Wahrscheinlichkeitsrechnung

Irene Winkler

Arbeitsgruppe Maschinelles Lernen

29. März 2012

Wahrscheinlichkeit

"nothing but common sense reduced to calculation." (Laplace)

Die Wahrscheinlichkeit P

- misst relative Häufigkeit auf einer Skala von 0 (nie) bis 1 (immer)
- $P(A \cup B) = P(A) + P(B)$ falls $A \cap B = \emptyset$

Für unsere Zwecke brauchen wir

- *Zufallsvariablen* - die Ergebnisse eines Zufallsexperiments - und
- deren *Verteilungen* - die relative Häufigkeit der Ergebnisse

Diskrete Zufallsvariablen

$X \in \{x_1, \ldots, x_n\}$, $Y \in \{y_1, \ldots, y_m\}$ diskrete Zufallsvariablen

- Randverteilung $P(X)$, $P(Y)$
- Gemeinsame Verteilung $P(X, Y)$
- Bedingte bzw. Posteriorverteilung $P(X|Y), P(Y|X)$

Diskrete Zufallsvariablen

- Summenregel:

$$P(X = x) = \sum_{y \in Y} P(X = x, Y = y)$$

- Produktregel:

$$P(X = x, Y = y) = P(Y = y | X = x) P(X = x)$$

- Bayesregel (2 x Anwendung Produktregel):

$$P(Y = y | X = x) = \frac{P(X = x | Y = y) P(Y = y)}{P(X = x)}$$

Kontinuierliche Zufallsvariablen

Sei $X \in \mathbb{R}$ eine reellwertige Zufallsvariable. Eine Funktion $p : \mathbb{R} \mapsto \mathbb{R}$ heisst *Dichte* von X falls für $a < b$

$$P(X \in [a, b]) = \int_a^b p(x)dx$$

Für jede Dichte gilt

$$\int_{-\infty}^{\infty} p(x)dx = 1 \qquad \text{und} \qquad p(x) \geq 0, \forall x \in \mathbb{R}$$

Kontinuierliche Zufallsvariablen

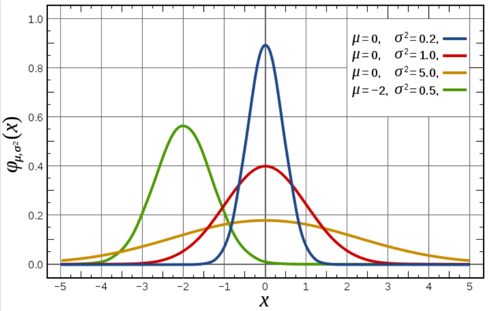
Sei $X \in \mathbb{R}$ eine reellwertige Zufallsvariable. Die (kumulative) *Verteilungsfunktion* $F_X : \mathbb{R} \mapsto \mathbb{R}$ ist definiert durch

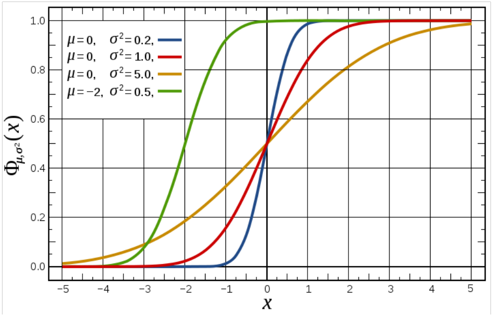
$$F_X(c) = P(X \leq c) = \int_{-\infty}^{c} p(x)dx$$

Es gilt

$$0 \leq F_X(c) \leq 1$$

$$P(X \in [a, b]) = F(b) - F(a)$$


$\mu=0, \quad \sigma^2=0.2,$
$\mu=0, \quad \sigma^2=1.0,$
$\mu=0, \quad \sigma^2=5.0,$
$\mu=-2, \quad \sigma^2=0.5,$
$\varphi_{\mu,\sigma^2}(x)$
x
1.0
0.8
0.6
0.4
0.2
0.0
−5 −4 −3 −2 −1 0 1 2 3 4 5


$\mu=0, \quad \sigma^2=0.2,$
$\mu=0, \quad \sigma^2=1.0,$
$\mu=0, \quad \sigma^2=5.0,$
$\mu=-2, \quad \sigma^2=0.5,$
$\Phi_{\mu,\sigma^2}(x)$
x
1.0
0.8
0.6
0.4
0.2
0.0
−5 −4 −3 −2 −1 0 1 2 3 4 5

Kontinuierliche Zufallsvariablen

Veallgemeinerung für mehrdimensionale Zufallsvariable $X \in \mathbb{R}^d$:

- $p : \mathbb{R}^d \mapsto \mathbb{R}$ Dichte von X falls

$$P(X \in [a_1, b_1] \times \cdots \times [a_d, b_d]) = \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} p(\mathbf{x}) d\mathbf{x}$$

Für jede Dichte gilt

$$\int_{\mathbb{R}^d} p(\mathbf{x}) d\mathbf{x} = 1 \quad \text{und} \quad p(\mathbf{x}) \geq 0, \forall \mathbf{x} \in \mathbb{R}^d$$

Kontinuierliche Zufallsvariablen

Seien $X \in \mathbb{R}^d, Y \in \mathbb{R}^l$ zwei Zufallsvariablen mit gemeinsamer Dichte p, und den Randverteilungen p_X und p_Y.

- Summenregel:

$$p_X(\mathbf{x}) = \int_{\mathbb{R}^l} p(\mathbf{x}, \mathbf{y}) d\mathbf{y}$$

- Produktregel:

$$p(\mathbf{x}, \mathbf{y}) = p_{Y|X=\mathbf{x}}(\mathbf{y}) p(\mathbf{x})$$

- Bayesregel:

$$p_{Y|X=\mathbf{x}}(\mathbf{y}) = \frac{p_{X|Y=\mathbf{y}}(\mathbf{x}) p_Y(\mathbf{y})}{p_X(\mathbf{x})}$$

Unabhängigkeit

Zwei Zufallsvariablen X und Y sind *unabhängig* wenn

- (Im diskreten:)

$$P(X = \mathbf{x}, Y = \mathbf{y}) = P(X = \mathbf{x})P(Y = \mathbf{y})$$

- (Im kontinuierlichen:)

$$p(\mathbf{x}, \mathbf{y}) = p_X(\mathbf{x}) \cdot p_Y(\mathbf{y})$$

Erwartungswert

$=$ mittlerer Wert einer Zufallsvariablen X

- Ist X eine diskrete Zufallsvariable so ist der Erwartungswert

$$\mathbb{E}(X) := \sum_x \mathbf{x} P(X = \mathbf{x})$$

- Ist $X \in \mathbb{R}^d$ eine kontinuierliche Zufallsvariable so ist der Erwartungswert

$$\mathbb{E}(X) := \int_{\mathbb{R}^d} \mathbf{x} p(\mathbf{x}) dx$$

  Dies ist ein d-dimensionaler Vektor

- Der Erwartungswert ist linear:

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

Varianz

Seien X und Y zwei eindimensionale Zufallsvariablen.

- Die *Varianz* von X ist definiert als

$$\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}(X))^2]$$

- Die *Standardabweichung* von X ist definiert als

$$Std(X) := \sqrt{\mathrm{Var}(X)}$$

Kovarianz

Seien X und Y zwei eindimensionale Zufallsvariablen.

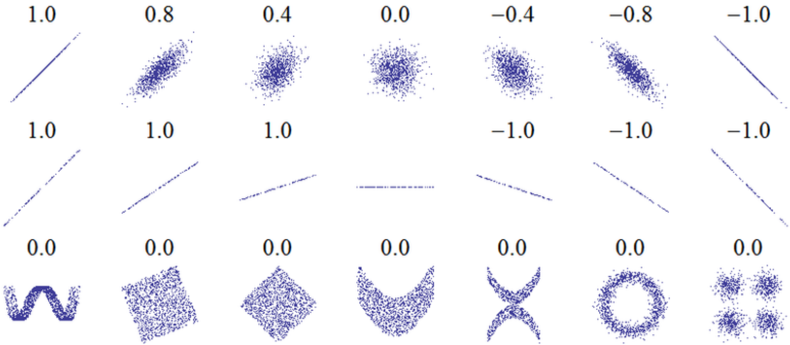
- Die *Kovarianz* von X und Y ist definiert als

$$\mathrm{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

- Die *Korrelation* von X und Y ist definiert als

$$\mathrm{Corr}(X, Y) := \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

  Es gilt $-1 \leq \mathrm{Corr}(X, Y) \leq 1$.
  Maß für den linearen Zusammenhang


1.0
0.8
0.4
0.0
−0.4
−0.8
−1.0
1.0
1.0
1.0
−1.0
−1.0
−1.0
0.0
0.0
0.0
0.0
0.0
0.0
0.0

Rechenregeln I

- Verschiebungssatz

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\ \mathrm{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \end{aligned}$$

- Die Kovarianz ist bilinear

$$\begin{aligned} \mathrm{Cov}(aX+b, cY+d) &= ac\,\mathrm{Cov}(X) \\ \mathrm{Var}(aX+b) &= a^2\,\mathrm{Var}(X) \\ \mathrm{Cov}(X+Z,Y) &= \mathrm{Cov}(X,Y) + \mathrm{Cov}(Z,Y) \end{aligned}$$

Rechenregeln II

$$\operatorname{Cov}(X,Y)^2 \leq \operatorname{Var}(X)\operatorname{Var}(Y)$$

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y)$$

Sind X, Y unabhängig, so gilt:

$$\begin{aligned} \mathbb{E}(XY) &= \mathbb{E}(X)\mathbb{E}(Y) \\ \operatorname{Cov}(X,Y) &= 0 \end{aligned}$$

Kovarianzmatrix

Sei $X = (X_1, \dots X_n)$ eine mehrdimensionale Zufallsvariable. Dann ist ihre Kovarianzmatrix gegeben durch

$$\begin{aligned}
\mathrm{Cov}(X) &= \mathbb{E}((X - \mathbb{E}(X)) \cdot (X - \mathbb{E}(X))^T) \\
&= \mathbb{E}(XX^T) - \mathbb{E}(X)\mathbb{E}(X)^T \\
&= \begin{pmatrix} \mathrm{Cov}(X_1, X_1) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \cdots & \mathrm{Cov}(X_n, X_n) \end{pmatrix}
\end{aligned}$$

Die Kovarianzmatrix ist

- symmetrisch
- positiv semi-definit

Die Normalverteilung

Eine Zufallsvariable $X \in \mathbb{R}$ ist normalverteilt, $X \sim \mathcal{N}(\mu, \sigma^2)$, wenn

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

ihre Dichtefunktion ist.

$$\mathbb{E}(X) = \mu$$

$$\mathrm{Var}(X) = \sigma^2$$

Die Normalverteilung

- Zentraler Grenzwertsatz : Die Summe einer großen Zahl von unabhängigen Zufallsvariablen ist annähernd normalverteilt.
- Die Summe normalverteilter Zufallsvarialben ist wieder normalverteilt
- Paarweise unkorrelierte normalverteilte Zufallsvariablen sind auch unabhängig.