

Proseminar Maschinelles Lernen und Experimentelles Design, UdS, SS11 Hausaufgaben 21.4.2011

Caroline Sporleder

April 20, 2011

Aufgabe 1 Max wohnt in Saarbrücken und ißt regelmäßig Eis. Ihr möchtet herausfinden, welche Faktoren Max beeinflussen, sich ein Eis zu kaufen. Die Beobachtungen, die ihr macht und die korrekte Targethypothese (c) (die ihr noch lernen müßt) sind:

	Regen-Mannh.	Sonne-SB	WE	geschl.	glüickl.	kein-Geld	Eis?
c	?	T	?	F	?	F	T
1	T	T	T	F	T	F	T
2	T	T	T	F	T	T	F
3	T	T	F	F	T	F	T
4	F	F	F	F	F	F	F
5	T	T	T	F	F	F	T
6	T	F	F	F	F	F	F
7	F	F	T	T	T	T	F

- T = true
- F = false
- Regen-Mannh = es regnet in Mannheim
- Sonne-SB = in SB scheint die Sonne
- WE = es ist Wochenende
- geschl. = die Eiscafé's sind geschlossen
- glüickl. = Max ist glücklich
- kein-Geld = Max hat kein Geld mehr

Wendet Find-S an und berechnet die spezifischste Hypothese, h , für die Trainingsbeispiele.

Welchen Ausgabewert sagt h für die folgenden neuen Beispiele vorher? Welchen sagt die korrekte Targethypothese, c , vorher?

1. $\langle TTTFTF \rangle$
2. $\langle FTFFFF \rangle$
3. $\langle FTFFTF \rangle$
4. $\langle TTTTTT \rangle$
5. $\langle TTFTF F \rangle$

Aufgabe 2 Zeichnet den partiell geordneten Hypothesenraum für zwei Attribute mit den möglichen Werten *true* und *false* (z.B. $\langle TT \rangle$, $\langle T? \rangle$, $\langle F? \rangle$). Berechnet den Version Space nachdem der Lerner das negative Beispiel $\langle TT \rangle$ gesehen hat. Markiert alle Hypothesen, die noch im Version Space vorhanden sind in rot. Berechnet dann den Version Space nachdem der Lerner auch das positive Beispiel $\langle FT \rangle$ gesehen hat, und markiert die verbleibenden Hypothesen in grün.

Zeichnet den Hypothesenraum noch einmal und berechnet wieder die Version Spaces, diesmal aber mit der umgekehrten Beispielfolge, d.h. zuerst das positive Beispiel $\langle FT \rangle$, dann das negative Beispiel $\langle TT \rangle$.

Was sagt euch das Ergebnis darüber, wie die Beispiele am besten geordnet werden sollten, wenn man die Version Spaces berechnet?

Aufgabe 3 Wendet Candidate-Elimination auf das Trainingsset aus Aufgabe 1 an (um es einfacher zu machen, ordnet die Trainingsbeispiele so um, dass die positiven zuerst kommen). Zeigt, dass der Version Space am Ende wie folgt aussieht:

- S: $\langle TT?F?F \rangle$
- G: $\langle ?T???F \rangle$

Berechnet die Ausgabewerte der ungesehenen Beispiele aus Aufgabe 1, indem ihr alle Hypothesen des Version Space anwendet. Falls die Hypothesen für eine Instanz verschiedene Ausgaben geben, dann wählt den Ausgabewert durch Mehrheitsentscheidung über alle Hypothesen in Version Space. Vergleicht die Ergebnisse mit denen, die ihr mit dem Find-S Algorithmus bekommen habt (Aufgabe 1).