Übungen zur Vorlesung
Grundlagen der Künstlichen Intelligenz
Sommersemester 2024

Institut für Informatik 4
Priv.-Doz. Dr. V. Steinhage
Friedrich-Hirzebruch-Allee 8
53115 Bonn
Email: steinhage@cs.uni-bonn.de
WWW: http://net.cs.uni-bonn.de/ivs/

Blatt 8 (10 Punkte + 1 Bonuspunkt)

Abgabe durch Hochladen (nur PDF-Format bzw. Python-Code) auf der eCampus-Seite bis **Sonntag, 09.06.2024, 12:00 Uhr**, in Gruppen von 3 Personen.

Beachten Sie, dass von Aufgaben 8.1 und 8.2 nur eine auszusuchen ist, die abgegeben werden kann/muss. Es wird dementsprechend nur eine Aufgabe korrigiert und bepunktet. Bei Abgabe beider Aufgaben bitte kenntlich machen, welche bewertet werden soll - ansonsten wird Aufgabe 8.1 bewertet.

Aufgabe 8.1: Entscheidungsbäume: Attributsauswahl (entweder hier $2 + 1 + 0 = 3$)

Gegeben sei folgende Trainingsmenge über acht Trainingsbeispielen, die jeweils über die drei Booleschen Attributen A, B und C beschrieben sind, mit ihren Booleschen Antwortwerten.

A	B	C	Antwort
Nein	Nein	Nein	Nein
Nein	Nein	Ja	Nein
Nein	Ja	Nein	Ja
Nein	Ja	Ja	Nein
Ja	Nein	Nein	Ja
Ja	Nein	Ja	Ja
Ja	Ja	Nein	Ja
Ja	Ja	Ja	Nein

a) Erstellen Sie einen vollständigen Entscheidungsbaum nach der in Vorlesung 13 beschriebenen Methode des Decision-Tree-Learning mit schrittweiser Auswahl der Attribute nach größtem Informationsgewinn. Gehen Sie davon aus, dass im ersten Schritt das Attribut A den größten Gewinn (GAIN) hat und somit als erstes Attribut ausgewählt wird. Fahren Sie ausgehend von diesem Punkt fort, den Algorithmus anzuwenden, indem Sie für *jeden* zu bestimmenden Entscheidungsknoten explizit die Gain-Werte für die möglichen Attribute rechnen.

b) Verwenden Sie nun nur die beiden kombinierten Attribute (B = JA $\wedge$ C = JA) und (A = NEIN $\wedge$ B = NEIN) und geben Sie einen optimalen Entscheidungsbaum an, der *ausschließlich* auf diesen beiden kombinierten Attributen basiert. (Der Baum beginnt also nicht mit der Wahl von A als erstem Attribut wie in Aufgabenteil a).) Sie müssen hier die Gain-Werte für die möglichen Attributauswahlen nicht wieder explizit berechnen.

c) Diskutieren Sie **in der Übung**, was dieses Vorgehen mit kombinierten Attributen bedeutet, sowie die Vor- und Nachteile davon.

Aufgabe 8.2: Programmieraufgabe: Entscheidungsbäume (oder hier 3)

Vorbereitung: Laden Sie bitte das ZIP-Archiv decision.zip herunter von unserer eCampus-Seite unter Kursunterlagen >> Python und AIMA Python >> Decision Trees. Das ZIP-Archiv enthält den Ordner *decision* mit dem Skript *decisionS.py* sowie dem vorgegebenen Restaurant-Trainingsdatensatz *restaurant.feat*. Fügen Sie den Ordner *decision* auf der gleichen Ebene wie den Ordner *aima* ein.

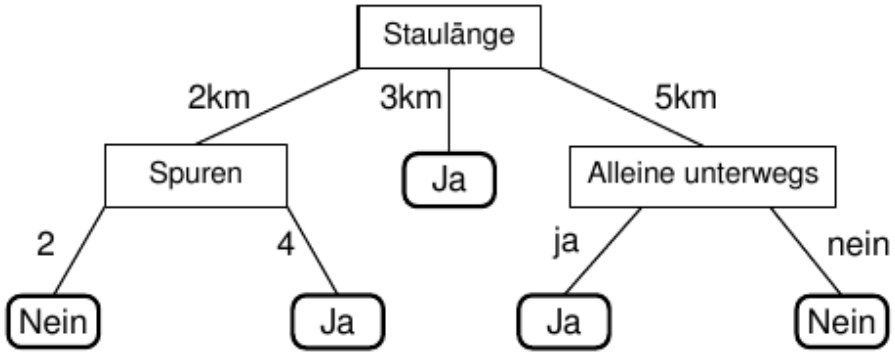
Aufgabe: Der Trainingsalgorithmus im vorgegeben Skript *decisionS.py* wählt in der Funktion *choose_attribute* die Attribute auf triviale Weise aus. Bitte erstellen Sie auf der Basis des gegebenen Skripts *decisionS.py* ein neues Skript *decision.py*, welches die Auswahl der Attribute nach Gain und Remainder umsetzt. Bitte erzeugen Sie mit Hilfe Ihres neu erstellten Skripts *decision.py* den Entscheidungsbaum für das Restaurantbeispiel auf der Basis des Restaurant-Trainingsdatensatzes *restaurant.feat* und reichen Sie auf jeden Fall einen Screenshot des erzeugten Entscheidungsbaums ein. Eine Visualisierung des Entscheidungsbaums ist über *Show Decisiontree* der AIMA-Py-GUI erzeugbar.

Aufgabe 8.3: Entscheidungsbäume und -listen $(1(\text{Bonus}) + 1)$

Instanz	Alleine unterwegs	Staulänge	Spuren	Tank voll	Abfahrt ?
#1	JA	5 KM	2	NEIN	JA
#2	NEIN	3 KM	3	NEIN	JA
#3	JA	3 KM	2	NEIN	JA
#4	JA	2 KM	2	NEIN	NEIN
#5	NEIN	5 KM	2	JA	NEIN
#6	JA	2 KM	4	NEIN	JA
#7	JA	2 KM	2	NEIN	NEIN

Stellen Sie sich vor, Sie fahren gerade nach Köln und hören im Radio, dass kurz vor Ihnen ein Stau ist. Wie entscheiden Sie sich, ob Sie von der Autobahn abfahren, um den Stau zu umgehen? In der obigen Tabelle sind einige Attribute aufgeführt, die Ihre Entscheidung beeinflussen könnten, und einige Beispieldaten von verschiedenen (fiktiven) Autofahrern.

Der resultierende Entscheidungsbaum habe folgende Gestalt:



a) Was passiert, wenn Sie Ihren Entscheidungsbaum auf ein Problem anwenden, bei dem die Staulänge 2 km beträgt und es 3 Spuren gibt? Tip: Schauen Sie sich dazu noch einmal genau die vier Fälle an, die im Algm. DECISION-TREE-LEARNING behandelt werden.

b) Konstruieren Sie nun eine Entscheidungsliste für das oben gegebene Problem mit möglichst kurzen Tests (also möglichst kleinem k bzgl. der Sprachklasse k-DL). Verwenden Sie für Ihre Lösung *Tank voll* als ersten Test. Um Aufzählungen zu umgehen, berücksichtigen Sie auch Tests von Attributwerten über Relationen wie „$<$“, „$>$“, „$\leq$ “ und „$\geq$“ mit Hilfe eines Schwellwertes (s. Vorl 14, Folie 23) wie z.B. „$\mathit{Spuren} > 1$“.

Aufgabe 8.4: Bewertung von Lernalgorithmen $(\mathbf{1 + 1 + 0,5 = 2,5})$

Für die Validierung eines trainierten Entscheidungsbaumes mit Booleschem Zielprädikat *infiziert* stehen 8 Stichproben von Ein-/Ausgabepaaren zur Verfügung. Die Stichproben $S_{11}, ..., S_{14}$ gehören der positiven Zielklasse 1 (*infiziert*) an. Die Stichproben $S_{21}, ..., S_{24}$ gehören der komplementären negativen Klasse 2 (*nicht infiziert*) an.
Der trainierte Entscheidungsbaum prädiziert beim Testen wie folgt: $S_{11}, S_{12}, S_{13}, S_{21}, S_{22}$ werden der Klasse 1 zugeordnet. Die restlichen Stichproben werden der Klasse 2 zugeordnet.

a) Erstellen Sie die Konfusionsmatrix.

b) Berechnen Sie mit Herleitung *P*recision und *R*ecall .

c) Der Entscheidungsbaum wird geändert, so dass S_{22} jetzt der Klasse 2 zugeordnet wird - alle andere Prädiktionen bleiben wie zuvor. Wird die *Precision* oder der *Recall* dadurch erhöht? Begründen Sie Ihre Antwort.

Aufgabe 8.5: AdaBoost **$(1 + 2 + 0,5 = 3,5)$**

Anhand der zwei Attribute A_1:*"linkes Knie juckt"* und A_2:*"Vögel fliegen tief"* soll entschieden werden, ob es regnet oder nicht. Dazu soll hier mittels des Algorithmus `AdaBoost` (Vorlesung 14) ein Ensemble von nur zwei gewichteten Entscheidungsstümpfen über genau diesen beiden Attributen gelernt werden.
Da keines der beiden Attribute alleine als Testknoten in einem Entscheidungsstumpf alle acht Beispiele eindeutig einordnen kann, ist anschaulich der Mehrheitsentscheid im Fall 3 vom Entscheidungsbaumlernen (Folie 25, Vorl. 13) anzuwenden.
In diesen Mehrheitsentscheid müssen in `AdaBoost` aber die Gewichte der Beispiele einfließen. Daher besteht das Hypothesen-Lernverfahren L(examples,w) jetzt darin, für jedes der beiden Attribute $A_m, m \in \{1, 2\}$, die Hypothese $h_m \in \{+1, -1\}$ zu lernen, die den gewichteten Gesamtfehler $e_m = 0.5 \cdot \sum_{j=1}^{8} w[j] \cdot |h_m \cdot I_j(A_m) - I_j(R)|$ über der Trainingsmenge minimiert. Danach sollte also gelten: $h_m = +1$, wenn es eher regnet, wenn A_m zutrifft; $h_m = -1$, wenn es eher regnet, wenn A_m nicht zutrifft.

Gegeben sind folg. Trainingsbeispiele:

Instanz	A_1	A_2	R egen
I_1	1	1	1
I_2	-1	1	1
I_3	1	-1	1
I_4	-1	1	1
I_5	1	-1	-1
I_6	-1	1	-1
I_7	1	1	-1
I_8	1	-1	-1

a) 1) Berechnen Sie für jede der zwei Iterationen zunächst für die beiden möglichen Hypothesen $h_m \in \{+1, -1\}$ den jeweiligen gewichteten Fehler e_m und nur für die den Fehler minimierende Hypothese dann das Hypothesengewicht z_m zum Attribut A_m. Verwenden Sie den natürlichen Logarithmus für log im Algorithmus `AdaBoost`.

Iteration m	h_m	e_m	z_m
#1	+		
#1	-		
#2	+		
#2	-		

Hinweis: Bearbeiten Sie Teile 1 und 2 zusammen, da die Einträge in der ersten Tabelle teilweise von den Einträgen in zweiten Tabelle abhängen.

2) Notieren Sie nach jeder Iteration m die neue Gewichtsverteilung über den Trainingsinstanzen w_m und den absoluten Fehler E_m des Ensembles auf der Trainingsmenge.

$$E_m = 0.5 \cdot \sum_{j=1}^{8} \left| \overbrace{\operatorname{sgn}\left[\sum_{i=1}^{m} z_i \cdot h_i \cdot I_j(A_i)\right]}^{\text{Klassifikator Ergebnis}} - \overbrace{I_j(R)}^{\text{ground truth}} \right|$$

	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	E_m
$w_0[I_j]$	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	-
$w_1[I_j]$ vor Normalisierung									
$w_1[I_j]$ nach Normalisierung									
$w_2[I_j]$ vor Normalisierung									
$w_2[I_j]$ nach Normalisierung									

b) Würde es laut dem gelernten Ensemble-Klassifizierer bei der Testinstanz $(-1, -1)$ regnen oder trocken bleiben?