

Maschinelles Lernen: Symbolische Ansätze



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Wintersemester 2008/2009

6. Übungsblatt für den 4.12.2008

Aufgabe 1: Evaluierungsmethoden

Ein Datenset enthält $2 \times n$ Beispiele, wobei genau n Beispiele positiv sind und n Beispiele negativ sind. Der einfache Algorithmus `ZeroRule` betrachtet nur die Klassenverteilung der Trainings-Daten und sagt für alle Beispiele die Klasse $+$ voraus, wenn mehr positive als negative Beispiele in den Trainings-Daten enthalten sind, und die Klasse $-$ falls es umgekehrt ist. Bei Gleichverteilung entscheidet er sich zufällig für eine der beiden Klassen, die er dann immer vorhersagt.

- Wie groß ist die Genauigkeit dieses Klassifizierers, wenn die Verteilung der Trainings-Daten der Gesamt-Verteilung entspricht (d.h., wenn die Trainings-Daten repräsentativ sind)?
- Schätzen Sie die Genauigkeit von `ZeroRule` mittels Leave-One-Out Cross-Validation ab.

Aufgabe 2: Vorzeichentest

Sie vergleichen zwei Algorithmen A und B auf 20 Datensets und beobachten folgende Genauigkeitswerte:

Datenset	1	2	3	4	5	6	7	8	9	10
Algorithm A	0,91	0,86	0,93	0,74	0,65	0,91	0,87	0,95	0,78	0,86
Algorithm B	0,94	0,80	0,96	0,88	0,84	0,94	0,97	0,67	0,86	0,89
Datenset	11	12	13	14	15	16	17	18	19	20
Algorithm A	0,98	0,96	0,74	0,53	0,95	0,67	0,98	0,96	0,97	0,91
Algorithm B	0,87	0,90	0,79	0,51	0,96	0,69	0,79	0,98	0,98	0,76

Läßt sich mit Hilfe des Vorzeichentests nachweisen, ob einer der beiden Algorithmen A oder B signifikant besser ist als der andere? Folgt daraus, daß er nicht besser ist?