

**Maschinelles Lernen und Data Mining**  
 Sommersemester 2010  
**Übungsblatt 4**

Besprechung des Übungsblattes am 10.06.2010

**Aufgabe 4-1**      Wahrscheinlichkeitsrechnung  
*schriftlich bearbeiten*

Gegeben sei die folgende Wahrscheinlichkeitstabelle:

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

Berechnen Sie:

- die Randwahrscheinlichkeiten  $P(X = i)$  für  $i \in \{1, 2\}$  und  $P(Y = j)$  für  $j \in \{1, 2, 3\}$ ,
- die Erwartungswerte  $E(X)$ ,  $E(Y)$ ,
- die Varianzen  $Var(X)$ ,  $Var(Y)$ , sowie die Kovarianz  $Cov(X, Y)$ ,
- und die Korrelation  $\rho = \frac{cov(X,Y)}{\sqrt{var(X) \cdot var(Y)}}$ .
- Sind die Variablen  $X$  und  $Y$  unabhängig?

**Aufgabe 4-2**      Interpretation der Standardabweichung

Skizzieren Sie den Graph der Standardnormalverteilung  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$ ; mit  $\mu = 0$  und  $\sigma = 1$  im Intervall  $x \in [-4, 4]$ . Markieren und interpretieren Sie die Bereiche  $0 \pm \sigma$ ;  $0 \pm 2\sigma$ ;  $0 \pm 3\sigma$

**Aufgabe 4-3**      Personengröße  
*schriftlich bearbeiten*

Angenommen die Größe einer Population sei normalverteilt:

$$P_{\mathbf{w}}(\mathbf{x}_i) = \mathcal{N}(\mathbf{x}_i; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{x}_i - \mu)^2\right)$$

Wenn  $\mathbf{w} = (\mu, \sigma)^T$ , so gilt

$$\begin{aligned} P_{\mathbf{w}}(\mathbf{x}_1, \dots, \mathbf{x}_N) &= \prod_{i=1}^N P_{\mathbf{w}}(\mathbf{x}_i) = \prod_{i=1}^N \mathcal{N}(\mathbf{x}_i; \mu, \sigma^2) = \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^N (\mathbf{x}_i - \mu)^2\right) \end{aligned}$$

b.w.

- a) Bestimmen Sie (*mit* Herleitung) den Maximum Likelihood Schätzer von  $P_{\mathbf{w}}(\mathbf{x}_1, \dots, \mathbf{x}_N)$ .
- b) Berechnen Sie die entsprechenden Schätzer für die Höhen-Datensätze unter `body_sizes.txt` und visualisieren Sie die dazugehörigen Verteilungen. Ist der Schätzer hilfreich für das Verständnis der Daten?