

# Maschinelles Lernen: Symbolische Ansätze

Musterlösung für das 5. Übungsblatt

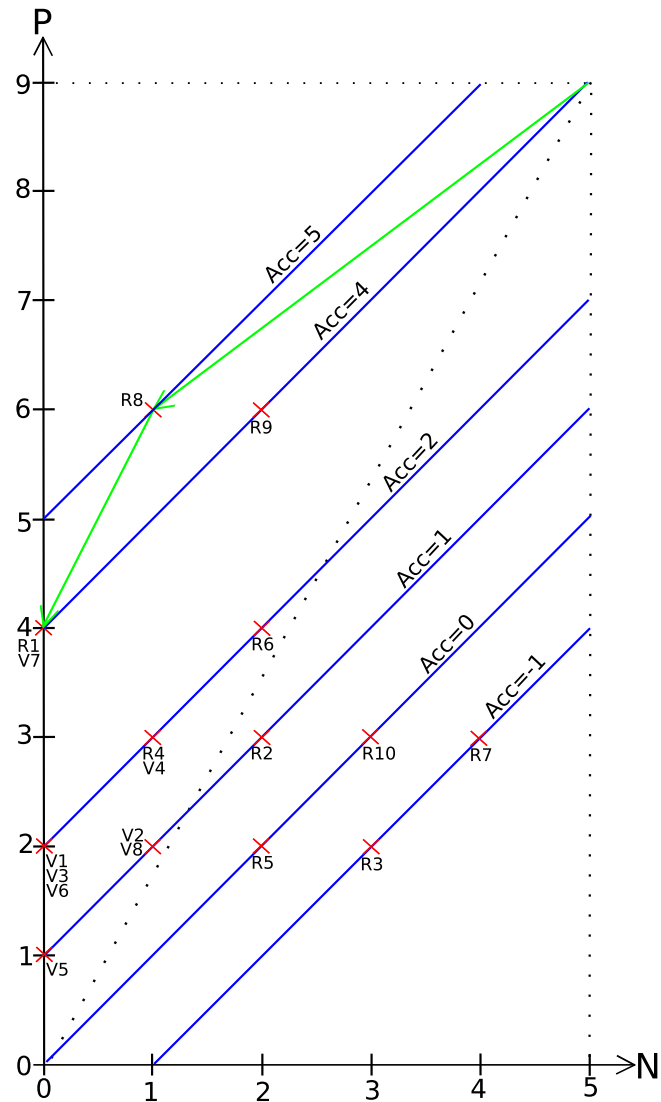
## Aufgabe 1

Visualisieren Sie den Ablauf des Covering-Algorithmus mit den Daten des letzten Übungsblatts Aufgabe 1.2. Veranschaulichen Sie das Lernen jeder einzelnen Regel im Coverage-Space. Zeichnen Sie auch alle Kandidaten-Regeln, die untersucht werden, ein und skizzieren Sie zusätzlich die Linien, die dem jeweiligen Bewertungsmaß entsprechen. Sie sollten sowohl einen Graphen für jede Regel als auch für das Lernen der gesamten Theorie anfertigen.

- für das Bewertungsmaß Accuracy, wobei die Regel mit der höchsten Bewertung ausgewählt wird

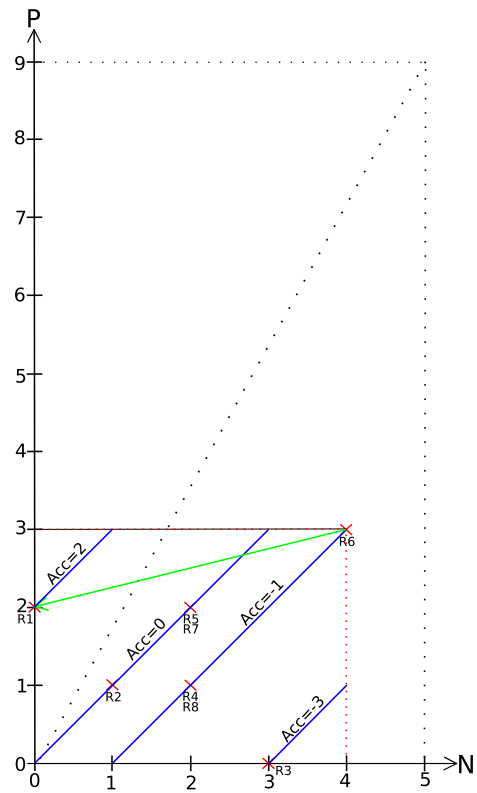
**Lösung:** Wir nummerieren die möglichen Bedingungen der Reihe nach durch und erhalten so alle Regeln (eine einzelne Bedingung entspricht einer Kandidaten-Regel). Somit haben wir für die Suche nach der ersten Regel bei Accuracy insgesamt 10 Kandidaten-Regeln, wobei folgendes gilt: R1  $\equiv$  outlook=overcast, R2  $\equiv$  outlook=rainy, ..., R10  $\equiv$  windy=TRUE. Diese Regeln zeichnen wir in den Coverage Space ein (repräsentiert durch die roten Kreuze). Da die erste gefundene Regel (R8  $\equiv$  humidity=normal) noch negative Beispiele abdeckt, muss weiter verfeinert werden. In der Grafik ist die Menge der ersten Verfeinerung mit "V plus Index" benannt, wobei wieder von oben nach unten nummeriert wird (so dass V1  $\equiv$  outlook=overcast, ..., V8  $\equiv$  windy=TRUE). Dann erstellen wir noch zusätzlich die Linien, die einer gleichen Evaluierung entsprechen (die blauen Linien).

Beginnend mit der allgemeinsten Regel zeichnen wir nun noch den Pfad der Verfeinerungen in den Coverage Space ein, der durch die grünen Linien repräsentiert wird. Es resultiert der folgende Graph für die 1. Regel:

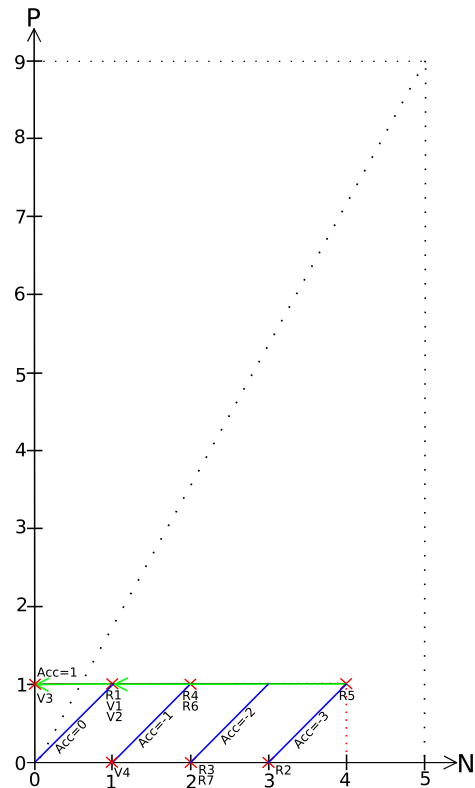


Je näher sich die Linien von Accuracy am Punkt  $(n, p) = (0, 9)$  befinden, desto höher ist deren Evaluierungswert. Da es als gleich gut gewertet wird, ein positives Beispiel abzudecken (das  $p$  in der Formel von Accuracy) und ein negatives Beispiel auszuschließen (das  $-n$  in der Formel von Accuracy), werden festgelegte Kosten angenommen. Wann immer man fixe Kosten vorgibt (diese können bekannt sein oder angenommen werden), erhält man Isometrien, deren Linien parallel sind. Aus diesem Grund wird bei der Heuristik Accuracy auch nicht die Regel V6 verwendet, sondern die vorher gefundene Regel R8, da diese unter der Annahme von gleichen Kosten eine höhere Bewertung erhält.

Nun werden alle Beispiele, die von der ersten Regel abgedeckt werden, entfernt und es verbleibt der durch die rot gepunktete Linie gekennzeichnete Unterraum. In diesem wird die 2. Regel gelernt (ohne Verfeinerungen V, da hier nicht verfeinert werden muss):



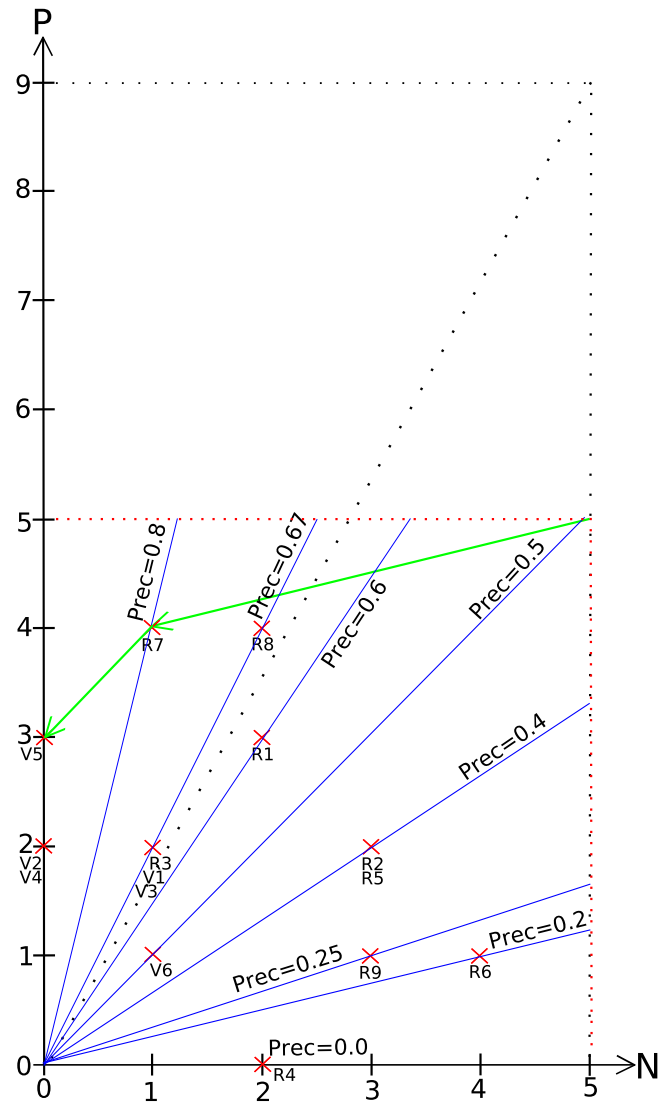
Es werden wieder die abgedeckten Beispiele entfernt und die 3. Regel wird gelernt:



Das Lernen der Theorie bei Accuracy ist im Graphen zu Precision visualisiert.

- für das Bewertungsmaß Precision (zumindest für die erste gelernte Regel)

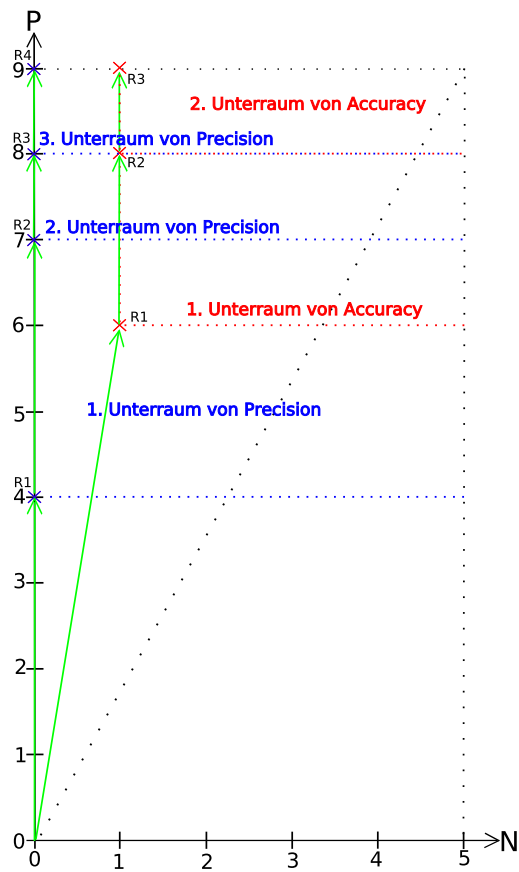
**Lösung:** Da die erste Regel bei Precision nicht verfeinert werden muss, haben wir das Lernen der 2. Regel visualisiert (nachdem die von der ersten Regel abgedeckten Beispiele bereits entfernt worden sind):



Bei der Heuristik Precision geht man davon aus, dass im Vorhinein keine Kosten festgelegt oder angenommen werden. Daher wählt diese Heuristik immer die Linie in der Isometrik aus, deren Steigung am größten ist. In unserem Beispiel ist dies bei der Auswahl der ersten Bedingung die Regel R7. Diese muss aber weiter verbessert werden. Danach gibt es 3 Verfeinerungen (V5, V2, V4), die jeweils die größte Steigung aufweisen (Evaluierungswert 1). Aus diesen wird nun zufällig die Verfeinerung V5 gewählt. Hieran wird der Unterschied zu Accuracy deutlich: Bei Accuracy würde R7 den Wert  $4 - 1 = 3$  erhalten und V5  $3 - 0 = 3$ . Es wäre also egal, welche Regel man wählen würde, was bei Precision nicht so ist. Wann

immer also eine Verfeinerung zu einer Abdeckung von keinem negativen Beispiel und einer beliebigen Anzahl von positiven Beispielen führt, bekommt diese den größtmöglichen Heuristikwert von Precision zugeordnet. Da dies immer möglich ist (verfeinere so lange, bis nur noch ein einzelnes positives Beispiel abgedeckt ist), neigt diese Heuristik eher zu Overfitting (Anpassung an die Trainingsmenge) als Accuracy.

Beim Lernen der Theorie geht man davon aus, dass diese anfangs leer ist. Danach werden sukzessive Regeln hinzugefügt und man bewegt sich immer in einen neuen Unterraum, da die Beispiele, die von der vorherigen Regel abgedeckt sind, entfernt werden. Die Unterräume von Accuracy sind durch blau gepunktete Linien und die blaue Schrift gekennzeichnet und die von Precision jeweils in rot.



## Aufgabe 2

Wie viele positive Beispiele muß eine Regel, die ein negatives Beispiel abdeckt, abdecken, damit CN2's likelihood ratio statistics zu 95% signifikant wird (threshold 3.84) bzw. zu 99% signifikant wird (threshold 6.64)?

Nehmen Sie an, daß die positiven und negativen Beispiele in der Trainingsmenge a) gleichverteilt sind, b) 90% positive und 10% negative Beispiel vorliegen. **Lösung:** Die Heuristik CN's likelihood ratio statistics (LRS) mißt, ob sich eine Regel, die eine Abdeckung von  $p$  und  $n$  hat, signifikant von der A-Priori-Verteilung unterscheidet. Das heißt, es wird getestet, ob das Verhältnis zwischen  $p$  und  $n$  sich signifikant von dem Verhältnis zwischen  $P$  und  $N$  unterscheidet. Dies ist genau dann der Fall, wenn der von LRS berechnete Wert einen durch ein Signifikanz-Niveau festgelegten Schwellenwert überschreitet.

Diese Heuristik ist  $\chi^2$ -verteilt und hat 1 Freiheitsgrad (da die "Beobachtungen" (Beispiele) nur positiv oder negativ sein können). Der Schwellenwert für ein gegebenes Signifikanzniveau kann dann aus der folgenden Tabelle entnommen werden.

Signifikanz-Niveau	0,005	0,01	0,025	0,05	0,1	0,5	0,9	<b>0,95</b>	0,975	<b>0,99</b>	0,995
Schwellen-Wert	0,00	0,00	0,00	0,00	0,02	0,45	2,71	<b>3,84</b>	5,02	<b>6,64</b>	7,88

Zur Erinnerung:

$$h_{LRS} = 2 \left( p \log \frac{p}{e_p} + n \log \frac{n}{e_n} \right) \quad (1)$$

mit

$$e_p = (p+n) \frac{P}{P+N}, e_n = (p+n) \frac{N}{P+N} \quad (2)$$

- a) Bestimmen wir zunächst die konkreten Werte der Parameter von  $h_{LRS}$ .  $n = 1$ ,  $p$  ist (sowohl für 95% als auch 99%) unbekannt und soll für beiden Niveaus bestimmt werden.  $P = N$ , da die Beispiele gleichverteilt sind. Setzen wir diese Werte in die obigen Formeln ein, erhalten wir:

$$e_p = e_n = \frac{p+1}{2} \quad (3)$$

und damit

$$h_{LRS} = 2 \left( p \log \frac{2 \cdot p}{p+1} + \log \frac{2}{p+1} \right) \quad (4)$$

Wir setzen

$$f(p) = 2 \left( p \log \frac{2 \cdot p}{p+1} + \log \frac{2}{p+1} \right) \quad (5)$$

$p$  können wir nun bestimmen, indem wir in  $f(p)$  Werte einsetzen. Sobald wir ein  $p$  gefunden, für das folgendes gilt:

- $f(p) >$  Schwellenwert und
- $f(p - 1) <$  Schwellenwert

dann wir haben das gesuchte  $p$  gefunden. Setzen wir nun also Werte ein:

$$f(1) = 0$$

...

$$f(5) = 2,911032$$

$$f(6) = 3,962432 > 3,84 \text{ (größer als der Schwellenwert für 95\%)}$$

...

$$f(8) = 6,197672$$

$$f(9) = 7,361284 > 6,64 \text{ (größer als der Schwellenwert für 99\%)}$$

Also müssen für ein Signifikanz-Niveau von 95% bzw. 99% mindestens 6 bzw. 9 positive Beispiele abgedeckt werden.

b) Diese Aufgabe ist analog zu a), bis auf die folgenden Parameter:

$$\frac{P}{P+N} = \frac{9}{10} = 1 - \frac{N}{P+N} \quad (6)$$

Damit erhalten wir die folgende Funktion:

$$g(p) = 2 \left( p \log \frac{10 \cdot p}{9 \cdot (p+1)} + \log \frac{10}{p+1} \right) \quad (7)$$

...

$$g(41) = 3,793394$$

$$g(42) = 3,956491 > 3,84 \text{ (größer als der Schwellenwert für 95\%)}$$

...

$$g(57) = 6,512724$$

$$g(58) = 6,688961 > 6,64 \text{ (größer als der Schwellenwert für 99\%)}$$

Also müssen für ein Signifikanz-Niveau von 95% bzw. 99% mindestens 42 bzw. 58 positive Beispiele abgedeckt werden.