Wahrscheinlichkeitsrechnung

Mustererkennung und Klassifikation, Vorlesung No. 2

M. O. Franz

18.10.2007

Überblick

1. Diskrete Zufallsvariablen
2. Paare von diskreten Zufallsvariablen
3. Bedingte Wahrscheinlichkeiten und der Satz von Bayes
4. Vektorielle Zufallsvariablen

Diskrete Zufallsvariablen

Eine Zufallsvariable x nimmt Werte aus einer Menge $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ mit den jeweiligen Wahrscheinlichkeiten $p_1, p_2, \ldots, p_n$ an.
Damit müssen die p_i folgende Bedingungen erfüllen:

$$p_i \geq 0 \quad \text{und} \quad \sum_{i=1}^{n} p_i = 1$$

Alternative Schreibweise mit Wahrscheinlichkeitsdichte $p(x)$

$$p(x) \geq 0 \quad \text{und} \quad \sum_{x_i \in \mathcal{X}} p(x) = 1$$

Wahrscheinlichkeit für Untermenge $V \in \mathcal{X}$

$$p(V) = \sum_{x_i \in V} p(x)$$

i	a_i	p_i
1	a	0.0575
2	b	0.0128
3	c	0.0263
4	d	0.0285
5	e	0.0913
6	f	0.0173
7	g	0.0133
8	h	0.0313
9	i	0.0599
10	j	0.0006
11	k	0.0084
12	l	0.0335
13	m	0.0235
14	n	0.0596
15	o	0.0689
16	p	0.0192
17	q	0.0008
18	r	0.0508
19	s	0.0567
20	t	0.0706
21	u	0.0334
22	v	0.0069
23	w	0.0119
24	x	0.0073
25	y	0.0164
26	z	0.0007
27	–	0.1928

[MacKay, 2003]

Erwartungswerte

Erwartungswert (Mittelwert) einer Zufallsvariable ($\cong$ Schwerpunkt) :

$$\mathcal{E}[x] = \mu = \sum_{x_i \in \mathcal{X}} x p(x) = \sum_{i=1}^{n} x_i p_i$$

Erwartungswert einer Funktion $f(x)$ der Zufallsvariable x:

$$\mathcal{E}[f(x)] = \sum_{x_i \in \mathcal{X}} f(x) p(x)$$

Erwartungswert ist linear:

$$\mathcal{E}[a_1 f_1(x) + a_2 f_2(x)] = a_1 \mathcal{E}[f_1(x)] + a_2 \mathcal{E}[f_2(x)]$$

Spezialfälle:

- Varianz: $\sigma^2 = \mathcal{E}[(x-\mu)^2] = \sum_{x_i \in \mathcal{X}} (x-\mu)^2 p(x)$ ($\cong$ Trägheitsmoment)
- Standardabweichung: $\sigma = \sqrt{\mathcal{E}[(x-\mu)^2]}$

Paare von diskreten Zufallsvariablen

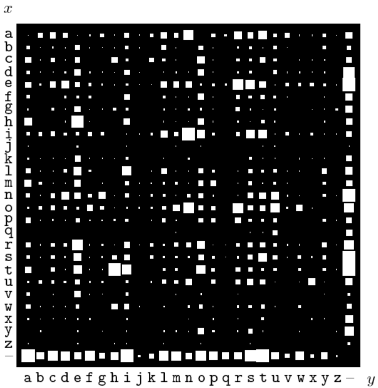
Zwei Zufallsvariablen x und y werden durch eine gemeinsame Verteilung $p(x, y)$ beschrieben, die jedem Wertepaar (x_i, y_j) eine Wahrscheinlichkeit p_{ij} zuordnet. Es gilt

$$p(x, y) \geq 0 \quad \text{und} \quad \sum_{x_i \in \mathcal{X}} \sum_{y_j \in \mathcal{Y}} p(x, y) = 1$$

Randwahrscheinlichkeitsfunktionen von x und y:

$$p(x) = \sum_{y_j \in \mathcal{Y}} p(x, y)$$

$$p(y) = \sum_{x_i \in \mathcal{X}} p(x, y)$$



[MacKay, 2003]

Statistische Unabhängigkeit und Erwartungswerte bei zwei Variablen

Zwei Variablen x und y sind statistisch unabhängig genau dann, wenn

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

Erwartungswert einer Funktion $f(x,y)$ von 2 Zufallsvariablen x und y:

$$\mathcal{E}[f(x,y)] = \sum_{x_i \in \mathcal{X}} \sum_{y_j \in \mathcal{Y}} f(x,y)p(x,y)$$

- Mittelwert von x: $\mu_x = \mathcal{E}[x] = \sum_{x_i \in \mathcal{X}} \sum_{y_j \in \mathcal{Y}} xp(x,y)$
- Mittelwert von y: $\mu_y = \mathcal{E}[y] = \sum_{x_i \in \mathcal{X}} \sum_{y_j \in \mathcal{Y}} yp(x,y)$
- Varianz von x: $\sigma_x^2 = \mathcal{E}[(x-\mu_x)^2] = \sum_{x_i \in \mathcal{X}} \sum_{y_j \in \mathcal{Y}} (x-\mu_x)^2 p(x,y)$
- Varianz von y: $\sigma_y^2 = \mathcal{E}[(x-\mu_x)^2] = \sum_{x_i \in \mathcal{X}} \sum_{y_j \in \mathcal{Y}} (y-\mu_y)^2 p(x,y)$
- Kovarianz von x und y:

$$\sigma_{xy} = \mathcal{E}[(x-\mu_x)(y-\mu_y)] = \sum_{x_i \in \mathcal{X}} \sum_{y_j \in \mathcal{Y}} (x-\mu_x)(y-\mu_y)p(x,y)$$

Bedingte Wahrscheinlichkeiten

Wenn zwei ZV voneinander statistisch abhängig sind, erhöht Kenntnis der einen ZV das Wissen über die andere:

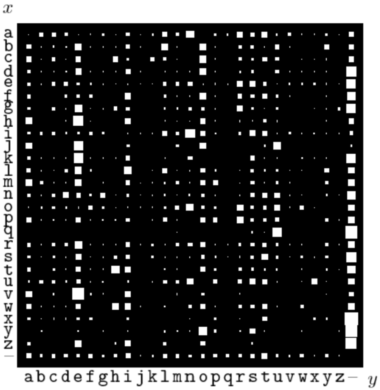
$$p(x|y) = \frac{p(x,y)}{p(y)} \quad \text{bzw.}$$

$$p(x,y) = p(x|y)p(y) \quad \text{(Kettenregel)}$$

$p(x|y)$: Wahrscheinlichkeit von x gegeben y.

Bei statistisch unabhängigen ZV gilt

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



(a) $P(y \mid x)$

[MacKay, 2003]

Der Satz von Bayes

Es gelten

$$
\begin{aligned}
p(y) &= \sum_{x_i \in \mathcal{X}} p(x,y) \quad \text{(Randwahrscheinlichkeit von } y\text{)} \\
p(x,y) &= p(y|x)p(x) \quad \text{und} \quad p(x,y) = p(x|y)p(y) \quad \text{(Kettenregel)}
\end{aligned}
$$

Substitution ergibt Satz von Bayes:

$$
p(x|y) = \frac{p(y|x)p(x)}{\sum_{x_i \in \mathcal{X}} p(y|x)p(x)}
$$

In Worten:

$$
\text{A-posteriori-Wahrscheinlichkeit} = \frac{\text{Likelihood} \times \text{A-priori-Wahrscheinlichkeit}}{\text{Evidenz}}
$$

- erlaubt das Umkehren von Schlussfolgerungen: Die Berechnung von P(Ereignis | Ursache) ist häufig einfach, aber oft ist P(Ursache | Ereignis) gesucht.
- beschreibt Lernen aus Erfahrung

Beispielaufgabe

- Bauer Fritz läßt seinen Truthahn Max auf Vogelgrippe testen. Maxs Gesundheitszustand sei a (1 = krank, 0 = gesund), das Testergebnis b (1 = positive, 0 = negativ).
- Der Test hat eine Verläßlichkeit von 95 %, d.h. bei 95% der infizierten Tiere schlägt der Test an, bei 95% der gesunden Tiere schlägt er nicht an.
- In ganz Deutschland sind 1% aller Truthähne erkrankt (stimmt zum Glück bisher nicht wirklich).

Der Test fällt positiv aus. Wie wahrscheinlich ist es, daß Max tatsächlich Vogelgrippe hat?

Vektorielle Zufallsvariablen

Schreibe d Zufallsvariablen x_i als Vektor $\mathbf{x} = (x_1, x_2, \ldots, x_d)$. Damit gilt für $p(\mathbf{x})$

$$p(\mathbf{x}) \geq 0 \quad \text{und} \quad \sum_{\mathbf{x} \in \mathcal{X}} p(\mathbf{x}) = 1.$$

Wenn alle x_i statistisch unabhängig sind, gilt

$$p(\mathbf{x}) = p_1(x_1)p_2(x_2) \cdots p_d(x_d) = \prod_{i=1}^{d} p_i(x_i).$$

Rand- und bedingte Wahrscheinlichkeiten werden analog berechnet, z.B. bei $p(x_1, x_2, x_3, x_4, x_5)$

$$\begin{aligned} p(x_1, x_4) &= \sum_{x_2} \sum_{x_3} \sum_{x_5} p(x_1, x_2, x_3, x_4, x_5) \\ p(x_1, x_2 | x_3, x_4, x_5) &= \frac{p(x_1, x_2, x_3, x_4, x_5)}{p(x_3, x_4, x_5)} \quad \text{oder} \quad p(\mathbf{x}_1 | \mathbf{x}_2) = \frac{p(\mathbf{x}_1, \mathbf{x}_2)}{p(\mathbf{x}_2)} \end{aligned}$$

Bayesformel und Erwartungswerte bei vektoriellen Zufallsvariablen

Satz von Bayes:

$$p(\mathbf{x}_1|\mathbf{x}_2) = \frac{p(\mathbf{x}_2|\mathbf{x}_1)p(\mathbf{x}_1)}{\sum_{\mathbf{x}_1 \in \mathcal{X}} p(\mathbf{x}_2|\mathbf{x}_1)p(\mathbf{x}_1)}$$

Erwartungswert einer vektorwertigen Funktion $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^n$, d.h. $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_n(\mathbf{x}))$

$$\mathcal{E}[\mathbf{f}] = \sum_{\mathbf{x}} \mathbf{f}(\mathbf{x}) p(\mathbf{x})$$

- Mittelwert: $\mu = \mathcal{E}[\mathbf{x}] = \sum_{\mathbf{x}} \mathbf{f}(\mathbf{x}) p(\mathbf{x})$
- Kovarianzmatrix: $\sigma_{ij} = \mathcal{E}[(x_i - \mu_i)(x_j - \mu_j)]$