

Maschinelles Lernen: Symbolische Ansätze



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Wintersemester 2008/2009

1. Übungsblatt für den 28.10.2008

Aufgabe 1: Anwendungsszenario

Überlegen Sie sich ein neues Szenario des klassifizierenden Lernens (kein aus der Vorlesung bekanntes).

- Bestimmen Sie die zu verwendenden Trainings- und Testdaten Ihres Klassifikationsproblems.
- Aus welchen Typen von Attributen (nominal, numerisch, ...) setzen sich die Beispiele zusammen?
- Welche Kriterien würden Sie verwenden, um die Performanz des resultierenden Klassifizierers zu bewerten? Bedenken Sie bei Ihren Überlegungen, dass die Performanz abhängig von dem gewählten Problem ist (bei der Klassifizierung von Spam Mail ist es beispielsweise wichtig, echte Mails nicht als Spam einzuordnen).

Aufgabe 2: Praktische Anwendung

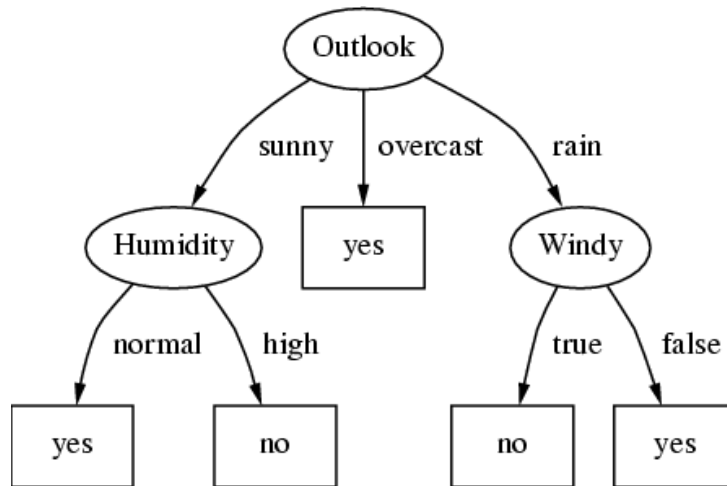
Betrachten wir nochmals den Beispieldatensatz aus der Vorlesung.

Temperature	Outlook	Humidity	Windy	PlayGolf?
hot	sunny	high	false	no
hot	sunny	high	true	no
hot	overcast	high	false	yes
cool	rain	normal	false	yes
cool	overcast	normal	true	yes
mild	sunny	high	false	no
cool	sunny	normal	false	yes
mild	rain	normal	false	yes
mild	sunny	normal	true	yes
mild	overcast	high	true	yes
hot	overcast	normal	false	yes
mild	rain	high	true	no
cool	rain	normal	true	no
mild	rain	high	false	yes

- Klassifizieren Sie die folgende Testmenge, deren Klassenlabel uns bereits bekannt sind, mit Hilfe des abgebildeten Entscheidungsbaums.

Temperature	Outlook	Humidity	Windy	PlayGolf?
hot	rain	high	true	yes
mild	sunny	normal	false	no
hot	rain	normal	false	yes
cool	overcast	high	true	yes
mild	rain	normal	true	no

Errechnen Sie die Genauigkeit des Klassifizierers auf der Testmenge (die korrekt klassifizierten Beispiele geteilt durch die Gesamtzahl der Beispiele). Wie schätzen Sie die Güte des Klassifizierers ein? Begründen Sie Ihre Aussage.



- b) Klassifizieren Sie nun dieselbe Testmenge mit dem Lernalgorithmus Nearest Neighbour aus der Vorlesung. Verwenden Sie als Distanzfunktion die Anzahl der Attributwerte, in denen sich die zu vergleichenden Beispiele unterscheiden. Bestimmen Sie alle Trainingsbeispiele mit minimaler Distanz zum jeweiligen Testbeispiel. Sagen Sie anhand der Klassenlabel dieser Trainingsbeispiele die Klasse des Testbeispiels voraus.