

Knowledge Discovery in Databases II
SoSe 2010

Übungsblatt 2: Feature Reduktion
Besprechung am 6.5.2010

Aufgabe 2-1 *Unterraumselektion mittels Inkonsistenz*

Wir haben in der Vorlesung konsistenzbasierte Kriterien für Feature-Selektion kennengelernt.

Bestimmen Sie für das folgende Klassifikationsproblem den informativsten Unterraum mittels Branch-and-Bound mit Inkonsistenzkriterium.

Jede Klasse umfasst 2 Punkte:

(a) Klasse 1: $A(2, 1, 0)$, $B(3, 1, 0)$

(b) Klasse 2: $E(1, 1, 0)$, $F(1, 0, 0)$

Aufgabe 2-2 *Hauptachsentransformation*

Rechnen Sie noch einmal das Beispiel zur Hauptachsentransformation durch.

Gegeben seien folgende Daten:

$$X = \left\{ \begin{array}{l} (-3, -2), (-2, -1), (-1, 0), (0, 1), (1, 2), (2, 3), \\ (-2, -2), (-1, -1), (0, 0), (1, 1), (2, 2), \\ (-2, -3), (-1, -2), (0, -1), (1, 0), (2, 1), (3, 2) \end{array} \right\}$$

(a) Berechnen Sie die Korrelationsmatrix M .

(b) Berechnen Sie die Eigenwerte und Eigenvektoren von M .

(c) Bestimmen Sie den kleinsten Eigenwert und streichen Sie dessen Eigenvektor. Der resultierende Eigenvektor bildet die Basis für den neuen Unterraum.

(d) Transformieren Sie die Vektoren der Daten X in diesen neuen Unterraum, indem Sie die Vektoren aus X nach dieser neuen Unterraubasis entwickeln.

Aufgabe 2-3 *Hauptachsentransformation*

Führen Sie die Hauptachsentransformation (PCA) auf dem folgenden Datensatz durch:

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