

Maschinelles Lernen und Data Mining
 Sommersemester 2008
Übungsblatt 6

Besprechung des Übungsblattes am 23.06.2008

Aufgabe 6-1 Wahrscheinlichkeitsrechnung
schriftlich bearbeiten

Gegeben seien die folgende Daten:

		Y		
		1	2	3
X	1	0,1	0,15	0,25
	2	0,05	0,3	0,15

Berechnen Sie

- die Randwahrscheinlichkeiten $P(X = x_i)$ und $P(Y = y_i)$
- die Erwartungswerte $E(X)$, $E(Y)$
- die Varianzen $Var(X)$, $Var(Y)$, sowie die Kovarianz $Cov(X,Y)$.
- die Korrelation $\rho = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}}$ (Achtung, im Skript fehlt die Wurzel des Nenners)
- Sind die Variablen X,Y unabhängig? Begründen Sie ihre Entscheidung.

Aufgabe 6-2 Bedingte Wahrscheinlichkeiten
schriftlich bearbeiten

Gegeben die Datentabelle aus Aufgabe 1, sowie die folgenden Ereignisse:

$$A \begin{cases} 0 & x \leq y \\ 1 & \text{sonst} \end{cases}$$

$$B \begin{cases} 0 & x = y \\ 1 & \text{sonst} \end{cases}$$

Berechnen Sie: $P(B = 0|A = 0)$, $P(A = 0|B = 0)$, $P(B = 1|A = 0)$

Aufgabe 6-3 Interpretation der Standardabweichung

Skizzieren Sie den Graph der Standardnormalverteilung

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right); \mu = 0; \sigma = 1$$

im Intervall $x \in [-4, 4]$. Markieren und interpretieren Sie die Bereiche $0 \pm \sigma$; $0 \pm 2\sigma$; $0 \pm 3\sigma$

Aufgabe 6-4 Spamfilter

Skizzieren Sie einen einfachen Bayes-Spamfilter. Gehen Sie davon aus, dass die Wörter der zu analysierenden Nachricht voneinander unabhängig sind.