Maschinelles Lernen: Symbolische Ansätze

TECHNISCHE UNIVERSITÄT DARMSTADT

Wintersemester 2013/2014
Musterlösung für das 12. Übungsblatt

Aufgabe 1 `RelieF`

Gegeben sind folgende 12 Beispiele der Wetter-Daten:

ID	outlook	temperature	humidity	windy	play
1	sunny	hot	high	FALSE	no
2	rainy	mild	high	FALSE	yes
3	rainy	cool	normal	FALSE	yes
4	rainy	cool	normal	TRUE	no
5	overcast	cool	normal	TRUE	yes
6	sunny	mild	high	FALSE	no
7	sunny	cool	normal	FALSE	yes
8	rainy	mild	normal	FALSE	yes
9	sunny	mild	normal	TRUE	yes
10	overcast	mild	high	TRUE	yes
11	overcast	hot	normal	FALSE	yes
12	rainy	mild	high	TRUE	no

Berechnen Sie die `RelieF` Feature-Gewichte (Foliensatz Instanzenbasiertes Lernen, s.23) für alle 4 Attribute (die ID ist nur zur leichten Identifizierung eines Beispiels vorhanden). Berechnen Sie den Nearest Hit und Nearest Miss für jedes Beispiel ($r = 12$). Gehen Sie davon aus, dass jedes Beispiel genau einmal gewählt wird. (Es kann aber natürlich mehrfach als Nearest-Neighbor auftauchen) Als Distanz-Funktion nehmen Sie einfach die Anzahl der verschiedenen Attribute.

Lösung: Der Algorithmus geht folgendermaßen vor: Zuerst setzt der Benutzer den Wert für r (in der Aufgabe gilt $r = 12$). Als nächstes wählt der Algorithmus ein Beispiel zufällig aus (hier wird **nicht** sichergestellt, dass das Beispiel nicht schon einmal verwendet wurde). Dann findet er den *nearest hit* und den *nearest miss* und macht ein Update auf den Gewichtswert jedes Attributs.

In der Aufgabe gehen wir davon aus, dass ein bereits verwendetes Beispiel nicht noch einmal genommen wird. Wir machen uns die Eigenschaft zu Nutze, dass man auch eine Tabelle der Distanzen für die *Hits* und die *Misses* erstellen kann und dann einfach direkt für jedes Beispiel den *nearest hit* und *nearest miss* ausrechnen kann (indem man die Werte für jedes Attribut einfach addiert, was dem Update-Schritt des Algorithmus entspricht).

Wir erstellen also 2 Tabellen für die *Hits* und eine für die *Misses* (zu beachten ist, dass die Tabellen an der Diagonalen gespiegelt sind):

Hits für +								
ID	2	3	5	7	8	9	10	11
2	0	2	4	3	1	3	2	3
3	2	0	2	1	1	3	4	2
5	4	2	0	2	3	2	2	2
7	3	1	2	0	2	2	4	2
8	1	1	3	2	0	2	3	2
9	3	3	2	2	2	0	2	3
10	2	4	2	4	3	2	0	3
11	3	2	2	2	2	3	3	0

Hits für -				
ID	1	4	6	12
1	0	4	1	3
4	4	0	4	2
6	1	4	0	2
12	3	2	2	0

Misses				
ID	1	4	6	12
2	2	3	1	1
3	3	1	3	3
5	4	1	4	3
7	2	2	2	4
8	3	2	2	2
9	3	2	2	2
10	3	3	2	1
11	2	3	3	4

In den Tabellen sind die *nearest neighbors* der gleichen Klasse rot markiert (wobei bei gleicher Distanz immer das erste Beispiel in der Tabelle ausgewählt wurde). Nun erstellt man eine große Tabelle in der man für jedes Attribut die Unterscheidungen festhält (gleicher Wert = Distanz 0, unterschiedlicher Wert = Distanz 1). In dieser Tabelle repräsentiert eine Zeile jeweils einen Durchlauf des Algorithmus (wobei wir die Gewichte erst am Ende updaten, bzw. berechnen).

Hits						Misses				
ID	ID	outl.	temp.	hum.	wind	ID	outl.	temp.	hum.	wind
2	8	0	0	1	0	6	1	0	0	0
3	7	1	0	0	0	4	0	0	0	1
5	3	1	0	0	1	4	1	0	0	0
7	3	1	0	0	0	1	0	1	1	0
8	2	0	0	1	0	4	0	1	0	1
9	5	1	1	0	0	4	1	1	0	0
10	2	1	0	0	1	12	1	0	0	0
11	3	1	1	0	0	1	1	0	1	0
1	6	0	1	0	0	2	1	1	0	0
4	12	0	1	1	0	3	0	0	0	1
6	1	0	1	0	0	2	1	0	0	0
12	4	0	1	1	0	2	0	0	0	1
		6	6	4	2		7	4	2	4

Da man nun alle Werte zu Verfügung hat, kann man die `RelieF` Feature-Gewichte errechnen:
$W(outlook) = -{}^{6}/_{12} + {}^{7}/_{12} = {}^{1}/_{12}$
$W(temperature) = -{}^{6}/_{12} + {}^{4}/_{12} = -{}^{1}/_{6}$ $(\rightarrow W(temperature) = 0)$
$W(humidity) = -{}^{4}/_{12} + {}^{2}/_{12} = -{}^{1}/_{6}$ $(\rightarrow W(humidity) = 0)$
$W(wind) = -{}^{2}/_{12} + {}^{4}/_{12} = {}^{1}/_{6}$

Aufgabe 2 Diskretisierungsmethoden

Gegeben sei folgende Version der Wetter-Daten mit 12 Trainings-Beispielen und 2 numerischen Attributen.

ID	outlook	temperature	humidity	windy	play
1	sunny	85	85	FALSE	no
2	rainy	70	96	FALSE	yes
3	rainy	68	80	FALSE	yes
4	rainy	65	70	TRUE	no
5	overcast	64	65	TRUE	yes
6	sunny	72	95	FALSE	no
7	sunny	69	70	FALSE	yes
8	rainy	75	80	FALSE	yes
9	sunny	75	70	TRUE	yes
10	overcast	72	90	TRUE	yes
11	overcast	81	75	FALSE	yes
12	rainy	71	91	TRUE	no

Diskretisieren Sie die beiden numerischen Attribute mit den Verfahren, die Sie in der Vorlesung kennen gelernt haben. Wählen Sie die Anzahl der Intervalle so, daß Sie die bekannten Daten erhalten könnten (drei Werte für Temperature, zwei für Humidity). Vergleichen Sie die Resultate miteinander und mit den bekannten Daten (die aus der vorherigen Aufgabe).

- equal-width

  Lösung: Bei equal-width wird versucht gleich große Intervalle zu bilden:

  - Attribut “temperature”: $85 - 64 = 21/3 = 7 \rightarrow$ Die Intervalle sollten also die Größe 7 haben. Es resultieren folgende Intervalle: $[64,70],[71,77],[78,85]$ ($[64,71],[72,78],[79,85]$ wäre auch zulässig, man muss sich hier eine geeignete Einteilung überlegen)

    Dies ergibt dann die Intervalle $(-\infty,70],(70,77],(77,\infty)$
  - Attribut “humidity”: $96 - 65 = \lfloor 31/2 \rfloor = 15 \rightarrow$ Die Intervalle sollten also die Größe 15 haben: $[65,80],[81,96]$

    Dies ergibt dann die Intervalle $(-\infty,80],(80,\infty)$.

- equal-frequency

  Lösung: Bei equal-frequency wird vesucht möglichst gleich viele Beispiele pro Intervall zu haben:

  - Attribut “temperature”: $12/3 = 4$ Es sollten also jeweils 4 Beispiele in einem Intervall liegen.

    Daher liegen die Beispiele

    Nr.5 temp.=64, Nr.4 temp.=65, Nr.3 temp.=68, Nr.7 temp.=69

    im ersten Intervall, die Beispiele

    Nr.2 temp.=70, Nr.12 temp.=71, Nr.6 temp.=72, Nr.10 temp.=72

    im zweiten und die Beispiele

    Nr.8 temp.=75, Nr.9 temp.=75, Nr.11 temp.=81, Nr.1 temp.=85

    im dritten Intervall.
  - Attribut “humidity”: $12/2 = 6$ Es sollten jeweils 6 Beispiele in einem Intervall liegen.

    Die Beispiele

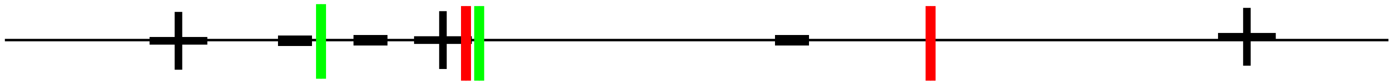
    Nr.5 hum.=65, Nr.4 hum.=70, Nr.7 hum.=70, Nr.9 hum.=70, Nr.11 hum.=75, Nr.3 hum.=80, Nr.8 hum.=80 (die beiden Beispiel mit humidity = 80 könnten auch im zweiten Intervall liegen)

    im ersten und die Beispiele

    Nr.1 hum.=85, Nr.10 hum.=90, Nr.12 hum.=91, Nr.6 hum.=95, Nr.2 hum.=96

    im zweiten Intervall.

Die folgende Grafik erklärt den Unterschied beider Methoden nochmals anschaulich, wobei es sich hier um ein allgemeines Beispiel handelt, das nicht den vorherigen Aufgaben entspricht (die roten Striche sind für equal-width und die grünen für equal-frequency):



- chi-merge

Lösung: Wir werden nicht auf die Berechnung jedes einzelnen χ^2-Wertes eingehen, wir werden jedoch zwei Beispiele exemplarisch berechnen und anschließend nur noch die zur Berechnung benötigten Werte und das Ergebnis auflisten.

Zur Erinnerung der χ^2-Wert zweier Intervalle I_1 und I_2 berechnet sich wie folgt:

$$\chi^s = \sum_{i=1}^{2} \sum_{j=1}^{c} \frac{\left(A_{ij} - E_{ij}\right)^2}{E_{ij}},$$

mit

A_{ij}: Anzahl der Beispiele in I_i, die zur Klasse j gehören

E_{ij}: zu erwartende Anzahl von Beispiele in I_i, die zur Klasse j gehören

$$E_{ij} = N_i \cdot \frac{C_j}{N_1 + N_2}$$

N_i: Anzahl der Beispiele in I_i

$$N_i = \sum_{j=1}^{c} A_{ij}$$

C_j: Anzahl der Beispiele in I_1 und I_2, die zur Klasse j gehören. Sollte dieser Wert 0 sein, wird diese Klasse bei der Berechnung des χ^2-Wertes nicht berücksichtigt (da Division durch 0), d.h. alle entsprechenden Terme werden bei der Berechnung auf 0 gesetzt.

$$C_j = A_{1j} + A_{2j}$$

Anmerkung:

Der χ^2-Wert zweier Intervalle, die exakt die gleiche Klassenverteilung aufweisen, ist minimal bzw. 0, diese Intervalle werden somit verbunden (falls noch kein Stopkriterium, wie z.B. minimale Intervallanzahl, erfüllt ist). D.h. natürlich auch, daß in der Initialphase reine Intervalle (also Intervalle mit nur einer auftretenden Klasse) sofort verbunden werden können.

Beispiel 1 (2 Klassen)

Intervall I_1 hat 3 positive Beispiele und 1 negatives, Intervall I_2 hat 4 positive und 2 negative Beispiele. Also haben wir folgende Werte

$$A_{1+} = 3 \qquad A_{1-} = 1 \qquad A_{2+} = 4 \qquad A_{2-} = 2$$

Hieraus berechnen wir zunächst

$$C_+ = 7 \qquad C_- = 3 \qquad N_1 = 4 \qquad N_2 = 6$$

und anschließend

$$E_{1+} = 4\frac{7}{4+6} = 2,8 \qquad E_{1-} = 4\frac{3}{4+6} = 1,2 \qquad E_{2+} = 6\frac{7}{4+6} = 4,2 \qquad E_{2-} = E_{2+} = 6\frac{3}{4+6} = 1,8$$

Mit diesen Werten können wir nun den χ^2-Wert berechnen:

$$\begin{aligned} \chi^2 &= \frac{(3-2,8)^2}{2,8} + \frac{(1-1,2)^2}{1,2} + \frac{(4-4,2)^2}{4,2} + \frac{(2-1,8)^2}{1,8} \\ &= \frac{1}{70} + \frac{1}{30} + \frac{1}{105} + \frac{1}{45} \\ &= \frac{9+21+6+14}{630} = \frac{5}{63} \end{aligned}$$

Beispiel 2 (3 Klassen)

Die Beispiele der 3 Klassen a, b und c verteilen sich auf das Intervall I_1 bzw. auf I_2 wie folgt

$$A_{1a} = 5 \qquad A_{1b} = 3 \qquad A_{1c} = 4 \qquad A_{2a} = 4 \qquad A_{2b} = 3 \qquad A_{2c} = 1$$

Hieraus berechnen wir

$$C_a = 9 \qquad C_b = 6 \qquad C_c = 5 \qquad N_1 = 12 \qquad N_2 = 8$$

und

$$\begin{aligned} E_{1a} &= 12\frac{9}{12+8} = 5,4 & E_{1b} &= 12\frac{6}{12+8} = 3,6 & E_{1c} &= 12\frac{5}{12+8} = 3 \\ E_{2a} &= 8\frac{9}{12+8} = 3,6 & E_{2b} &= 8\frac{6}{12+8} = 2,4 & E_{2c} &= 8\frac{5}{12+8} = 2 \end{aligned}$$

Daraus berechnen wir

$$\begin{aligned} \chi^2 &= \frac{(5-5,4)^2}{5,4} + \frac{(3-3,6)^2}{3,6} + \frac{(4-3)^2}{3} + \frac{(4-3,6)^2}{3,6} + \frac{(3-2,4)^2}{2,4} + \frac{(1-2)^2}{2} \\ &\approx 1,16 \end{aligned}$$

Temperature: Wir möchten mittels χ-Merge Temperature in 3 diskrete Werte aufteilen. Zuerst sortieren wir die Werte des Attributs *Temperature*. Je nach auf- bzw. absteigender Sortierung erhält man unterschiedliche Ergebnisse. Wir entscheiden uns für eine aufsteigende Sortierung. Als erstes verbinden wir alle reinen Intervalle (s. Anmerkung).

I_1	I_2	A_{1y}	A_{1n}	A_{2y}	A_{2n}	N_1	N_2	C_y	C_n	E_{1y}	E_{1n}	E_{2y}	E_{2n}	χ^2
64	65	1	0	0	1	1	1	1	1	0,50	0,50	0,50	0,50	2,00
65	68-70	0	1	3	0	1	3	3	1	0,75	0,25	2,25	0,75	4,00
68-70	71	3	0	0	1	3	1	3	1	2,25	0,75	0,75	0,25	4,00
71	**72**	**0**	**1**	**1**	**1**	**1**	**2**	**1**	**2**	**0,33**	**0,67**	**0,67**	**1,33**	**0,75**
72	75-81	1	1	3	0	2	3	4	1	1,60	0,40	2,40	0,60	1,88
75-81	85	3	0	0	1	3	1	3	1	2,25	0,75	0,75	0,25	4,00

Wir wählen die Intervalle mit geringstem χ^2-Wert aus, also die Intervalle $[71,71]$ und $[72,72]$, und verbinden diese zu $[71,72]$. Anschließend berechnen wir die χ^2-Werte von $[71,72]$ und seinen benachbarten Intervalle neu. Alle anderen Werte können übernommen werden.

I_1	I_2	A_{1y}	A_{1n}	A_{2y}	A_{2n}	N_1	N_2	C_y	C_n	E_{1y}	E_{1n}	E_{2y}	E_{2n}	χ^2
64	**65**	**1**	**0**	**0**	**1**	**1**	**1**	**1**	**1**	**0,50**	**0,50**	**0,50**	**0,50**	**2**
65	68-70	0	1	3	0	1	3	3	1	0,75	0,25	2,25	0,75	4
68-70	71-72	3	0	1	2	3	3	4	2	2,00	1,00	2,00	1,00	3
71-72	75-81	1	2	3	0	3	3	4	2	2,00	1,00	2,00	1,00	3
75-81	85	3	0	0	1	3	1	3	1	2,25	0,75	0,75	0,25	4

Diesmal verschmelzen wir $[64,64]$ und $[65,65]$ zu $[64,65]$und berechnen anhand der veränderten Werte die benötigten Werte (s.o.) neu.

I_1	I_2	A_{1y}	A_{1n}	A_{2y}	A_{2n}	N_1	N_2	C_y	C_n	E_{1y}	E_{1n}	E_{2y}	E_{2n}	χ^2
64-65	**68-70**	**1**	**1**	**3**	**0**	**2**	**3**	**4**	**1**	**1,6**	**0,4**	**2,4**	**0,6**	**1,88**
68-70	71-72	3	0	1	2	3	3	4	2	2	1	2	1	3
71-72	75-81	1	2	3	0	3	3	4	2	2	1	2	1	3
75-81	85	3	0	0	1	3	1	3	1	2,25	0,75	0,75	0,25	4

Der χ^2-Wert von $[64, 65]$ und $[68, 70]$ ist minimal, deshalb verschmelzen wir diese Intervalle zu $[64, 70]$ und berechnen neu.

I_1	I_2	A_{1y}	A_{1n}	A_{2y}	A_{2n}	N_1	N_2	C_y	C_n	E_{1y}	E_{1n}	E_{2y}	E_{2n}	χ^2
64-70	**71-72**	**4**	**1**	**1**	**2**	**5**	**3**	**5**	**3**	**3,13**	**1,88**	**1,88**	**1,13**	**1,74**
71-72	75-81	1	2	3	0	3	3	4	2	2	1	2	1	3
75-81	85	3	0	0	1	3	1	3	1	2,25	0,75	0,75	0,25	4

Wir verschmelzen $[64, 70]$ und $[71, 72]$ und erhalten damit die folgenden Intervalle: $[64, 72]$, $[75, 81]$ und $[85, 85]$. Offensichtlich können nicht alle möglichen Temperaturen einem dieser Intervalle zugeordnet werden. Aus diesem Grund müssen die enstandenden Intervalle angepaßt werden. Hierfür gibt es verschiedene Möglichkeiten diese Intervalle zu bearbeiten, u.a.

- $(-\infty, 72]$, $(72, 81]$ und $(81, \infty)$ oder
- $(-\infty, 75)$, $[75, 85)$ und $[85, \infty)$ oder
- $(-\infty, 73.5)$, $[73.5, 83)$ und $[83, \infty)$ usw.

Humidity:

Für Humidity gehen wir analog vor, wobei diesmal die Diskretisierung nur 2 Werte ergeben soll.

I_1	I_2	A_{1y}	A_{1n}	A_{2y}	A_{2n}	N_1	N_2	C_y	C_n	E_{1y}	E_{1n}	E_{2y}	E_{2n}	χ^2
65	**70**	**1**	**0**	**2**	**1**	**1**	**3**	**3**	**1**	**0,75**	**0,25**	**2,25**	**0,75**	**0,44**
70	75-80	2	1	3	0	3	3	5	1	2,50	0,50	2,50	0,50	1,20
75-80	85	3	0	0	1	3	1	3	1	2,25	0,75	0,75	0,25	4,00
85	90	0	1	1	0	1	1	1	1	0,50	0,50	0,50	0,50	2,00
90	91-95	1	0	0	2	1	2	1	2	0,33	0,67	0,67	1,33	3,00
91-95	96	0	2	1	0	2	1	1	2	0,67	1,33	0,33	0,67	3,00

Es werden $[65, 65]$ und $[70, 70]$ zu $[65, 70]$ verbunden, anschließend wird neu berechnet.

I_1	I_2	A_{1y}	A_{1n}	A_{2y}	A_{2n}	N_1	N_2	C_y	C_n	E_{1y}	E_{1n}	E_{2y}	E_{2n}	χ^2
65-70	**75-80**	**3**	**1**	**3**	**0**	**4**	**3**	**6**	**1**	**3,43**	**0,57**	**2,57**	**0,43**	**0,88**
75-80	85	3	0	0	1	3	1	3	1	2,25	0,75	0,75	0,25	4,00
85	90	0	1	1	0	1	1	1	1	0,50	0,50	0,50	0,50	2,00
90	91-95	1	0	0	2	1	2	1	2	0,33	0,67	0,67	1,33	3,00
91-95	96	0	2	1	0	2	1	1	2	0,67	1,33	0,33	0,67	3,00

Diesmal verschmelzen wir $[65, 70]$ und $[75, 80]$ zu $[65, 80]$ und berechnen neu.

I_1	I_2	A_{1y}	A_{1n}	A_{2y}	A_{2n}	N_1	N_2	C_y	C_n	E_{1y}	E_{1n}	E_{2y}	E_{2n}	χ^2
65-80	85	6	1	0	1	7	1	6	2	5,25	1,75	0,75	0,25	3,43
85	**90**	**0**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**0,50**	**0,50**	**0,50**	**0,50**	**2,00**
90	91-95	1	0	0	2	1	2	1	2	0,33	0,67	0,67	1,33	3,00
91-95	96	0	2	1	0	2	1	1	2	0,67	1,33	0,33	0,67	3,00

Wir wählen $[85, 85]$ und $[90, 90]$ aus und verschmelzen sie zu $[85, 90]$. Anschließend berechnen wir die fehlenden Werte. Es werden $[65, 80]$ und $[85, 90]$ zu $[65, 90]$ verschmolzen und die benötigten Werte berechet.

I_1	I_2	A_{1y}	A_{1n}	A_{2y}	A_{2n}	N_1	N_2	C_y	C_n	E_{1y}	E_{1n}	E_{2y}	E_{2n}	χ^2
65-80	**85-90**	**6**	**1**	**1**	**1**	**7**	**2**	**7**	**2**	**5,44**	**1,56**	**1,56**	**0,44**	**1,15**
85-90	91-95	1	1	0	2	2	2	1	3	0,5	1,5	0,5	1,5	1,33
91-95	96	0	2	1	0	2	1	1	2	0,67	1,33	0,33	0,67	3

I_1	I_2	A_{1y}	A_{1n}	A_{2y}	A_{2n}	N_1	N_2	C_y	C_n	E_{1y}	E_{1n}	E_{2y}	E_{2n}	χ^2
65-90	91-95	7	2	0	2	9	2	7	4	5,73	3,27	1,27	0,73	4,28
91-95	**96**	**0**	**2**	**1**	**0**	**2**	**1**	**1**	**2**	**0,67**	**1,33**	**0,33**	**0,67**	**3**

Als letzes verschmelzen wir $[91,95]$ und $[96,96]$ und erhalten als Ergebnis die Intervalle $[65,90]$ und $[91,96]$.

- Entropy-split

Lösung: Bei Entropy-split zerlegen wir eines der bestehenden Intervalle in jeweils zwei Teilintervalle, falls das zu zerlegende Intervall den niedristen E-Score (Berechnung von Entropy-split ist äquivalent zu diesem Maß) aufweist. E-Score ist wiederum äquivalent zu Information-Gain. Beide Maße und deren Berechnung wurden bereits ausführlich in den Übungen über Entscheidungsbäume besprochen. Aus diesem Grund werden wir nicht weiter auf deren Berechnung eingehen.

Temperature: Wir sortieren zunächst die Attributwerte (aufsteigend).

	A<Wert		A≥Wert		
Wert	positiv	negativ	positiv	negativ	E-Score
64	0	0	8	4	0,918
65	1	0	7	4	0,867
68	1	1	7	3	0,901
69	2	1	6	3	0,918
70	3	1	5	3	0,907
71	4	1	4	3	0,876
72	4	2	4	2	0,918
75	5	3	3	1	0,907
81	7	3	1	1	0,901
85	8	3	0	1	**0,775**

Der Teilungspunkt 85 hat den besten (niedrigsten) E-Score. Wir teilen das Intervall $(-\infty, \infty)$ in $(-\infty, 85)$ und $[85, \infty)$ auf.

Wert	A<Wert		A≥Wert		E-Score
	positiv	negativ	positiv	negativ	
64	0	0	8	3	0,845
65	1	0	7	3	0,801
68	1	1	7	3	0,807
69	2	1	6	2	0,840
70	3	1	5	2	0,844
71	4	1	4	2	0,829
72	4	2	4	1	0,829
75	5	3	3	0	**0,694**
81	7	3	1	0	0,801
85	0	0	0	1	-

Jetzt ist 75 der beste Teilungspunkt. Wir zerlegen das Intervall $(-\infty, 85)$ in $(-\infty, 75)$ und $[75, 85)$. Demnach sieht die Diskretisierung von Temperature wie folgt aus: $(-\infty, 75)$, $[75, 85)$ und $[85, \infty)$.

Humidity:

Wert	A<Wert		A≥Wert		E-Score
	positiv	negativ	positiv	negativ	
65	0	0	8	3	0,918
70	1	0	7	4	0,867
75	3	1	5	3	0,907
80	4	1	4	3	0,876
85	6	1	2	3	**0,750**
90	6	2	2	2	0,874
91	7	2	1	2	0,803
95	7	3	1	1	0,867
96	7	4	1	0	0,801

Wir teilen das Intervall $(-\infty, \infty)$ in $(-\infty, 86)$ und $[86, \infty)$ auf.

Nun gehen wir noch kurz auf den Vergleich der Ergebnisse ein: Da man nicht weiß, wie die numerischen Daten des originalen Datensatzes diskretisiert wurden, kann man hier keine genaue Aussage treffen. Daher kann man sich nur die Unterschiede zwischen den einzelnen Methoden und den Originaldaten anschauen (diese sind rot markiert):

Tabelle 1: Vergleich: temperature

ID	outlook	temperature	equal-width	equal-frequency	chi-merge	entropy-split
1	sunny	hot	hot	hot	hot	hot
2	rainy	mild	cool	mild	cool	cool
3	rainy	cool	cool	cool	cool	cool
4	rainy	cool	cool	cool	cool	cool
5	overcast	cool	cool	cool	cool	cool
6	sunny	mild	mild	mild	cool	cool
7	sunny	cool	cool	cool	cool	cool
8	rainy	mild	mild	hot	mild	mild
9	sunny	mild	mild	hot	mild	mild
10	overcast	mild	mild	mild	cool	cool
11	overcast	hot	hot	hot	mild	mild
12	rainy	mild	mild	mild	cool	cool

Tabelle 2: Vergleich: humidity

ID	outlook	humidity	equal-width	equal-frequency	chi-merge	entropy-split
1	sunny	high	high	high	normal	high
2	rainy	high	high	high	high	high
3	rainy	normal	normal	normal	normal	normal
4	rainy	normal	normal	normal	normal	normal
5	overcast	normal	normal	normal	normal	normal
6	sunny	high	high	high	high	high
7	sunny	normal	normal	normal	normal	normal
8	rainy	normal	normal	normal	normal	normal
9	sunny	normal	normal	normal	normal	normal
10	overcast	high	high	high	normal	high
11	overcast	normal	normal	normal	normal	normal
12	rainy	high	high	high	high	high