

Mustererkennung: Wahrscheinlichkeitstheorie II

D. Schlesinger – TUD/INF/KI/IS

Zwei Ereignisse $A \subset \Omega$ und $B \subset \Omega$ sind unabhängig, wenn

$$P(A \cap B) = P(A) \cdot P(B)$$

Interessant: Ereignisse A und $\bar{B} = \Omega/B$ sind unabhängig, falls A und B unabhängig sind.

Bedingte Wahrscheinlichkeit:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Eine andere Definition der Unabhängigkeit: A und B sind unabhängig, wenn

$$P(A|B) = P(A) \quad \text{bzw.} \quad P(B|A) = P(B)$$

(äquivalent, wenn $P(B) \neq 0$ bzw. $P(A) \neq 0$).

Bayessche Formel:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Mehrdimensionale (reellwertige) statistische Größen – $\xi : \Omega \rightarrow \mathbb{R}^n$

Äquivalent (Einfachheit halber $n = 2$): $\xi = (\xi_1, \xi_2)$ mit $\xi_1 : \Omega \rightarrow \mathbb{R}$ und $\xi_2 : \Omega \rightarrow \mathbb{R}$

Verteilungsfunktion:

$$F_{\xi}(r, s) = P(\{\omega : \xi_1(\omega) \leq r\} \cap \{\omega : \xi_2(\omega) \leq s\})$$

Zwei statistische Größen sind unabhängig, wenn

$$F_{\xi=(\xi_1, \xi_2)}(r, s) = F_{\xi_1}(r) \cdot F_{\xi_2}(s) \quad \forall r, s$$

Verbundwahrscheinlichkeitsverteilung (diskret):

$$p_{\xi=(\xi_1, \xi_2)}(r, s) = P(\{\omega : \xi_1(\omega) = r\} \cap \{\omega : \xi_2(\omega) = s\})$$

Verbundwahrscheinlichkeitsdichte (kontinuierlich):

$$p_{\xi=(\xi_1, \xi_2)}(r, s) = \frac{\partial F_{\xi}(r, s)}{\partial r \partial s}$$

Definitionen

Kürzel: $p(x, y) \equiv p_{\xi}(x, y)$

Marginale Wahrscheinlichkeit:

$$p(x) = \sum_y p(x, y)$$

Bedingte Wahrscheinlichkeit:

$$p(x|y) = \frac{p(x, y)}{p(y)}$$

Merke: $\sum_x p(x|y) = 1$.

Unabhängige Wahrscheinlichkeitsverteilungen:

$$p(x, y) = p(x) \cdot p(y)$$

oder (analog zu Ereignissen)

$$p(x|y) = p(x)$$

Interessant: In das Intervall $[0, 1]$ werden zufällig und nacheinander 3 Punkte geworfen. Wie groß ist die Wahrscheinlichkeit, dass die Koordinate des dritten Punktes größer ist, als die der beiden ersten?

Gegeben sei zwei statistische Größen.

Typischerweise ist eine davon diskret (d.h. $k \in K$) und heißt „Klasse“.

Die andere ist allgemein (sehr oft kontinuierlich, d.h. $x \in \mathbb{R}^n$) und heißt „Beobachtung“.

„Gegeben“ sei die Verbundwahrscheinlichkeitsverteilung $p(x, k)$.

Da die k diskret ist, wird oft $p(x, k)$ durch $p(x, k) = p(k) \cdot p(x|k)$ spezifiziert.

Die Erkennungsaufgabe:

man beobachtet x , man sage etwas über k

– „welche Klasse hat die Beobachtung x verursacht“.

Schwierigkeiten (Fragen):

- Wie ist k anhand der x zu wählen?
- Die Wahrscheinlichkeitsverteilung ist meist nicht explizit angegeben
→ dafür sorgen, dass das eine Wahrscheinlichkeitsverteilung ist.
- Die Menge K ist manchmal sehr groß (erinnere an Hopfield-Netze).

Lernen:

Das Wahrscheinlichkeitsmodell ist oft bis auf unbekannte freie Parameter definiert.

Wie sind diese zu wählen?