yes$; IF $outlook = sunny \rightarrow no$ (diese Regel ist kürzer als die folgende); IF $outlook = sunny \wedge wind = strong \rightarrow no$; IF $outlook = sunny \wedge temperature = cool \rightarrow yes$ (die beiden letzten Regeln decken keine Beispiele mehr ab, da die betreffenden Beispiele durch die vorherigen Regeln bereits klassifiziert wurden).

Eine andere Möglichkeit um das Abdecken eines Beispiels von mehreren Regeln in den Griff zu bekommen wäre, die Klassenvorhersagen aller möglichen Regeln zu zählen und die Klasse vorherzusagen, die von den meisten Regeln vorhergesagt wird. Diese Methode wird allerdings in dieser Übung nicht mehr beschrieben.

In der Aufgabe 2b) sollte nun die Bedingung einer Regel so entfernt werden, dass nicht mehr die Genauigkeit dieser Regel, sondern die der gesamten Regelmenge maximiert wird. Wir testen wieder alle möglichen geprunten Regeln wie in Aufgabe 2a). Angenommen wir benennen die Regeln wieder mit 1.,...,7. und die verschiedenen geprunten Regeln mit a,b,c (3b. würde dann beispielsweise bedeuten, dass die letzte Bedingung der Regel geprunt wird, also resultiert die Regel IF $outlook = rain \to yes$ entsprechen) dann erhalten wir folgende Genauigkeiten der Regelmenge (unter Beibehaltung der Reihenfolge der Regeln):

1. $^3/_5$ mit der *Default-Regel* $\to no$, die am Ende der Menge eingefügt wird und 2a. $^4/_5$; 2b. $^2/_5$ und 3a. $^3/_5$; 3b. $^3/_5$ und 4a. $^3/_5$; 4b. $^2/_5$ und 5a. $^3/_5$; 5b. $^3/_5$ und 6a. $^3/_5$; 6b. $^3/_5$; 6c. $^3/_5$ und schließlich 7a. $^3/_5$; 7b. $^3/_5$; 7c. $^3/_5$

Zur Illustration nochmal die Berechnung für 1., 2a. und 5b.:

Regelmenge bei 1.: $R = \{$IF false $\to$ yes, Regel 2 bis 7, IF true $\to$ no$\}$

- Beispiel 1: korrekt abgedeckt von Regel 7,
- Beispiel 2: korrekt abgedeckt von Regel 3,
- Beispiel 3: korrekt abgedeckt von der *Default-Regel*,
- Beispiel 4: falsch abgedeckt von der *Default-Regel* und
- Beispiel 5: falsch abgedeckt von Regel 4.

Regelmenge bei 2a.: $R = \{$1. Regel, IF wind = strong $\to$ no, Regel 3 bis 7$\}$

- Beispiel 1: korrekt abgedeckt von Regel 7,
- Beispiel 2: korrekt abgedeckt von Regel 3,
- Beispiel 3: falsch abgedeckt von Regel 1,
- Beispiel 4: korrekt abgedeckt von Regel 1 und
- Beispiel 5: korrekt abgedeckt von Regel 2.

Regelmenge bei 5b.: $R = \{$Regel 1 bis 4, IF outlook=sunny $\to$ no, Regel 6 bis 7$\}$

- Beispiel 1: korrekt abgedeckt von Regel 5,
- Beispiel 2: korrekt abgedeckt von Regel 3,
- Beispiel 3: falsch abgedeckt von Regel 1,
- Beispiel 4: korrekt abgedeckt von Regel 1,
- Beispiel 5: falsch abgedeckt von Regel 4.

Wir stellen fest, dass die höchste Genauigkeit der Regelmenge bei $\frac{4}{5}$ liegt. Damit würden wir die erste mögliche geprunte Regel auswählen, die diese Genauigkeit erreicht (die grün markierte) und mit der nächsten Iteration beginnen. Als Regelmenge haben wir bis jetzt: $\mathcal{R} = \{$Regel 1, IF wind=strong $\to$ no, Regel 3 bis 7$\}$

Wir errechnen die neuen Genauigkeiten der geprunten Regelmengen:

1. $^4/_5$ und 2. $^3/_5$ und 3a. $^3/_5$; 3b. $^3/_5$ und 4a. $^3/_5$; 4b. $^3/_5$ und 5a. $^3/_5$; 5b. $^3/_5$ und 6a. $^3/_5$; 6b. $^3/_5$; 6c. $^3/_5$ und schließlich 7a. $^3/_5$; 7b. $^3/_5$; 7c. $^3/_5$

Es wird die grün markierte geprunte Regel ausgewählt. Man stellt nun fest, dass man die Genauigkeit der Regelmenge nicht weiter erhöhen kann (da sie sich zum vorherigen Schritt nicht verbessert hat). Man könnte noch die Regeln, die keine Beispiele mehr abdecken, entfernen (Regel 4,5 und 7). Die Menge sieht wie folgt aus (wobei Regeln mit einer lokalen Abdeckung von null bereits entfernt sind, also auch die *Default-Regel*):

$\mathcal{R}$ = {IF wind = strong $\rightarrow$ no, IF outlook = rain $\wedge$ wind = weak $\rightarrow$ yes, IF outlook = sunny $\wedge$ temperature = mild $\wedge$ wind = strong $\rightarrow$ no}.

3. Wiederholen Sie Aufgabe 3.2 mit den entstandenen Regelmengen (jeweils einmal für die einzelne Regel und für die Regelmenge).

   Diese Aufgabe wurde bereits oben bearbeitet.