Data Mining und Maschinelles Lernen

Wintersemester 2015/2016

Lösungsvorschlag für das 11. Übungsblatt

TECHNISCHE UNIVERSITÄT DARMSTADT

AdaBoost (1)

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	*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.
Wählen Sie bei Gleichstand den zuerst gefundenen Test beginnend mit vertikalen Splits mit aufsteigenden Thresholds.

AdaBoost (2)

a) Berechnen Sie die ersten 3 AdaBoost-Iterationen.

Generelles Vorgehen In jeder Iteration sind jeweils 10 horizontale (y $\leq$ Wert) und 10 vertikale Splits möglich. Jeder Split ermöglicht 2 Vergleiche, z.B. y $\leq$ 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.*

AdaBoost
Erste Iteration (1)

Beginnen wir nun mit den Berechnungen. Am Anfang haben alle Beispiele das Gewicht $1/10$ (10 Beispiele). Betrachten wir nun alle vertikalen Splits (links), sowie horizontalen Splits (rechts).

Wert	Fehler	
	$x \leq$ Wert $\Rightarrow +$	$x >$ Wert $\Rightarrow +$
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

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 5 Splits mit minimalen Fehler gibt (rot markiert). Wir entscheiden uns für den zuerst gefundenen Split ($x \leq 2 \Rightarrow +$).

AdaBoost
Erste Iteration (2)

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:

$$w_i \leftarrow \begin{cases} w_i \cdot e^{-\alpha_1} \approx 0,0654, & \text{falls } w_i \text{ korrekt klassifiziert wird} \\ w_i \cdot e^{\alpha_1} \approx 0,1528, & \text{falls } w_i \text{ falsch klassifiziert wird} \end{cases}$$

Da in der ersten Iteration alle Anfangsgewichte gleich waren, sind dies auch die möglichen neuen Gewichte (mit Faktor 10).

Normalisierung: Da sieben Beispiele korrekt und drei falsch klassifiziert wurden, erhalten wir:

$$3 \cdot 0,1528 + 7 \cdot 0,0654 = 0,9162$$

als Gesamtsumme der Gewichte. Diese wollen wir auf eins normieren, aus diesem Grund teilen wir alle Gewichte durch 0,9162.

AdaBoost
Erste Iteration (3)

TECHNISCHE UNIVERSITÄT DARMSTADT

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	Klasse
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	-

Rundungsbedingt ergibt die Gesamtsumme der Gewichte den Wert 0,998.

AdaBoost
Zweite Iteration (1)

Suchen wir nun den nächsten Split.
Zur Veranschaulichung, die Berechnung des Fehlers für den ersten möglichen Stump $x \leq 1 \Rightarrow +$:

x	y	Klasse	(Fehler)	x	y	Klasse	(Fehler)
1	5	+	(0)	3	1	-	(0)
2	2	+	(0,071)	4	6	-	(0)
5	8	+	(0,167)	7	4	-	(0)
6	10	+	(0,167)	9	3	-	(0)
8	7	+	(0,167)	10	9	-	(0)

Summe der Fehlerterme: 0,572

AdaBoost
Zweite Iteration (2)

TECHNISCHE UNIVERSITÄT DARMSTADT

In Tabellenform:

Wert	Fehler $x \leq$ Wert $\Rightarrow +$	Fehler $x >$ Wert $\Rightarrow +$
1	0,572	0,426
2	0,501	0,497
3	0,572	0,426
4	0,643	0,355
5	0,476	0,522
6	0,309	0,689
7	0,38	0,618
8	0,213	0,785
9	0,284	0,714
10	0,355	0,643

Wert	Fehler $y \leq$ Wert $\Rightarrow +$	Fehler $y >$ Wert $\Rightarrow +$
1	0,714	0,284
2	0,643	0,355
3	0,714	0,284
4	0,785	0,213
5	0,714	0,284
6	0,785	0,213
7	0,618	0,380
8	0,451	0,547
9	0,522	0,476
10	0,355	0,643

Den Tabellen entnehmen wir, daß 3 mögliche Vergleiche optimal sind. Wiederum entschieden wir uns für den zuerst gefundenen Split, $x \leq 8 \Rightarrow +$. Punkte (3, 1), (4, 6) und (7, 4) werden falsch und alle anderen richtig klassifiziert.

AdaBoost
Zweite Iteration (3)

TECHNISCHE UNIVERSITÄT DARMSTADT

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 Rechenhilfe-Tabelle:

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

AdaBoost
Zweite Iteration (4)

TECHNISCHE
UNIVERSITÄT
DARMSTADT

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

			Neues Gewicht		
x	y	Altes Gewicht	Nicht normiert	Normiert	Klasse
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	-

Erneut ist zu beachten, daß aufgrund der Rundung die Gesamtsumme den Wert 0,996 hat.

AdaBoost
Dritte Iteration (1)

TECHNISCHE UNIVERSITÄT DARMSTADT

Für den letzten Klassifizierer nun wiederum die Splits:

	Fehler	
Wert	$x \leq$ Wert $\Rightarrow +$	$x >$ Wert $\Rightarrow +$
1	0,363	0,633
2	0,318	0,682
3	0,484	0,512
4	0,650	0,346
5	0,544	0,452
6	0,438	0,558
7	0,604	0,392
8	0,498	0,498
9	0,543	0,453
10	0,588	0,408

	Fehler	
Wert	$y \leq$ Wert $\Rightarrow +$	$y >$ Wert $\Rightarrow +$
1	0,574	0,422
2	0,529	0,467
3	0,574	0,422
4	0,740	0,256
5	0,695	0,301
6	0,861	0,135
7	0,755	0,241
8	0,649	0,347
9	0,694	0,302
10	0,588	0,408

Der beste Split ist $y > 6 \Rightarrow +$.

AdaBoost
Dritte Iteration (2)

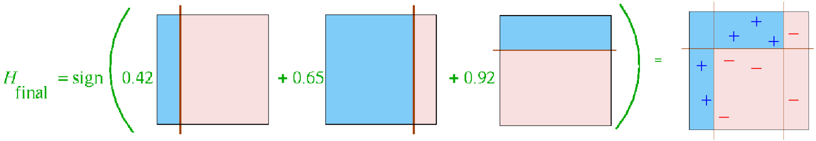
Berechnen wir nun das Gewicht des Klassifizierers. Es gilt

$$err_3 = 0,135$$

und damit

$$\alpha_3 = 0,929.$$

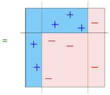
Finales Modell: Damit haben wir die drei Klassifizierer (und deren Gewichte) des Beispiels aus der Vorlesung berechnet! Eine Illustration des resultierenden Klassifizierers befindet sich auf Folie "*Final Hypothesis*".



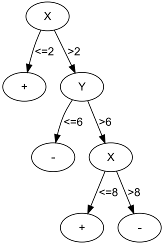
AdaBoost
Darstellung (1)

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

Lösung: Basierend auf der finalen Hypothese



ist eine von mehreren möglichen Lösungen folgende:



Stacking (1)

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 Vorhersagen des entsprechenden Klassifizierers verwendet.

Stacking
Datensatz (1)

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 ursprünglichen Klasse besteht.
Wenden wir dies auf unseren Datensatz an, erhalten wir folgendes:

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

Stacking
Datensatz (2)

Die ersten beiden Zeilen stellen die ursprünglichen Attributwerte dar und können (bzw., im üblichen Szenario, sollen!) entfernt werden. Damit erhalten wir den folgenden konvertierten Datensatz:

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

Stacking
Klassifizierer (1)

b) Bestimmen Sie nun auf dem konvertierten Datensatz einen Entscheidungsbaum mittels des Verfahrens ID3 und Maß Information Gains. Entscheiden Sie sich bei Gleichstand z.B. für den als erstes gefundenen Test.

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 Gain
C1	+	2	0	1,00	0,00	0,2	0,00	0,00	0,24
	-	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	0,24
	-	0	2	0,00	1,00	0,2	0,00	0,00	
C3	+	3	1	0,75	0,25	0,4	0,81	0,32	0,13
	-	2	4	0,33	0,67	0,6	0,92	0,55	

Die Attribute C1 und C2 sind beide gleichwertig, wir entscheiden uns für den ersten Test, also für C1.

Stacking
Klassifizierer (2)

TECHNISCHE UNIVERSITÄT DARMSTADT

Wir müssen nur die Beispiele für C1=- betrachten (die Beispiele von C1=+ gehören alle zur Klasse +):

		C1	C2	C3	
X	Y	$x \leq 2$	$x \leq 8$	$y > 6$	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 Gain
S		3	5	0,38	0,63	-	0,95	-	-
C2	+	3	3	0,5	0,5	0,75	1	0,75	0,2
	-	0	2	0	1	0,25	0	0	
C3	+	3	1	0,75	0,25	0,5	0,81	0,41	0,54
	-	0	4	0	1	0,5	0	0	

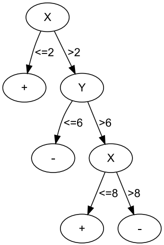
Stacking
Klassifizierer (3)

		+	-	P+	P–	$\lvert S_i\rvert / \lvert S\rvert$	$E(S_i)$	$\lvert S_i\rvert / \lvert S\rvert \cdot E(S_i)$	Total Gain
S		3	5	0,38	0,63	-	0,95	-	-
C2	+	3	3	0,5	0,5	0,75	1	0,75	0,2
	-	0	2	0	1	0,25	0	0	
C3	+	3	1	0,75	0,25	0,5	0,81	0,41	0,54
	-	0	4	0	1	0,5	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.

Stacking Klassifizierer (4)

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



Dieser Baum entspricht dem Baum aus Aufgabe 1b).

Stacking Klassifizierer (5)

Natürlich kann sich durch eine andere Wahl im Wurzelknoten ein anderer Baum ergeben, z.B: