

Wahrscheinlichkeitsrechnung

Irene Winkler

Arbeitsgruppe Maschinelles Lernen

Zufallsvariablen

Erwartungswert und Kovarianz

Normalverteilung

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 $A \cap B = \emptyset$

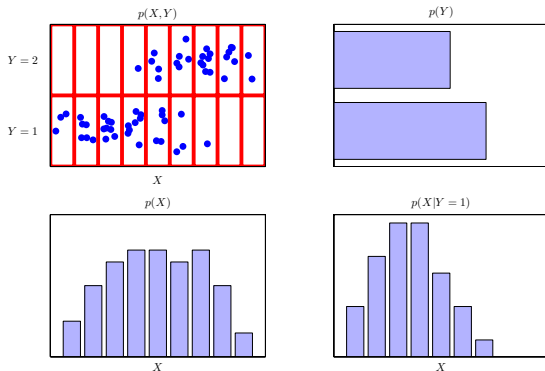
Für unsere Zwecke brauchen wir

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

Diskrete Zufallsvariablen

$X \in \{x_1, \dots, x_n\}$, $Y \in \{y_1, \dots, y_m\}$ 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

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

Kontinuierliche Zufallsvariablen

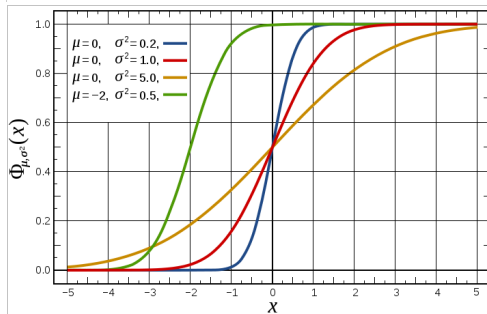
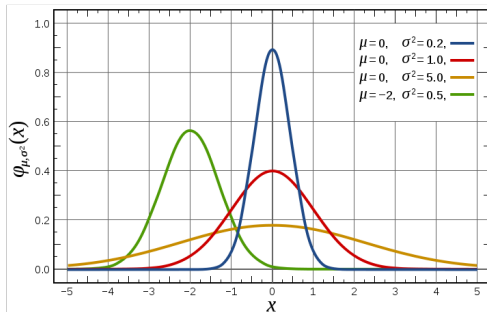
Sei $X \in \mathbb{R}$ eine reellwertige Zufallsvariable. Die (kumulative) *Verteilungsfunktion* $F_X : \mathbb{R} \mapsto \mathbb{R}$ ist definiert durch

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

Es gilt

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

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



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 der Funktion einer Zufallsvariablen $f(X)$

- ▶ Ist X eine diskrete Zufallsvariable so ist der Erwartungswert

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

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

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

- ▶ 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

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

- ▶ Die *Standardabweichung* von X ist definiert als

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

Kovarianz

Seien X und Y zwei eindimensionale Zufallsvariablen.

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

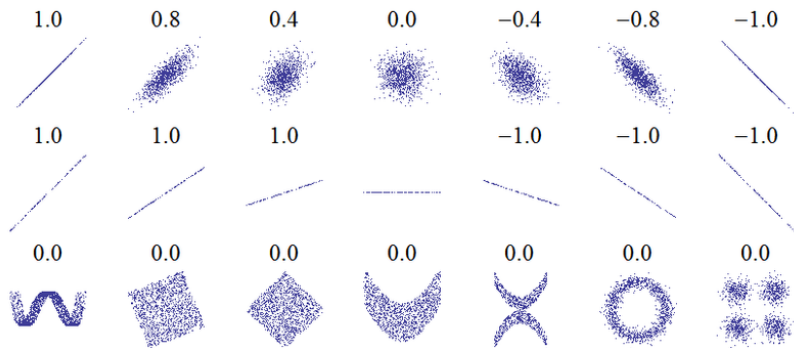
$$\text{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

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

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

Es gilt $-1 \leq \text{Corr}(X, Y) \leq 1$.

Maß für den linearen Zusammenhang



Rechenregeln I

► Verschiebungssatz

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

► Die Kovarianz ist bilinear

$$\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y)$$

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

$$\text{Cov}(X + Z, Y) = \text{Cov}(X, Y) + \text{Cov}(Z, Y)$$

Rechenregeln II

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

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

Sind X, Y unabhängig, so gilt:

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

$$\text{Cov}(X, Y) = 0$$

Kovarianzmatrix

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

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

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.

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

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

Die Normalverteilung

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