

Formale Systeme, Wintersemester 2020/21 Übung 5

Aufgabe 1

Es werde mit dem Hypothesenraum aus Def. 5.1 des Vorlesungsskripts (S. 128) gearbeitet.

Gegeben sind die folgenden Trainingsdaten:

#	Himmel	Lufttemperatur	Luftfeuchtigkeit	Wind	Wasser	Vorhersage	Sport_macht_Spaß
1	sonnig	warm	hoch	stark	kalt	wechselhaft	ja
2	regnerisch	kalt	hoch	stark	warm	wechselhaft	nein
3	sonnig	warm	hoch	stark	warm	gleichbleibend	ja
4	sonnig	warm	normal	stark	warm	gleichbleibend	ja

Das Attribut Sport_macht_Spaß hängt von verschiedenen Umweltfaktoren ab. Insbesondere gibt es für das Attribut Himmel drei, für Lufttemperatur, Luftfeuchtigkeit, Wind, Wasser und Vorhersage jeweils zwei verschiedene mögliche Werte.

Erklären Sie, warum die Anzahl der semantisch verschiedenen möglichen Hypothesen für das Attribut Sport_macht_Spaß 973 beträgt!

Angenommen, es gäbe ein weiteres Attribut Wasserströmung, welches drei verschiedene Werte annehmen kann, wie groß wären dann die Anzahl der möglichen Instanzen und der (syntaktisch und semantisch verschiedenen) möglichen Hypothesen ?

Wie würden diese Werte sich ändern, wenn ganz allgemein ein Attribut A mit k verschiedenen Werten hinzukäme ?

Aufgabe 2

Gegeben sind die folgenden Trainingsdaten:

Origin	Manufacturer	Color	Decade	Type	Example Type
Japan	Honda	Blue	1980	Economy	Positive
Japan	Toyota	Green	1970	Sports	Negative
Japan	Toyota	Blue	1990	Economy	Positive
USA	Chrysler	Red	1980	Economy	Negative
Japan	Honda	White	1980	Economy	Positive

Führen Sie den Algorithmus der Candidate-Elimination mit den Trainingsdaten in der angegebenen Reihenfolge durch und geben Sie die dabei erzeugte Sequenz der Begrenzungsmengen S und G an.

Aufgabe 3

Geben Sie zu folgenden Booleschen Funktionen einen Entscheidungsbaum an:

- (a) $A \text{ xor } B$ ("xor" = exklusives Oder)
- (b) $(A \wedge B) \vee (C \wedge D)$

Aufgabe 4

Man betrachte folgende Menge von Trainingsbeispielen:

Instanz	a_1	a_2	Klassifikation
1	W	W	+
2	W	W	+
3	W	F	-
4	F	F	+
5	F	W	-
6	F	W	-

Wie groß ist die Entropie der Trainingsbeispiele in Bezug auf die Klassifikation als Zielfunktion? Wie groß ist der Informationsgewinn (*information gain*) von a_2 relativ zu diesen Trainingsbeispielen?