

Knowledge Discovery in Databases II
WS 2014/2015

Übungsblatt 2: Feature Reduction

Aufgabe 2-1 Subspace Selection by Inconsistency

Determine the most informative subspace using Branch-and-Bound in combination with the inconsistency criterion.

ID	attribute X	attribute Y	attribute Z	class
A	2	red	yes	1
B	3	red	yes	1
C	3	green	yes	1
D	4	green	yes	2
E	1	red	yes	2
F	1	green	yes	2

Aufgabe 2-2 Potential of inconsistencies in different domains

Given attributes $A_i \in \mathbb{N}$, attributes $B_i \in \{\text{red, green, blue}\}$, and attributes $C_i \in \{0, 1\}$.

Is it possible for all n elements in a data set to be mutually distinct, when considering a feature space consisting of the following attributes:

- A_1
- B_1
- C_1
- $C_1 \times C_2 \times C_3$
- $B_1 \times C_2$
- $B_i^k \times C_j^l$
- $B_1 \times C_2 \times A_3$

Aufgabe 2-3 Principal Component Analysis

Consider the following example on principal axis transformation.

Given:

$$X = \{(-3, -2), (-2, -1), (-1, 0), (0, 1), (1, 2), (2, 3), \\ (-2, -2), (-1, -1), (0, 0), (1, 1), (2, 2), \\ (-2, -3), (-1, -2), (0, -1), (1, 0), (2, 1), (3, 2)\}$$

- (a) Calculate the covariance matrix M .
- (b) Calculate eigenvalues and eigenvectors of M .
- (c) Determine the smallest eigenvalue and remove its corresponding eigenvector. The remaining eigenvector is the basis of a new sub-space.
- (d) Transform all vectors in X in this new sub-space by expressing all vectors in X in this new basis.

Aufgabe 2-4 Principal Component Analysis

Conduct a principal axis transformation on the following data set:

- $A(1, 0, 3), B(0, 0, 3), C(1, 0, 1), D(0, 0, 1)$

What problem comes up? How can it be solved?