Foundations of Artificial Intelligence

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

University of Freiburg
Department of Computer Science

Exercise Sheet 4

Due: Tuesday, June 1, 2010

Exercise 4.1 (CBR)

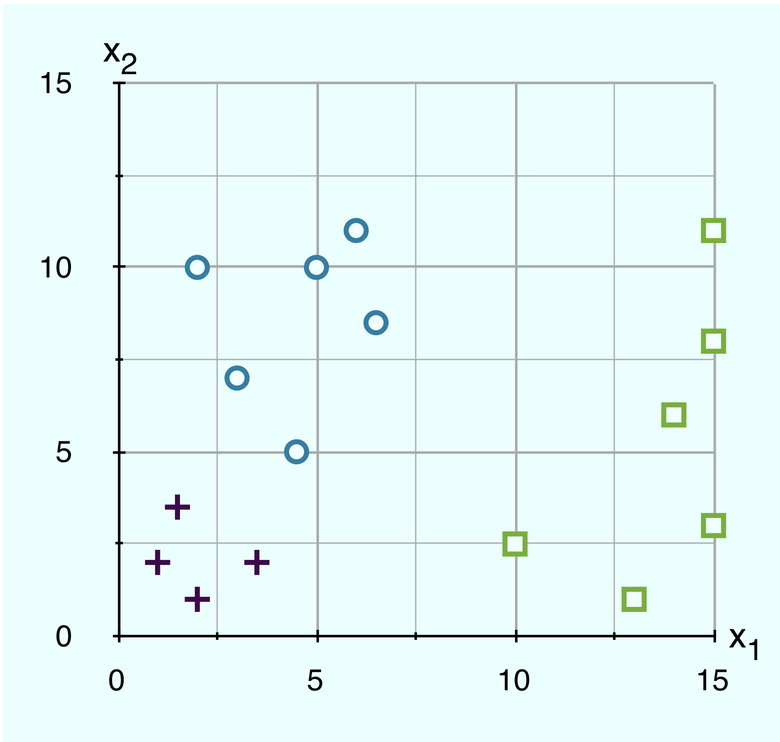
Given the depicted case-base of 16 instances of three different classes, each having two real-valued attributes (x_1, x_2), solve the following excercises.

For your reference:

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

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

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



(a) Construct the decision boundaries between the three classes A (cross), B (circle), and C (square) for a 1-NN classifier based on the euclidian distance measure. Use a print-out or copy of the above diagram for your drawing.

(b) The use of kd-trees as index structure allows for a more efficient retrieval in a case-based reasoning system. In contrast to simple binary search trees, here more than one dimension can be used for partitioning the data space. Therefore, when constructing kd-trees, it must be decided with respect to which dimension (which attribute) and with respect to which value of this attribute a splitting shall be performed. Which attribute is useful for the first partitioning of the data set? Confirm this by calculating the inter-quartile distances for both dimensions.

(c) When using median splitting, the set of cases considered is partitioned into two subsets of nearly equal size. The set of cases "below the median" ($\leq$ the median element) must contain maximally one more element than the set of cases "above the median" ($>$ the median element). Construct a partitioning (bucket size 2) for the given case base using the median splitting method and alternating the dimension, starting with dimension 1. Use a print-out or copy of the above diagram to draw your solution and number the sequence of splitting steps.

Exercise 4.2 (Joint Probability Distribution)

Given the joint probability distribution table

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

where cell A,B specifies the probability for $P(A \wedge B)$[1] $= 0.4$, calculate the following probabilities:

(a) $P(A)$, $P(B)$, $P(\neg A)$, and $P(\neg B)$

(b) $P(A \vee B)$ and $P((A \vee B) \wedge \neg(A \wedge B))$

(c) $P(A|B)$ and $P(B|A)$

Exercise 4.3 (Posterior Probability)

After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease, and that the test is 99% accurate (i.e., the probability of testing positive given that you have the disease is 0.99, as is the probability of testing negative given that you don't have the disease). The good news is that this is a rare disease, striking only one in 10,000 people. Why is it good news that the disease is rare? What are the chances that you actually have the disease?

The exercise sheets may and should be handed in and be worked on in groups of three (3) students. Please fill the cover sheet[2] and attach it to your solution.

[1] shorthand for $P(X_1 = A \text{ and } X_2 = B)$

[2] `http://www.informatik.uni-freiburg.de/~ki/teaching/ss10/gki/coverSheet-english.pdf`