Maschinelles Lernen: Symbolische Ansätze

TECHNISCHE UNIVERSITÄT DARMSTADT

Wintersemester 2008/2009
Musterlösung für das 10. Übungsblatt

Aufgabe 1: AdaBoost

Rechnen Sie das AdaBoost-Beispiel aus der Vorlesung (Ensemble-Methoden, Folien 15ff) nach. Verwenden Sie für die einzelnen Datenpunkte die folgenden Koordinaten (x, y, Klasse):

```
1,  5,  +              3,  1,  -
2,  2,  +              4,  6,  -
5,  8,  +              7,  4,  -
6, 10,  +              9,  3,  -
8,  7,  +             10,  9,  -
```

Als Basis-Lerner sollen Decision Stumps (also waagrechte bzw. senkrechte Splits, z.B. $x > 4 \rightarrow +$) verwendet werden. Der Basislerner wählt unter allen möglichen Splits jenen aus, bei dem die Gesamtsumme der Gewichte der falsch klassifizierten Beispiele minimiert wird.

a) Berechnen Sie die ersten 3 AdaBoost-Iterationen.

Lösung: In jeder Iteration sind jeweils 10 horizontale (y ≤ Wert) und 10 vertikale Splits möglich. Jeder Split ermöglicht 2 Vergleiche, z.B. y ≤ 4 und y > 4. Für jeden dieser Vergleiche addieren wir die Gewichte der Beispiele auf, die durch den Vergleich falsch klassifiziert werden. Anschließend wählen wir denjenigen Vergleich aus, der die geringste Summe aufweist. Wir berechnen danach die Gewichtung α_m des aus dem Vergleich resultierenden Klassifizierer (Decision Stump). Die Gewichte der Beispiele werden unter Verwendung der Gewichtung α_m erhöht bzw. gesenkt, falls sie falsch bzw. richtig klassifiziert werden. Am Ende jeder Iteration werden die Gewichte der Beispiel so normiert, daß ihre Summe eins ergibt. Die Folien 11-15 illustrieren diese Vorgehensweise.

Hinweis: *Bei den berechneten Fehlern können abhängig von der verwendeten Methode (Berechnung einer oder beider Tabellenspalte(n)) Abweichungen auftreten, da die Gesamtsumme der Beispielsgewichte bedingt durch die Rundung der Werte selten 1 beträgt.*

Beginnen wir nun mit den Berechnungen. Am Anfang haben alle Beispiele das Gewicht $1/10$ (10 Beispiele). Betrachten wir nun zuerst alle vertikalen Splits.

	Fehler	
Wert	x ≤ Wert ⇒ +	x > Wert ⇒ +
1	$4/10$	$6/10$
2	$3/10$	$7/10$
3	$4/10$	$6/10$
4	$5/10$	$5/10$
5	$4/10$	$6/10$
6	$3/10$	$7/10$
7	$4/10$	$6/10$
8	$3/10$	$7/10$
9	$4/10$	$6/10$
10	$5/10$	$5/10$

Entsprechend erhalten für die horizontalen Splits die folgenden Fehler.

Wert	Fehler	
	$y \leq$ Wert $\Rightarrow +$	$y >$ Wert $\Rightarrow +$
1	$6/10$	$4/10$
2	$5/10$	$5/10$
3	$6/10$	$4/10$
4	$7/10$	$3/10$
5	$6/10$	$4/10$
6	$7/10$	$3/10$
7	$6/10$	$4/10$
8	$5/10$	$5/10$
9	$6/10$	$4/10$
10	$5/10$	$5/10$

Betrachten wir beide Tabellen, sehen wir, daß es 4 Splits mit minimalen Fehler gibt (rot markiert). Wir entscheiden uns für den zuerst gefundenen Split ($x \leq 2 \Rightarrow +$). Berechnen wir nun das Gewicht des resultierenden Klassifizierers. Hierfür benötigen wir zunächst den Fehler err_1:

$$err_1 = \frac{3}{10}$$

Hiermit können wir nun das Gewicht α_1 des Klassifizierers berechnen:

$$\alpha_1 = \frac{1}{2}\log\left(\frac{1-err_m}{err_m}\right) = \frac{1}{2}\log\left(\frac{7}{3}\right) \approx 0,424$$

Damit ergeben sich die folgenden Faktoren, mit denen die einzelnen Gewichte multipliziert werden:

$$w_i \leftarrow \begin{cases} w_i \cdot e^{-\alpha_1} \approx 0,654, & \text{falls } w_i \text{ korrekt klassifiziert wird} \\ w_i \cdot e^{\alpha_1} \approx 1,528, & \text{falls } w_i \text{ falsch klassifiziert wird} \end{cases}$$

Zu beachten ist, dass die Werte eigentlich nicht mit dem alten Gewicht multipliziert sind, da es sich im ersten Schritt über die Normalisierung sowieso wieder herauskürzen würde. Da sieben Beispiele korrekt und drei falsch klassifizert wurden, erhalten wir:

$$3 \cdot 1,528 + 7 \cdot 0,654 = 9,162$$

als Gesamtsumme der Gewichte. Diese wollen wir auf eins normieren, aus diesem Grund teilen wir alle Gewichte durch 9,162. Damit erhalten wir folgende Gewichte:

$$w_i = \begin{cases} 0,071, & \text{falls } w_i \text{ korrekt klassifiziert wird} \\ 0,167, & \text{falls } w_i \text{ falsch klassifiziert wird} \end{cases}$$

und folgende Tabelle:

x	y	Gewicht
1	5	0,071
2	2	0,071
3	1	0,071
4	6	0,071
5	8	0,167
6	10	0,167
7	4	0,071
8	7	0,167
9	3	0,071
10	9	0,071

Suchen wir nun den nächsten Split. Wir betrachten zuerst vertikale Splits

Wert	Fehler	
	x ≤ Wert ⇒ +	x > Wert ⇒ +
1	0,572	0,428
2	0,501	0,499
3	0,572	0,428
4	0,643	0,357
5	0,476	0,524
6	0,309	0,691
7	0,38	0,62
8	0,213	0,787
9	0,284	0,716
10	0,355	0,645

und anschließend die horizontalen.

Wert	Fehler	
	y ≤ Wert ⇒ +	y > Wert ⇒ +
1	0,286	0,714
2	0,643	0,357
3	0,286	0,714
4	0,215	0,785
5	0,286	0,714
6	0,215	0,785
7	0,382	0,618
8	0,549	0,451
9	0,478	0,522
10	0,645	0,355

Der beste Split ist $x \leq 8 \Rightarrow +$. Das heißt die Punkte (3, 1), (4, 6) und (7, 4) werden falsch und alle anderen richtig klassifiziert. Demnach hat err_2 den folgenden Wert:

$$err_2 \approx 0,213$$

Mit err_2 können wir α_2 berechnen:

$$\alpha_2 = \frac{1}{2} \log \left(\frac{0,787}{0,213} \right) \approx 0,652$$

Berechnen wir nun die Faktoren e^{α_2} bzw. $e^{-\alpha_2}$, mit denen wir die Gewichte multiplizieren:

$$e^{-\alpha_2} = 0,521$$
$$e^{\alpha_2} = 1,919$$

Damit ergibt sich die folgende Tabelle:

Altes Gewicht	Neues Gewicht	
	Korrekt klassifiziert	falsch klassifiziert
0,071	0,037	0,136
0,167	0,087	0,32

Multiplizieren wir die Gewichte der korrekt (blau) und falsch (rot) klassifizierten Beispiele mit den entsprechenden Faktoren und normieren diese, erhalten wir folgende Gewichte:

x	y	Altes Gewicht	Neues Gewicht	
			Nicht normiert	Normiert
1	5	0,071	0,037	0,045
2	2	0,071	0,037	0,045
3	1	0,071	0,136	0,166
4	6	0,071	0,136	0,166
5	8	0,167	0,087	0,106
6	10	0,167	0,087	0,106
7	4	0,071	0,136	0,166
8	7	0,167	0,087	0,106
9	3	0,071	0,037	0,045
10	9	0,071	0,037	0,045

Für den letzten Klassifizierer betrachten wir wiederum zuerst die vertikalen Splits

	Fehler	
Wert	$x \leq$ Wert $\Rightarrow +$	$x >$ Wert $\Rightarrow +$
1	0,363	0,637
2	0,318	0,682
3	0,484	0,516
4	0,650	0,350
5	0,544	0,456
6	0,438	0,562
7	0,604	0,396
8	0,498	0,502
9	0,543	0,457
10	0,588	0,412

und anschließend die horizontalen

	Fehler	
Wert	$y \leq$ Wert $\Rightarrow +$	$y >$ Wert $\Rightarrow +$
1	0,574	0,426
2	0,529	0,471
3	0,574	0,426
4	0,740	0,260
5	0,695	0,305
6	0,861	0,139
7	0,755	0,245
8	0,649	0,351
9	0,694	0,306
10	0,588	0,412

Der beste Split ist $y > 6 \Rightarrow +$. Berechnen wir nun das Gewicht des Klassifizierers. Es gilt

$$err_3 = 0,139$$

und damit

$$\alpha_3 = 0,912.$$

Damit haben wir die drei Klassifizierer (und deren Gewichte) des Beispiels aus der Vorlesung berechnet. Eine Illustration des resultierenden Klassifizierers befindet sich wie bereits erwähnt auf Folie 15.

b) Generieren Sie aus den eben berechneten Decision Stumps einen Entscheidungsbaum.

Lösung: Eine von mehreren möglichen Lösungen ist folgende:

x ≤ 2: +
x > 2
| y < 6: -
| y ≥ 6
| | x ≤ 8: +
| | x > 8: -

Aufgabe 2: Stacking

In dieser Aufgabe sollen Sie unter Verwendung mehrerer Basislerner und der Ensemble-Methode Stacking einen Entscheidungsbaum lernen. Verwenden Sie hierfür den Datensatz und die drei Decision Stumps aus der vorherigen Aufgabe.

a) Konvertieren Sie diesen Datensatz, d.h. ersetzen sie die Attribute durch eine neue Attributsmenge, die jeweils ein Attribut für jeden Decision Stump beinhaltet. Als Attributwerte werden die Vorhersage des entsprechenden Klassifzierers verwendet.

Lösung: In der vorherigen Aufgabe haben wir drei Klassifizierer generiert, diese verwenden wir nun als Basisklassifizierer für die Ensemble-Methode Stacking. Hierfür müssen wir zunächst den Datensatz konvertieren. Jeder Basisklassifizierer wird durch ein Attribut repräsentiert und dessen Vorhersage für eine Instanz stellt deren Attributwert dar. Das heißt wir bestimmen für jede Instanz einen Vektor, der aus den Vorhersagen des ersten, zweiten und dritten Klassifizierers und der ursprüngliche Klasse besteht. Wenden wir dies auf unseren Datensatz erhalten wir folgendes:

		C1	C2	C3	
X	Y	$x \leq 2$	$x \leq 8$	$y \geq 6$	Klasse
1	5	+	+	-	+
2	2	+	+	-	+
3	1	-	+	-	-
4	6	-	+	+	-
5	8	-	+	+	+
6	10	-	+	+	+
7	4	-	+	-	-
8	7	-	+	+	+
9	3	-	-	-	-
10	9	-	-	+	-

Die ersten beiden Zeilen stellen die ursprünglichen Attributwerte dar und können entfernt werden. Damit erhalten wir den folgenden konvertierten Datensatz:

C1	C2	C3	
$x \leq 2$	$x \leq 8$	$y \geq 6$	Klasse
+	+	-	+
+	+	-	+
-	+	-	-
-	+	+	-
-	+	+	+
-	+	+	+
-	+	-	-
-	+	+	+
-	-	-	-
-	-	+	-

b) Bestimmen Sie nun auf dem konvertierten Datensatz einen Entscheidungsbaum mittels des Verfahrens ID3 und Maß Information Gains.

Lösung: Auf ID3 im speziellen gehen wir nicht mehr ein. Für die Berechnung des maximalen Information Gains bestimmen wir nur die gewichtete Summe der Entropien und minimieren diese:

		+	-	P+	P–	$\lvert S_i\rvert/\lvert S\rvert$	$E(S_i)$	$\lvert S_i\rvert/\lvert S\rvert \cdot E(S_i)$	Total
C1	+	2	0	1,00	0,00	0,2	0,00	0,00	0,76
	-	3	5	0,38	0,63	0,8	0,95	0,76	
C2	+	5	3	0,63	0,38	0,8	0,95	0,76	0,76
	-	0	2	0,00	1,00	0,2	0,00	0,00	
C3	+	3	2	0,60	0,40	0,5	0,97	0,49	0,97
	-	2	3	0,40	0,60	0,5	0,97	0,49	

Die Attribute C1 und C2 sind beide gleichwertig, wir entscheiden uns für den ersten Test, also für C1. Wir müssen nur die Beispiele für C1=- betrachten (die Beispiele von C1=+ gehören alle zur Klasse +):

		C1	C2	C3	
X	Y	$x \leq 2$	$x \leq 8$	$y \geq 6$	Klasse
3	1	-	+	-	-
4	6	-	+	+	-
5	8	-	+	+	+
6	10	-	+	+	+
7	4	-	+	-	-
8	7	-	+	+	+
9	3	-	-	-	-
10	9	-	-	+	-

Bestimmen wir also den nächsten Test:

		+	-	P+	P–	$\lvert S_i\rvert/\lvert S\rvert$	$E(S_i)$	$\lvert S_i\rvert/\lvert S\rvert \cdot E(S_i)$	Total
C2	+	3	3	0,5	0,5	0,75	1	0,75	0,75
	-	0	2	0	1	0,25	0	0	
C3	+	3	2	0,6	0,4	0,63	0,97	0,61	0,61
	-	0	3	0	1	0,38	0	0	

Diesmal ist C3 der beste Test. Für den Ast C3=+ wird noch ein Test auf C2 angehängt, der die verbleibenden Instanzen perfekt auf die Klassen + und – aufteilt.

c) Zeichnen Sie diese Baum und vergleichen Sie ihn mit dem Entscheidungsbaum aus Aufgabe 1.

Lösung:

x ≤ 2: +
x > 2
| y < 6: -
| y ≥ 6
| | x ≤ 8: +
| | x > 8: -

Dieser Baum entspricht dem Baum aus Aufgabe 1b). Natürlich kann sich durch eine andere Wahl im Wurzelknoten ein anderer Baum ergeben, z.B:

x ≤ 8
| x ≤ 2: +
| x > 2
| | y ≥ 6: +
| | y < 6:-
x > 8: -