
Data Mining WS 2012/13

Institut für Neuroinformatik

Dr. F. Schwenker

1. Aufgabenblatt (Abgabe bis zum 6.11.2012)

1. Aufgabe (3): Darstellung von 1D-Daten

Gegeben sei der folgende Datensatz:

2, 4, 6, 3, 7, 4, 1, 0, 4, 3, 2, 4, 5, 5, 1, 5, 4, 4, 2, 7, 1, 3, 3, 4, 4, 4, 6, 6, 3, 1, 2, 3, 4, 5,
7, 4, 1, 4, 5, 3, 6, 5, 3, 3, 5, 2, 4, 0, 1, 6.

Stellen Sie die Daten als Häufigkeitstabelle, als Balkendiagramm und als Kreisdiagramm dar.

2. Aufgabe (8): Statistische Kenngrößen

Bestimmen Sie für den obigen Datensatz die folgenden statistischen Kenngrößen und zeichnen Sie ferner einen *box plot*:

1. Modalwert,
2. Median,
3. $\frac{1}{5}$ -Quantil,
4. arithmetischer Mittelwert,
5. Spannweite,
6. Interquartilbereich,
7. Varianz und Standardabweichung.

3. Aufgabe (4 Punkte): Korrelationskoeffizient

Bestimmen Sie den Korrelationskoeffizienten r_{xy} für die folgenden Daten:

Preis €/Liter (X)	Menge in Liter (Y)
4,70	70
4,30	75
3,80	80
4,50	75
5,40	50
5,00	60
4,10	70
4,30	65
3,90	75
4,00	85

4. Aufgabe (6 Punkte): Eigenschaften von r_{xy}

Zeigen Sie:

1. $r_{xy} \in [-1, 1]$
2. $r_{xy} = 1$, gdw. $a > 0$ und $b \in \mathbb{R}$ mit $y_i = ax_i + b$ für alle $i = 1, \dots, n$.
3. $r_{xy} = -1$, gdw. $a < 0$ und $b \in \mathbb{R}$ mit $y_i = ax_i + b$ für alle $i = 1, \dots, n$.