Ludwig-Maximilians-Universität München 31.05.2012
Institut für Informatik
Prof. Dr. Volker Tresp
Tobias Emrich

Maschinelles Lernen und Data Mining

Sommersemester 2013

Übungsblatt 4

Besprechung des Übungsblattes am 06.06.2013

Aufgabe 4-1 Wahrscheinlichkeitsrechnung

Gegeben seien die folgende Daten über die Zufallsvariablen X und Y:

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

Berechnen Sie

a) die Randwahrscheinlichkeiten $P(X = x_i)$ und $P(Y = y_i)$

b) die Erwartungswerte $E(X)$, $E(Y)$

c) die Varianzen $var(X)$, $var(Y)$, sowie die Kovarianz $cov(X,Y)$.

d) die Korrelation $\rho = \frac{cov(X,Y)}{\sqrt{var(X) \cdot var(Y)}}$

e) Sind die Variablen X,Y unabhängig?

Lösungsvorschlag:

a)

		Y 1	2	3	
X	1	0,1	0,15	0,25	0,5
	2	0,05	0,3	0,15	0,5
		0,15	0,45	0,4	1

b)

$$\begin{aligned} E(X) &= E_{P(x)}(X) = \sum_i x_i P(X = x_i) \\ E(X) &= 1 \cdot 0{,}5 + 2 \cdot 0{,}5 = 1{,}5 \\ E(Y) &= 1 \cdot 0{,}15 + 2 \cdot 0{,}45 + 3 \cdot 0{,}4 = 2{,}25 \end{aligned}$$

c) $var(X) = \sum\limits_i [x_i - E(X)]^2 \cdot P(X = x_i) = E((X - E(X))^2) = E(X^2) - (E(X))^2$

$$\begin{aligned} var(X) =& (1 - 1{,}5)^2 \cdot 0{,}5 + (2 - 1{,}5)^2 \cdot 0{,}5 = 0{,}25 \\ var(Y) =& (1 - 2{,}25)^2 \cdot 0{,}15 + (2 - 2{,}25)^2 \cdot 0{,}45 + (3 - 2{,}25)^2 \cdot 0{,}4 = 39/80 \approx 0{,}4875 \\ cov(X{,}Y) =& \sum_i \sum_j (x_i - E(X))(y_i - E(Y)) P(X = x_i, Y = y_i) \\ =& 0.1(1 - 1.5)(1 - 2.25) + 0.15(1 - 1.5)(2 - 2.25) + 0.25(1 - 1.5)(3 - 2.25) + \\ & 0.05(2 - 1.5)(1 - 2.25) + 0.3(2 - 1.5)(2 - 2.25) + 0.15(2 - 1.5)(3 - 2.25) \\ =& -0{,}025 \\ & oder\ mit\ Verschiebungssatz : cov(X{,}Y) = E(XY) - E(X)E(Y) \\ E(XY) =& \underbrace{1 * 0.1 + 2 * 0.15 + 3 * 0.25}_{Zeile1} + \underbrace{2 * 0.05 + 4 * 0.3 + 6 * 0.15}_{Zeile2} = 3{,}35 \\ cov(XY) =& 3{,}35 - 1{,}5 * 2{,}25 = -0{,}025 \end{aligned}$$

d) $\rho = \frac{cov(X,Y)}{\sqrt{var(X) \cdot var(Y)}} = \frac{-0{,}025}{\sqrt{0{,}25 \cdot 0{,}4875}} \approx -0{,}07$

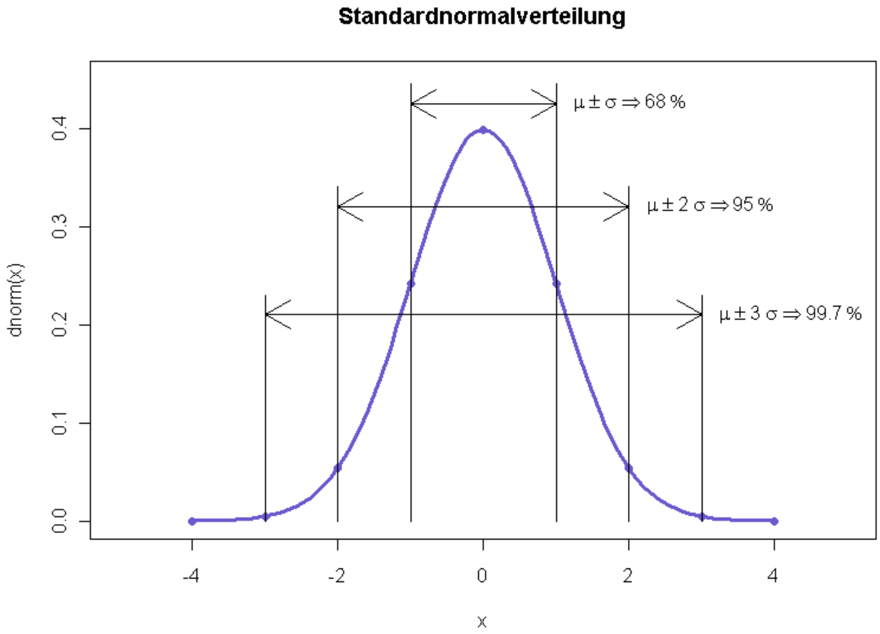
e) Nein, weil:

- $\rho \neq 0$ nicht aussagekräftig! Weil: $X{,}Y$ unabhängig $\Rightarrow$ $X{,}Y$ unkorreliert (keine Rückrichtung)
- $P(X = 2{,}Y = 1) = 0{,}05 \neq 0{,}15 * 0{,}5 = 0{,}075$, damit keine Unabhängigkeit

Aufgabe 4-2 Interpretation der Standardabweichung

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

Lösungsvorschlag:



Verteilungsfunktionen in `R` unter `dnorm`:

Dichteverteilung: $P(X = x)$ (`dnorm`)

Wahrscheinlichkeitsverteilung: $P(X <= x)$ (`pnorm`)

Quantilverteilung: $\arg_x P(X <= x) = q$ (`qnorm`)

Zufällige Stichprobe aus der Verteilung: Datensatz der entsprechend der Verteilung zufallsverteilt ist. (`rnorm`)

Folgen die Daten einer Standardnormalverteilung, gilt die 68-95-99.7-Prozent-Regel:

- 68% der Beobachtungen liegen im Intervall $\mu \pm \sigma$
  (also $P(X \geq \mu - \sigma \wedge X \leq \mu + \sigma) = P(X \leq \mu + \sigma) - P(X \geq \mu - \sigma)$)
- 95% der Beobachtungen liegen im Intervall $\mu \pm 2\sigma$
- 99.7% der Beobachtungen liegen im Intervall $\mu \pm 3\sigma$

Wenn sich eine Häufigkeitsverteilung für die Daten eines Merkmals X gut durch eine Normalverteilung mit $\mu = 0, \sigma = s$ approximieren lässt, gelten die Aussagen wie die 68-95-99.7-Prozent-Regel auch approximativ für diese Daten.
Quelle: Statistik (Ludwig Fahrmeir, Rita Künstler, Iris Pigeot) S. 92f

Aufgabe 4-3 Kernkombinationen

Um einen selbstdefinierten Kern $k(\mathbf{x}_i, \mathbf{x}_j)$ für $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^n$ anzuwenden, muss gezeigt werden, dass es sich auch tatsächlich um einen legitimen Kern handelt. Da es recht aufwendig sein kann zu zeigen, dass für k das *Mercer Theorem* zutrifft, wird häufig explizit das Mapping der impliziten Basistransformationen angegeben: $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$.

Eine weitere beliebte Variante, die Gültigkeit einer Kernfunktion zu zeigen, ist die Rückführung auf eine Kombination aus Kernen, da für einige Operationen $\circ$ gilt, dass $k(\mathbf{x}_i, \mathbf{x}_j) = k_1(\mathbf{x}_i, \mathbf{x}_j) \circ k_2(\mathbf{x}_i, \mathbf{x}_j)$ ein legitimer Kern ist.

Zeigen Sie, dass für valide Kernfunktionen $k_l(\mathbf{x}_i,\mathbf{x}_j)$ mit $l \in \mathbb{N}_+$ gilt:

b.w.

a) **Skalierung**: Für $a > 0$ ist $k(\mathbf{x}_i,\mathbf{x}_j) := a \cdot k_1(\mathbf{x}_i,\mathbf{x}_j)$ ein Kern.

b) **Summe**: $k(\mathbf{x}_i,\mathbf{x}_j) := k_1(\mathbf{x}_i,\mathbf{x}_j) + k_2(\mathbf{x}_i,\mathbf{x}_j)$ ist ein Kern.

c) **Linearkombination**: Für $w \in \mathbb{R}^d_+$ ist $k(\mathbf{x}_i,\mathbf{x}_j) := \sum_{l=1}^{d} w_l \cdot k_l(\mathbf{x}_i,\mathbf{x}_j)$ ein Kern.

d) **Produkt**: $k(\mathbf{x}_i,\mathbf{x}_j) := k_1(\mathbf{x}_i,\mathbf{x}_j) \cdot k_2(\mathbf{x}_i,\mathbf{x}_j)$ ist ein Kern.

e) **Potenz**: Für ein $p \in \mathbb{N}_+ : k(\mathbf{x}_i,\mathbf{x}_j) := (k_1(\mathbf{x}_i,\mathbf{x}_j))^p$ ist ein Kern.

Lösungsvorschlag:

a)

$$ak_1(\mathbf{x}_i,\mathbf{x}_j) = a\phi_1(\mathbf{x}_i)^T\phi_1(\mathbf{x}_j) = a\sum_{m=1}^{M}\phi_1(\mathbf{x}_i)_m\phi_1(\mathbf{x}_j)_m = \sum_{m=1}^{M}\sqrt{a}\phi_1(\mathbf{x}_i)_m\sqrt{a}\phi_1(\mathbf{x}_j)_m =$$

$$= \sum_{m=1}^{M}\phi'_1(\mathbf{x}_i)_m\phi'_1(\mathbf{x}_j)_m = \phi'_1(\mathbf{x}_i)^T\phi'_1(\mathbf{x}_j)$$

mit $\phi'_1(\mathbf{x}_i) = \sqrt{a}\phi_1(\mathbf{x}_i)$

b)

$$k_1(\mathbf{x}_i,\mathbf{x}_j) + k_2(\mathbf{x}_i,\mathbf{x}_j) = \phi_1(\mathbf{x}_i)^T\phi_1(\mathbf{x}_j) + \phi_2(\mathbf{x}_i)^T\phi_2(\mathbf{x}_j) = \phi''_{12}(\mathbf{x}_i)^T\phi''_{12}(\mathbf{x}_j)$$

mit $\phi''_{12}(\mathbf{x}_i) = \begin{pmatrix} \phi_1(\mathbf{x}_i) \\ \phi_2(\mathbf{x}_i) \end{pmatrix}$

c)

$$\sum_{l=1}^{n} w_l k_l(\mathbf{x}_i,\mathbf{x}_j) \overset{\text{a)}}{=} \sum_{l=1}^{n}\phi'_l(\mathbf{x}_i)\phi'_l(\mathbf{x}_j) \overset{\text{b)}}{=} \phi''_{1\ldots l}(\mathbf{x}_i)^T\phi''_{1\ldots l}(\mathbf{x}_j)$$

mit $\phi'_l(\mathbf{x}_i) = \sqrt{w_l}\phi_l(\mathbf{x}_i)$ und $\phi''_{1\ldots l}(\mathbf{x}_i) = \begin{pmatrix} \phi'_1(\mathbf{x}_i) \\ \phi'_2(\mathbf{x}_i) \\ \vdots \\ \phi'_l(\mathbf{x}_i) \end{pmatrix}$

b.w.

Lösungsvorschlag:

d)

$$\begin{aligned}
k_1(\mathbf{x}_i,\mathbf{x}_j)\cdot k_2(\mathbf{x}_i,\mathbf{x}_j) &= \left(\sum_{m=1}^{M_1}\phi_1(\mathbf{x}_i)_m\cdot\phi_1(\mathbf{x}_j)_m\right)\cdot\left(\sum_{n=1}^{M_2}\phi_2(\mathbf{x}_i)_n\cdot\phi_2(\mathbf{x}_j)_n\right) = \\
&= \sum_{m=1}^{M_1}\left(\phi_1(\mathbf{x}_i)_m\phi_1(\mathbf{x}_j)_m\cdot\left(\sum_{n=1}^{M_2}\phi_2(\mathbf{x}_i)_n\phi_2(\mathbf{x}_j)_n\right)\right) = \\
&= \sum_{m=1}^{M_1}\sum_{n=1}^{M_2}\left(\phi_1(\mathbf{x}_i)_m\phi_1(\mathbf{x}_j)_m\cdot\phi_2(\mathbf{x}_i)_n\phi_2(\mathbf{x}_j)_n\right) = \\
&= \sum_{m=1}^{M_1}\sum_{n=1}^{M_2}\left(\phi_1(\mathbf{x}_i)_m\phi_2(\mathbf{x}_i)_n\right)\cdot\left(\phi_1(\mathbf{x}_j)_m\phi_2(\mathbf{x}_j)_n\right) = \\
&= \phi_{12}'''(\mathbf{x}_i)^T\phi_{12}'''(\mathbf{x}_j) = k(x_i,x_j)
\end{aligned}$$

$$\text{mit } \phi_{12}'''(\mathbf{x}_i) = \begin{pmatrix} \phi_1(\mathbf{x}_i)_1\cdot\phi_2(\mathbf{x}_i)_1 \\ \phi_1(\mathbf{x}_i)_1\cdot\phi_2(\mathbf{x}_i)_2 \\ \vdots \\ \phi_1(\mathbf{x}_i)_1\cdot\phi_2(\mathbf{x}_i)_{M_2} \\ \phi_1(\mathbf{x}_i)_2\cdot\phi_2(\mathbf{x}_i)_1 \\ \vdots \\ \phi_1(\mathbf{x}_i)_{M_1}\cdot\phi_2(\mathbf{x}_i)_{M_2} \end{pmatrix}$$

e) Beweis per Induktion:

I.V.: $k(\mathbf{x}_i,\mathbf{x}_j) := k_1(\mathbf{x}_i,\mathbf{x}_j)^p$ ist ein Kern (für p positives Integer).

I.A.: $k_1(x_i,x_j)\cdot k_1(x_i,x_j)$ ist ein legaler Kern. Gezeigt in d).

I.S.: $k_1(\mathbf{x}_i,\mathbf{x}_j)^{p+1} = k_1(\mathbf{x}_i,\mathbf{x}_j)^p\cdot k_1(\mathbf{x}_i,\mathbf{x}_j) \overset{\text{I.V.}}{=} \text{ Kern }\cdot k_1(\mathbf{x}_i,\mathbf{x}_j) \overset{\text{d)}}{=}$ valider Kern.