

Grundlagen der Künstlichen Intelligenz

Prof. Dr. B. Nebel, Prof. Dr. M. Riedmiller
S. Lange, J. Witkowski, D. Zhang
Sommersemester 2010

Universität Freiburg
Institut für Informatik

Übungsblatt 4

Abgabe: Dienstag, 1. Juni 2010

Aufgabe 4.1 (CBR)

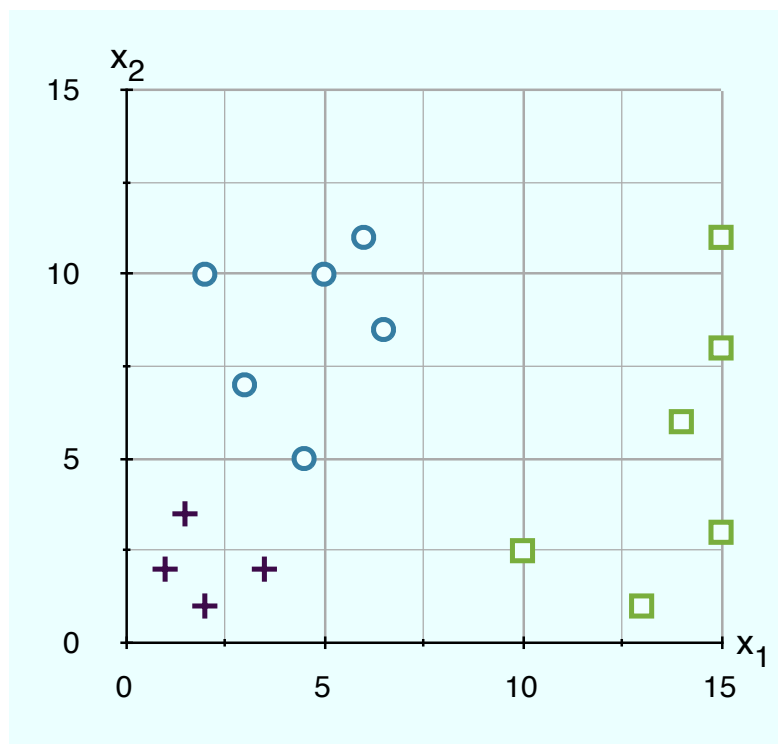
Lösen Sie folgende Teilaufgaben auf Basis der abgebildeten Fallbasis mit 16 Instanzen aus drei verschiedenen Klassen, wobei jede Instanz zwei reellwertige Attribute (x_1, x_2) besitzt.

In der Abbildung dargestellte Instanzen:

Klasse A: (1, 2), (1.5, 3.5), (2, 1), (3.5, 2)

Klasse B: (2, 10), (3, 7), (4.5, 5), (5, 10), (6, 11), (6.5, 8.5)

Klasse C: (10, 2.5), (13, 1), (14, 6), (15, 3), (15, 8), (15, 11)



- (a) Konstruieren Sie die Entscheidungsgrenzen zwischen den drei Klassen A (Kreuze), B (Kreise) und C (Quadrate) für einen 1-NN Klassifikator auf Basis des euklidischen Abstands. Benutzen Sie einen Ausdruck oder eine Kopie des obigen Diagramms für Ihre Zeichnung.

- (b) Die Verwendung von kd-Bäumen als Indexstruktur erlaubt einen effizienteren Zugriff (retrieval) auf eine Fallbasis. Im Gegensatz zu einfachen binären Suchbäumen können mittels kd-Bäumen mehr als eine Dimension zur Unterteilung des Datenraumes verwendet werden. Deshalb muss bei der Konstruktion eines kd-Baums entschieden werden, entlang welcher Dimension (welchem Attribut) und bei welchem Wert dieses Attributs eine Aufteilung vorgenommen werden soll. Welches Attribut ist bei der ersten Aufteilung der gegebenen Fallbasis nützlich? Belegen Sie Ihre Aussage an Hand einer Berechnung der Interquartilabstände beider Dimensionen.
- (c) Bei einer Aufteilung durch den Median (median splitting) wird die bearbeitete Menge in zwei Teilmengen annähernd gleicher Größe aufgeteilt. Die Teilmenge der Fälle "unterhalb des Medians" (\leq Medianelement) darf maximal ein Element mehr enthalten als die Teilmenge der Fälle "oberhalb des Medians" ($>$ Medianelement). Konstruieren Sie eine Partitionierung (Eimergröße 2) für die gegebene Fallbasis unter Verwendung der Median-Aufteilungsmethode, wobei die erste Aufteilung entlang Dimension 1 vorgenommen und danach alterniert wird. Verwenden Sie eine Kopie oder einen Ausdruck des obigen Diagramms für Ihre Zeichnung und nummerieren Sie die vorgenommenen Aufteilungsschritte.

Aufgabe 4.2 (Multivariate Verteilung)

Gegeben die multivariate Verteilung entsprechend der Tabelle

	A	$\neg A$
B	0.4	0.2
$\neg B$	0.1	0.3

berechnen Sie die folgenden Wahrscheinlichkeiten, wobei die Zelle A,B die Wahrscheinlichkeit für $P(A \wedge B)^1 = 0.4$ spezifiziert:

- (a) $P(A)$, $P(B)$, $P(\neg A)$ und $P(\neg B)$
- (b) $P(A \vee B)$ und $P((A \vee B) \wedge \neg(A \wedge B))$
- (c) $P(A|B)$ und $P(B|A)$

Aufgabe 4.3 (A-posteriori Wahrscheinlichkeit)

Nach Ihrer jährlichen Vorsorgeuntersuchung hat der Arzt gute und schlechte Neuigkeiten. Die schlechte Nachricht ist, dass der Test für eine ernste Krankheit positiv war und dass dieser Test zu 99% zuverlässig ist (d.h., die Wahrscheinlichkeit bei einer tatsächlichen Erkrankung positiv getestet zu werden beträgt 0.99, genau wie die Wahrscheinlichkeit negativ getestet zu werden, wenn man nicht erkrankt ist). Die gute Nachricht ist, dass diese Krankheit extrem selten auftritt und nur eine in 10 000 Personen trifft. Warum ist es eine gute Nachricht, dass die Krankheit so selten auftritt? Was ist die Wahrscheinlichkeit, dass Sie die Krankheit tatsächlich haben?

Die Übungsblätter dürfen und sollten in Gruppen von drei (3) Studenten bearbeitet werden. Bitte füllen Sie das Deckblatt² aus und heften Sie es an Ihre Lösung.

¹Kurzschreibweise für $P(X_1 = A \text{ and } X_2 = B)$

²<http://www.informatik.uni-freiburg.de/~ki/teaching/ss10/gki/coverSheet-german.pdf>