

Maschinelles Lernen und Data Mining

1. Übungsblatt

10. 11. 2005

Aufgabe 1

Gegeben sei ein Beispielraum, der durch n binäre Attribute aufgespannt wird. Hypothesenraum sind alle möglichen Regeln, die sich durch Konjunktionen von Tests der Form *Attribut = Wert* ergeben.

1. Wie viele mögliche Konzepte gibt es?
2. Wie viele mögliche Regeln gibt es?
3. Unter der Voraussetzung, daß das zu lernende Konzept im Hypothesenraum darstellbar ist, wie viele Fehler kann der Algorithmus Find-S beim Lernen dieses Konzepts maximal machen?

[**Anm:** Ein Fehler ist hier ein (Trainings-)Beispiel, das von der momentanen Theorie falsch klassifiziert wird.]

4. Gegeben sei ein Hypothesenraum, der nur *Disjunktionen* von Attribut-Wert-Paaren erlaubt. Eine gültige Regel wäre also z.B.

if ($att_i = t$) **or** ($att_j = f$) **or** ($att_k = f$) **then** +

Überlegen Sie sich Verallgemeinerungs- und Spezialisierungsvorschriften für diesen Hypothesenraum und geben Sie einen geeigneten Lernalgorithmus an.

Aufgabe 2

Sie wollen den Find-S Algorithmus auf numerische Daten anwenden, indem Sie Intervalle definieren.

1. Wie sieht das spezifischste und generellste Element der Sprache aus, wenn Sie offene Intervalle verwenden?
2. Wie sieht das spezifischste und generellste Element der Sprache aus, wenn Sie geschlossene Intervalle verwenden?

3. Finden Sie eine passende Generalisierungsvorschrift und simulieren den Algorithmus Find-S auf folgenden Beispielen:

A1	Klasse
0.5	–
1.0	+
2.1	–
0.8	–
1.5	+
1.8	+

4. Finden Sie eine passende Spezialisierungsvorschrift und simulieren Sie den Algorithmus Find-G auf denselben Beispielen. Vergleichen Sie die Allgemeinheit der gefundenen Lösungen.