Maschinelles Lernen: Symbolische Ansätze

TECHNISCHE UNIVERSITÄT DARMSTADT

Wintersemester 2009/2010
Musterlösung für das 5. Übungsblatt

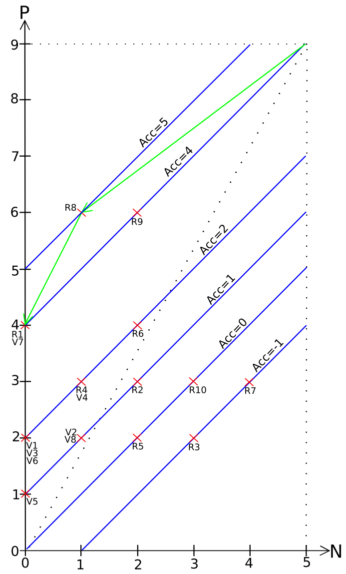
Aufgabe 1: Covering-Algorithmus und Coverage-Space

Visualisieren Sie den Ablauf des Covering-Algorithmus mit den Daten des letzten Übungsblatts Aufgabe 1b). 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

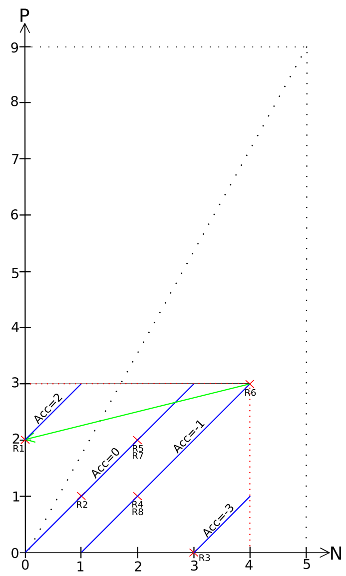
  Lösung: Wir nummerieren die möglichen Bedingungen der Reihe nach durch und erhalten so alle Regeln (eine einzelne Bedingung entspricht einer Kandidaten-Regel). Somit haben wir für die Suche nach der ersten Regel bei Accuracy insgesamt 10 Kandidaten-Regeln, wobei folgendes gilt: R1 ≡ outlook=overcast, R2 ≡ outlook=rainy, ..., R10 ≡ windy=TRUE. Diese Regeln zeichnen wir in den Coverage Space ein (repräsentiert durch die roten Kreuze). Da die erste gefundene Regel (R8 ≡ 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 nummereriert wird (so dass V1 ≡ outlook=overcast, ..., V8 ≡ 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:

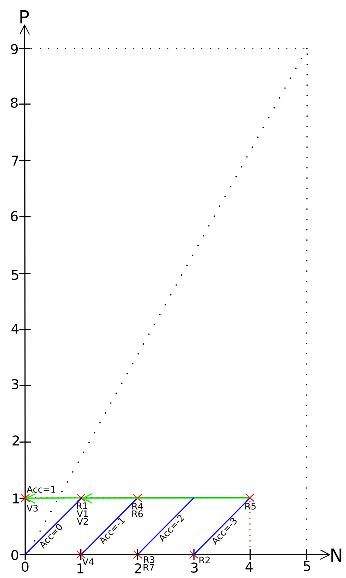


  Je näher sich die Linien von Accuracy am Punkt $(n, p) = (0, 9)$ befinden, desto höher ist deren Evaluierungswert. Da es als gleich gut gewertet wird, ein positives Beispiel abzudecken (das p in der Formel von Accuracy) und ein negatives Beispiel auszuschließen (das $-n$ in der Formel von Accuracy), werden festgelegte Kosten angenommen. Wann immer man fixe Kosten vorgibt (diese können bekannt sein oder angenommen werden), erhält man Isometriken, 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):



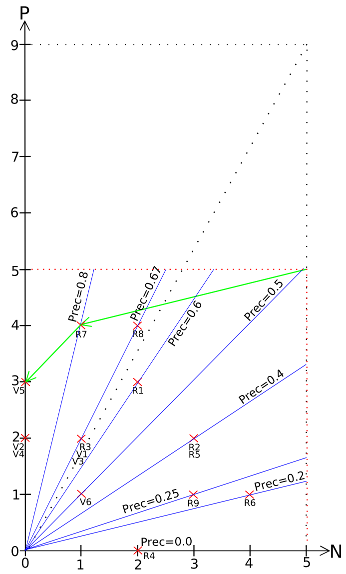
Es werden wieder die abgedeckten Beispiele entfernt und die 3. Regel wird gelernt:



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

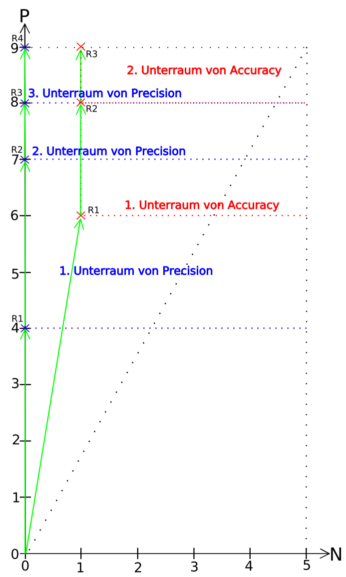
- für das Bewertungsmaß Precision (zumindest für die zweite gelernte Regel, da die erste Regel nur einmal verfeinert wird).

  Lösung: Die Beispiele, die von der ersten Regel abgedeckt sind, werden entfernt. Dann wird das Lernen der 2. Regel visualisiert:



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: Heuristiken und Äquivalenzen

In der Vorlesung haben Sie die Heuristiken Precision, Accuracy, Weighted Relative Accuracy, Gini-Index und ihre äquivalenten Berechnungen kennengelernt.

a) Zeigen Sie die Äquivalenz von

- Accuracy $h_{Acc} = \frac{p+(N-n)}{P+N}$ und $p - n$

Lösung:

$$\begin{aligned} h_{Acc} &= \frac{p+(N-n)}{P+N} \\ &= \frac{p-n}{P+N} + \frac{N}{P+N} \\ &\equiv \frac{p-n}{P+N} \\ &\equiv p-n \end{aligned}$$

- WRA $h_{WRA} = \frac{p+n}{P+N}\left(\frac{p}{p+n} - \frac{P}{P+N}\right)$ und $\frac{p}{P} - \frac{n}{N}$

Lösung:

$$\begin{aligned} h_{WRA} &= \frac{p+n}{P+N}\left(\frac{p}{p+n} - \frac{P}{P+N}\right) \\ &= \frac{1}{P+N}\left(p - \frac{(p+n)P}{P+N}\right) \\ &= \frac{1}{P+N}\left(p - p\frac{P}{P+N} - n\frac{P}{P+N}\right) \\ &= \frac{1}{P+N}\left(p\left(1 - \frac{P}{P+N}\right) - n\frac{P}{P+N}\right) \\ &= \frac{1}{P+N}\left(p\left(\frac{N}{P+N}\right) - n\frac{P}{P+N}\right) \\ &= \left(\frac{1}{P+N}\right)^2 (pN - nP) \\ &= \frac{PN}{(P+N)^2}\left(\frac{p}{P} - \frac{n}{N}\right) \\ &\equiv \frac{p}{P} - \frac{n}{N} \end{aligned}$$

- Gini-Index $h_{Gini} = 1 - \left(\frac{p}{p+n}\right)^2 - \left(\frac{n}{p+n}\right)^2$ und $\frac{pn}{(p+n)^2}$

Lösung:

$$\begin{aligned} h_{Gini} &= 1 - \left(\frac{p}{p+n}\right)^2 - \left(\frac{n}{p+n}\right)^2 \\ &= \left(\frac{1}{p+n}\right)^2 \left((p+n)^2 - p^2 - n^2\right) \\ &= \left(\frac{1}{p+n}\right)^2 (2np) \\ &\equiv \frac{np}{(p+n)^2} \end{aligned}$$

b) In der vorangegangenen Teilaufgabe haben Sie die folgende Äquivalenz

$$h_{Gini} = 1 - \left(\frac{p}{p+n}\right)^2 - \left(\frac{n}{p+n}\right)^2 \equiv \frac{pn}{(p+n)^2}$$

bewiesen. Zeigen Sie anhand dieses Ergebnisses, daß der Gini-Index äquivalent zu Precision ist, falls $p < n$ gilt, bzw. äquivalent zur negierten Precision, falls $p \geq n$ gilt.

Lösung: Variante 1

- Schritt 1: Zeigen, daß die Isometriken gleich sind (beides Geraden durch den Ursprung).

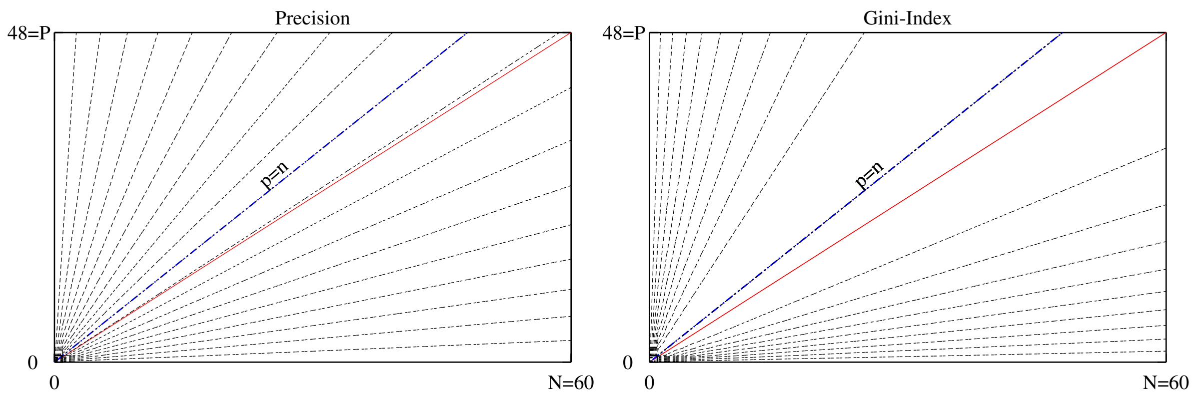
$$h_{Gini}(p,n) \equiv \frac{pn}{(p+n)^2} = \frac{k^2pn}{(kp+kn)^2} \equiv h_{Gini}(kp,kn)$$

$$h_{Prec}(p,n) \equiv \frac{p}{p+n} = \frac{kp}{kp+kn} \equiv h_{Prec}(kp,kn)$$

Beispiele:

– (p,n) = (2,3) und (3,7): $h_{Prec}(2,3) = 0,4 > 0,3 = h_{Prec}(3,7)$ und $h_{Gini}(2,3) = 0,48 > 0,42 = h_{Gini}(3,7)$

– (p,n) = (3,2) und (7,3): $h_{Prec}(3,2) = 0,6 < 0,7 = h_{Prec}(7,3)$ und $h_{Gini}(3,2) = 0,48 > 0,42 = h_{Gini}(7,3)$



Aufgabe 3: CN2's likelihood ratio statistics

Signifikanz-Niveau	0,9	0,95	0,975	0,99	0,995
Schwellen-Wert	2,71	3,84	5,02	6,64	7,88

Gegeben sei ein Datensatz, der aus 60 positiven und 40 negativen Beispielen besteht.

a) Berechnen Sie für die folgenden Regeln, von denen Ihnen nur die Abdeckung bekannt ist, die "CN2's likelihood ratio statistics" und bestimmen Sie für die oben gegebenen Signifikanz-Niveaus, ob die Regeln gepruned werden würde.

Lösung: Zur Berechnung der "CN2's likelihood ratio statistics" benötigen wir zuerst die Apriori-Verteilung der positiven und negativen Beispiele.

$$\frac{P}{P+N} = \frac{60}{60+40} = 0,6 \qquad \frac{N}{P+N} = \frac{40}{60+40} = 0,4$$

Anhand dieser Verteilungen können wir nun die für $(p+n)$ Beispiele zu erwartenden Beispiele e_p und e_n berechnen.

- R1: p=11 und n=3

Die Regel deckt (p+n)= 14 Beispiele, also sind

$$e_p = (p+n) * \frac{P}{P+N} = 14 * 0,6 = 8,4 \quad und \quad e_n = \frac{N}{P+N} = 5,6$$

Beispiele zu erwarten. Setzen wir nun diese Werte in die Formel ein, erhalten wir

$$\begin{aligned} h_{LRS} &= 2(p * \ln \frac{p}{e_p} + n * \ln \frac{n}{e_n}) \\ &= 2(11 * \ln \frac{11}{8,4} + 3 * \ln \frac{3}{5,6}) \\ &\approx 2,19 \end{aligned}$$

Da 2,19 kleiner ist als alle angegebenen Schwellenwerte, wird die Regel für alle Signifikanz-Niveaus gepruned.

- R2: p=15 und n=2

Die Regel deckt (p+n)= 17 Beispiele, also sind

$$e_p = (p+n) * \frac{P}{P+N} = 17 * 0,6 = 10,2 \quad und \quad e_n = \frac{N}{P+N} = 6,8$$

Beispiele zu erwarten. Es gilt also

$$\begin{aligned} h_{LRS} &= 2(15 * \ln \frac{15}{10,2} + 2 * \ln \frac{2}{6,8}) \\ &\approx 6,67 \end{aligned}$$

6,67 liegt zwischen den Schwellenwerten für die Signifikanzniveaus 0,99 und 0,995, also würde die Regel für das Signifikanzniveau 0,995 gepruned und für alle anderen nicht.

- R3: p=22 und n=6

Die Regel deckt (p+n)= 28 Beispiele, also sind

$$e_p = (p+n) * \frac{P}{P+N} = 28 * 0,6 = 16,8 \quad und \quad e_n = \frac{N}{P+N} = 11,2$$

Beispiele zu erwarten. Es gilt also

$$\begin{aligned} h_{LRS} &= 2(22 * \ln \frac{22}{16,8} + 6 * \ln \frac{6}{11,2}) \\ &\approx 4,38 \end{aligned}$$

4,38 liegt zwischen den Schwellenwerten für die Signifikanzniveaus 0,95 und 0,975, also würde die Regel für die Signifikanzniveaus 0,975 und größer gepruned und für 0,95 und kleiner nicht.

b) Überlegen Sie sich ohne Berechnung der "CN2's likelihood ratio statistics", warum eine Regel, die 9 positive und 6 negative Beispiele abdeckt, für alle Signifikanz-Niveaus gepruned wird.

Lösung: Die "CN2's likelihood ratio statistics" mißt, ob die Verteilung der abgedeckten Beispiele einer Regel sich signifikant von der Verteilung in den Daten unterscheidet. Aus der vorherigen Teilaufgaben wissen wir, daß die Verteilung in den Daten 0,6 zu 0,4 bzw. 3 zu 2 ist. Da 9 zu 6 Beispielen genau dieser Verteilung entspricht, existiert für kein Signifikanzniveau eine signifikante Abweichung. Demnach wird die Regel immer gepruned.