

Algorithms for the Social Web

Simon Ginzinger / Markus Zanker

2. Vorlesung

Bewertungsschema

- Punkte aller Tasks werden summiert

Notenschema		
von	bis	Note
0%	49%	5
50%	66%	4
67%	80%	3
81%	91%	2
92%	100%	1
Beispiel:		
Gesamtpunkte	40	Note
0	19	5
20	26	4
27	32	3
33	36	2
37	40	1

Grundlegendes zur Wahrscheinlichkeitsrechnung

Ein **Zufallsexperiment** ist ein Vorgang, der

- unter ***gleichen*** Bedingungen beliebig oft wiederholt werden kann.
- dessen Ergebnis nicht mit Sicherheit vorhergesagt werden kann.

Theoretisches Konstrukt
ABER
Relevante Bezüge zur Praxis

Grundlegendes zur Wahrscheinlichkeitsrechnung

Die Menge aller möglichen (sich gegenseitig ausschließenden) Ergebnisse eines Zufallsexperimentes wird Ergebnisraum genannt und üblicherweise mit Ω bezeichnet.

Beispiele:

- **Würfel:** $\Omega = \{1, 2, 3, 4, 5, 6\}$
- **Lotto:** $\Omega = \{ \{x_1, \dots, x_6\} \mid x_i \in \{1, \dots, 45\} \wedge \forall j \neq i: x_i \neq x_j \wedge i, j \in \{1, \dots, 6\} \}$

Grundlegendes zur Wahrscheinlichkeitsrechnung

Ein Ereignis A ist eine Teilmenge von Ω . Das Ereignis A ist eingetreten, wenn das Ergebnis des Zufallsexperiments ein Element der Menge A ist. Die einelementigen Teilmengen von Ω sind die möglichen Ergebnisse des Experimentes und werden auch *Elementarereignisse* genannt.

Man nennt:

- Ω das sichere Ereignis
- $\{\}$ das unmögliche Ereignis

Mehrere Ereignisse

Seien $A, B \in \Omega$ zwei Ereignisse:

- Die Aussage: "*Mindestens eines der beiden Ereignisse tritt ein*" entspricht der Menge $A \cup B$
- Die Aussage "*Beide Ereignisse treten ein*" entspricht der Menge $A \cap B$
- Beispiel (Würfeln):
 - $A = \text{"Gerade Zahl"} = \{2, 4, 6\}$
 - $B = \text{"Mehr als 3"} = \{4, 5, 6\}$

Gegenereignis

$$\bar{A} = \Omega \setminus A$$

Unvereinbarkeit von Ereignissen

Seien $A, B \in \Omega$ zwei Ereignisse:

- A und B sind ***unvereinbar***, wenn $A \cap B = \{\}$
- Beispiel (Würfeln):
 - $A = \text{"6"} = \{6\}$
 - $B = \text{"Höchstens 3"} = \{1, 2, 3\}$

Laplace-Experiment und Wahrscheinlichkeit

- **Endlich viele** mögliche Elementarereignisse
- Jedes Elementarereignis ist **gleich wahrscheinlich**

Unter der Laplace-Annahme kann man die Wahrscheinlichkeit eines Ereignisses $P(A)$ als

$$P(A) = \frac{|A|}{|\Omega|}$$

berechnen.

Laplace Annahme
hält nicht im
Allgemeinen.

Verbindung zur Statistik

Die **Statistik** versucht $P(A)$ durch die Analyse von Stichproben möglichst genau abzuschätzen.

Beispiel: Abschätzen der Wahrscheinlichkeit, dass ein Benutzer eine bestimmte Bewertung abgibt.

Eigenschaften von $P(A)$

- $0 \leq P(A) \leq 1$
- $P(\Omega) = 1$
- $P(A \cup B) = P(A) + P(B)$ falls $A \cap B = \{\}$
- $P(\bar{A}) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n)$ falls $A_i \cap A_j = \{\}$ für $i \neq j$
- $P(A) \leq P(B)$ für $A \subseteq B$



Was ist mit $P(\{\})$?

Beispiel: Geburtstagsparadoxon

Wie viele Menschen braucht man damit die Wahrscheinlichkeit, dass zwei von Ihnen am gleichen Tag im Jahr Geburtstag haben größer als 50% ist?

Genaue Beschreibung
des Zufallsexperimentes
ist essentiell!

Analogie: Hashing

Beispiel: Schätzen von W-Keiten

- W-keiten für Alter im Master MMT

Beispiel: Anders verteilte Zufallszahlen aus gleichverteilten Zufallszahlen

- *Inversionsmethode*

Task 2 (50 Punkte)

Deadline 6.11.12 – 10:00 AM

- Identifizieren Sie die 50 Benutzer mit den meisten Bewertungen ($\neq 0$).
- Für diese 50 Benutzer:
 - Schätzen Sie die Wahrscheinlichkeit für jede Bewertung (1 bis 10) durch Berechnung der entsprechenden relativen Häufigkeit.
 - Ersetzen sie nun die fehlenden Werte durch Zufallswerte, die der geschätzten Verteilung folgen.
 - Identifizieren das Paar von Benutzern mit der höchsten paarweisen Pearson-Korrelation.
- Laden Sie die erstellte Daten-Matrix, die Matrix der paarweisen Korrelationskoeffizienten (50×50), den Source Code und einen Report ins Wiki.

Format:

- Report -> PDF
- Matrizen -> ";"-separated CSV
- Source Code -> Text File(s)
- Hochladen als **ein** Zip-File (siehe Wiki)