

Übungen zur Vorlesung
Wissensentdeckung in Datenbanken
Sommersemester 2008
Blatt 6

Aufgabe 6.1 (5 Punkte)

Bestimmen Sie eine Annäherung der Zahl π , in dem Sie zunächst gleichverteilte Zufallsvektoren $X = (X_1, X_2)'$ in $[-1, 1] \times [-1, 1]$ erzeugen (in \mathbb{R} ist dies mit der Funktion `runif()` möglich) und die Wahrscheinlichkeit für $P(\{\|X\| \leq 1\})$ schätzen.

Hinweis: Für die Dichtefunktion von zweidimensional gleichverteilten Zufallsvektoren in $[a, b] \times [c, d]$ gilt:

$$f(x_1, x_2) = \begin{cases} \frac{1}{(b-a)(d-c)}, & x_1 \in [a, b] \text{ und } x_2 \in [c, d] \\ 0, & \text{sonst} \end{cases}$$

sowie

$$P(\{a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2\}) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x_1, x_2) dx_2 dx_1.$$

- Zeigen Sie zunächst, dass $\hat{\pi} := 4\hat{P}(\{\|X\| \leq 1\})$ eine sinnvolle Schätzung für π darstellt!
- Schätzen Sie $\hat{P}(\{\|X\| \leq 1\})$ durch $\frac{1}{n} \sum_{i=1}^n I_{\{\|x_i\| \leq 1\}}(x_i)$, den Anteil an Realisierungen $x_i = (x_{i1}, x_{i2})'$ innerhalb des Einheitskreises für unterschiedliche Anzahlen $n \in \{10, 20, 50, 100, 1000, 5000, 10000\}$ an erzeugten Zufallszahlen!
- Stellen Sie $\hat{\pi}$ in Abhängigkeit von n grafisch dar. Wiederholen Sie nun für jedes n die Schätzung zehn mal und berechnen Sie die Standardabweichung in Abhängigkeit von n .

Aufgabe 6.2 (5 Punkte)

Im Netz unter

<http://www.statistik.tu-dortmund.de/download/vorlesungen/wissensentdeckung/Daten/>
finden Sie den Datensatz `auto-mpg.txt` sowie ein File mit der zugehörigen Beschreibung.

- Was für Daten beinhaltet dieser Datensatz?
- Lesen sie den Datensatz in \mathbb{R} ein. Achten Sie dabei insbesondere auf fehlende Werte. Entfernen Sie die Beobachtungen mit fehlenden Werten aus dem Datensatz! Tragen Sie in zweidimensionalen Grafiken die zu modellierende Variable `mpg` gegen die Variablen `displacement`, `horsepower`, `weight` sowie `acceleration` ab. Welche Zusammenhänge lassen sich vermuten?
- Berechnen Sie die Korrelationskoeffizienten zwischen der Variablen `mpg` und den anderen Merkmalen aus Aufgabe **b**!
- Bilden Sie ein lineares Modell zur Vorhersage von `mpg` durch die anderen vier Merkmale

(in R ist dies mit der Funktion `lm()` möglich). Interpretieren Sie den Output. Wie sind die geschätzten Einflüsse. Bestehen signifikante Effekte? Was sagt der Wert von R^2 aus? Wie beurteilen Sie das angepasste Modell?

Allgemeines

ACHTUNG: Abgabe an andere eMail-Adresse (s.o.) (und bitte nur eine Datei!)