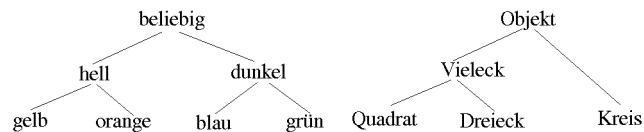


Maschinelles Lernen und Data Mining

Übungsblatt für den 17.11.2005

Aufgabe 1

Gegeben sei folgende Hierarchie von Begriffen:



Beobachtet werden Objekte, die durch Begriffspaare charakterisiert werden, die man an der untersten Ebene dieser Taxonomien finden kann (also z.B. “blaues Dreieck”). Konzepte können auch höherliegende Begriffe verwenden (also z.B. “dunkles Vieleck”). Überlegen Sie sich eine Generalisierungsvorschrift, die diese Taxonomien verwendet.

1. Wie sieht die minimale Generalisierung der Objekte “blauer Kreis” und “grünes Dreieck” aus?
2. Wie sehen minimale Spezialisierungen des Konzepts “helles Objekt” aus, sodaß das Beispiel “oranger Kreis” nicht mehr abgedeckt wird?
3. Gegeben seien folgende S und G-Sets:
G: { dunkles Vieleck, beliebiges Quadrat }
S: { blaues Quadrat }

Skizzieren Sie den Version Space, der durch diese Mengen definiert wird.

4. Wie würden Sie mit Hilfe des oben gegebenen Version Spaces die folgenden Beispiele klassifizieren (mit Begründung):

| Objekt | Klasse |
|----------------|---------------|
| blaues Quadrat | |
| blauer Kreis | |
| blaues Dreieck | |

5. Gegeben seien wiederum die S- und G-sets aus Punkt 3. Wie verändern sich die Sets nach Eintreffen des Beispiels

gelbes Dreieck +

Wie interpretieren Sie dieses Ergebnis?

Aufgabe 2

Überlegen Sie sich eine geeignete Sprache, um den Candidate Elimination Algorithmus um die Behandlung von numerischen Daten zu erweitern.

1. Wie sieht eine passende Generalisierungs/Spezialisierungsvorschrift aus?
2. Berechnen Sie den Version Space für folgende Beispiele:

| Nr. | A1 | A2 | Klasse |
|-----|-----|-----|--------|
| 1 | 0.5 | 1.5 | – |
| 2 | 1.1 | 1.2 | + |
| 3 | 1.8 | 1.0 | + |
| 4 | 1.5 | 2.1 | – |
| 5 | 2.1 | 1.2 | – |

3. Skizzieren Sie das S-Set, das G-Set, und den Version Space im \mathbb{R}^2 .
4. Vertauschen Sie die Rolle der positiven und negativen Beispiele. Was passiert dann?