Formale Grundlagen: Einführung

Yannick Versley

SoSe 2014

Überblick

- Zeit und Ort: Mi 14-16, SR 19 (INF 306)
- Webseite und weitere Materialien:
  `www.cl.uni-heidelberg.de/courses/ss15/fg/`
- Voraussetzungen:
  keine (empfohlen: Logik, ECL)
- Arbeitsaufwand:
  6 LP $\Rightarrow$ ca. 12h/Woche

Tutorien

Tutoren:

- Julian Hitschler
- Maximilian Müller-Eberstein

…verantwortlich für: Tutorien (Teilnahme empfohlen), Korrektur/Benotung der Übungsaufgaben

Ziele

Die “Formale-Grundlagen”-Vorlesung soll dabei helfen, dass Sie

- **Mathematische Erklärungen/Beweise** in Artikeln zu Computerlinguistischen Themen lesen können und eine grobe Idee vom Inhalt bekommen
- **Eigene Probleme** mit Hilfe von mathematischen Konzepten wie Vektoren, Wahrscheinlichkeitsverteilungen, linearen Abbildungen (z.B. Matrizen) modellieren und lösen können.
- Mathematische **Verfahren der Computerlinguistik** wie gerichtete und ungerichtete graphische Modelle (z.B. HMMs, PCFG) kompetent auf typische Probleme anwenden können.
- mit geeigneten **Maßen und Statistiken** Aussagen über Verteilungen in großen Datenmengen machen können (Informationstheorie, Signifikanztests).

Formale Abstraktionen

In einem Formalismus:

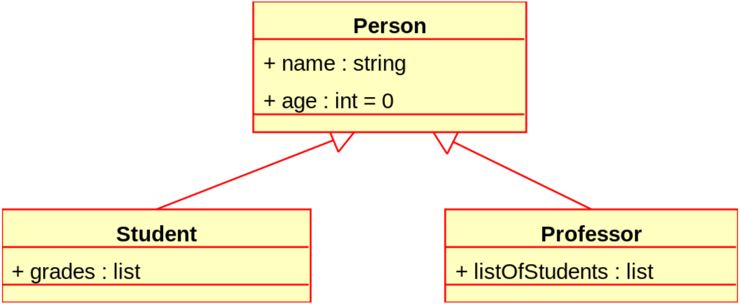
- stelle ich Dinge (Zahlen, Datenstruturen, Personen) durch Symbole dar, die bestimmten Regeln genügen
- sage ich etwas über diese Dinge, indem ich diese Symbole gemäß der (Spiel-)regeln bearbeite

Mathematische Formalismen:

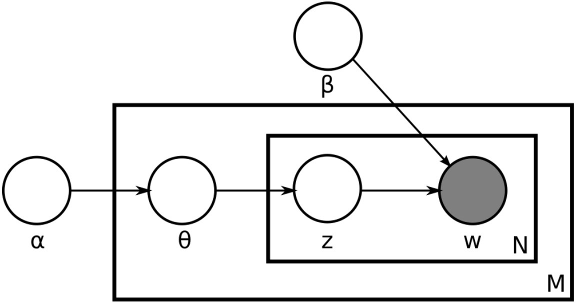
- ... basieren auf (Prädikaten-)Logik
- Beweise (logische Aussagen über Dinge bestimmter Art)
- "Ausrechnen" (Übertragen vom Abstrakten auf konkretere Objekte)

Terence Tao: *The point of rigour is <u>not</u> to destroy all intuition; instead it should be used to destroy <u>bad</u> intuition while clarifying and elevating <u>good</u> intuition.*

Grafische Formalismen: UML



Grafische Formalismen: Probabilistische Modelle



Mathematische Probleme lösen

- **Abstraktion**: Welchen Teil des Problems kann ich mathematisch modellieren?
- **Modellierung**: Welche Eigenschaften hat das (mathematische) Modell, das ich von meinem Problem gemacht habe?
- **Implementation**: Wie kann ich nützliche Inferenzen für das reale Problem machen?
  Unser Brot- und Butter-Werkzeug hier:
  numpy (Numerical Python)

Mathematische Verfahren der Computerlinguistik

- Naïve Bayes, Markov-Modelle
- HMMs
- WFSA, WFST
- und noch mehr!

In dieser Vorlesung interessieren uns *allgemeine* Techniken, um mit dieser Art Modell (*probabilistic graphical model*) umzugehen.

Deskriptive Statistiken

Signifikanztests

- Funktionieren HMMs bei meinem Problem (signifikant) besser als Entscheidungsbäume?

Entropie / Cross-Entropie / Divergenz

- Enthalten morphologische Tags mehr Information als POS-Tags, wenn ich das Lemma schon weiss?

Kriterien

- Erfolgreiche Bearbeitung der Übungsaufgaben (Mindestpunktzahl)
- Klausur am Ende des Semesters

Axiome vs. Sätze/Korollare (1)

Mathematische Aussagen sind **Tautologien**

- Aussagen die immer (in jeder Welt) wahr sind
- Die meisten interessanten mathematischen Aussagen haben die Form:
  Wenn leicht überprüfbare Aussage, dann interessante Aussage

Üblicherweise verpackt man die leicht überprüfbaren Aussagen in **Definitionen** mathematischer Begriffe (alternativ: man definiert Begriffe durch **Axiome**)

Axiome vs. Sätze/Korollare

- Oft lassen Leute die leicht überprüfbare Aussage oder Teile davon weg (weil die Definition von reellen Zahlen bekannt ist, oder weil alle vorkommenden Zahlen relle Zahlen sein sollen)
- Das spart Platz
- Das stiftet Verwirrung, wenn man als Leser nicht weiss, welche Voraussetzungen gemeint sind

Alte Bekannte: Mengen

Eine **Menge** ist eine mathematische Struktur, in der Dinge enthalten sein können (sie sind dann **Element** einer Menge, etwa: $x \in A$) oder auch nicht.

$$\begin{array}{ccc} A & = & \{1,2,4,7,12\} \\ 1 & \in & A \\ 3 & \notin & A \end{array}$$

Mengen, die endlich viele Elemente haben, sind **endlich**. Manche Mengen sind **unendlich**.

Schon bekannt: Tupel

Ein *n*-Tupel (Paar, Tripel, ...) besteht aus *n* Stellen, in denen etwas steht.

$$(1, 2),\ (1, 4, 12)\ (1, \{2, 3\})$$

Wir können Tupel dadurch beschreiben, dass wir schreiben, aus welcher Menge jede Stelle des Tupels kommt:

$$(1, 2, q) \in A \times A \times \{q, r, s\}$$

Die Menge $A \times B$ von Tupeln mit einer Stelle aus A und einer Stelle aus B heisst *kartesisches Produkt* von A und B.

Alte Bekannte: Untermengen, Relationen

Eine Menge ist **Untermenge** einer anderen Menge, wenn jedes Element der einen Menge auch ein Element der anderen Menge ist:

$$\begin{aligned} \{1,4,12\} &\subseteq A \\ \{1,3,12\} &\not\subseteq A \end{aligned}$$

Eine Menge von *n*-Tupeln ist eine *n*-stellige **Relation**

$$\{(1,2),(1,12)\} \subseteq A \times A$$

Alte Bekannte: Funktionen

Eine Funktion $f : A \to B$ bildet jedes Element der Menge A auf ein Element aus der Menge B ab.

$$f : \begin{array}{rcl} 1 & \mapsto & 2 \\ 2 & \mapsto & 12 \\ 3 & \mapsto & 1 \end{array}$$

Eine n-stellige Funktion $f : A \times B \to C$ bildet jedes Tupel aus der Menge $A \times B$ auf einen Wert aus C ab.

Mathematische Struktur: Eine Gruppe

Definition: $(G, \odot)$ ist eine Gruppe, wenn folgendes gilt:

- $\odot : G \times G \to G$
  G ist **abgeschlossen** bezüglich $\odot$
- $\forall x, y, z \in G : (x \odot y) \odot z = x \odot (y \odot z)$
  $\odot$ ist **assoziativ**
- $\exists e \in G : a \odot e = e \odot a = a$
  Es gibt ein **neutrales Element** e
  Warum ist das neutrale Element eindeutig?
- $\forall x \in G : \exists x^{-1} \in G : x \odot x^{-1} = x^{-1} \odot x = e$
  (mit dem neutralen Element e)
  Jedes Element hat ein **inverses Element**

Beispiele für Gruppen

- $(\mathbb{Z}, +)$ (mit 0 als neutrales Element)
- $(\mathbb{R} \setminus 0, \cdot)$
- $(\mathbb{Z}_n, +)$ (Zahlen modulo einer bestimmten Zahl)

Keine Gruppe:

- $(\mathbb{B}, \rightarrow)$ (kein neutrales Element)
- $(\mathbb{B}, \wedge)$ (kein Inverses)
- $(\mathbb{N}_0, -)$ (nicht abgeschlossen)
- $(\Sigma^*, \circ)$ (Zeichenketten: kein Inverses)

Andere mathematische Strukturen (1)

- **Halbgruppe** (nur assoziative Operation)
- **Monoid** Gruppe ohne Inverses
- **Körper** $(\mathbb{K}, +, \cdot)$:
  $(\mathbb{K}, +)$ ist eine kommutative Gruppe
  mit neutralem Element 0
  $(\mathbb{K} \setminus 0, \cdot)$ ist eine kommutative Gruppe
  (mit neutralem Element 1)
  $\cdot$ distribuiert über $+$, d.h. $(a + b) \cdot c = a \cdot c + b \cdot c$
  Bsp.: $\mathbb{Q}, \mathbb{R}, \mathbb{C}$

Andere mathematische Strukturen (2)

- **Halbring** $(G, \oplus, \otimes)$:
  $(G, \oplus)$ ist eine kommutative Halbgruppe
  (d.h. $a + b = b + a$ für alle $a, b \in G$)
  $(G, \otimes)$ ist eine Halbgruppe
  $\otimes$ distribuiert über $+$, d.h.

$$\begin{aligned} (a \oplus b) \otimes c &= a{\otimes}c \oplus b{\otimes}c \\ c \otimes (a \oplus b) &= c{\otimes}a \oplus c{\otimes}b \end{aligned}$$

Beispiel: für die Menge der Sprachen über ein Alphabet $\mathcal{L} = \mathcal{P}(\Sigma^*)$ ist $(\mathcal{L}, \cup, +\!\!+)$ ein Halbring

Axiome für $\mathbb{R}$

- $(\mathbb{R}, +, \cdot)$ ist ein Körper.

Axiome für $\mathbb{R}$

- $(\mathbb{R}, +, \cdot)$ ist ein Körper.
- …und wird durch $\leq$ **geordnet**:
  - Wenn $a \leq b$, dann auch $a + c \leq b + c$
  - Wenn $0 \leq a$ und $0 \leq b$, dann auch $0 \leq a \cdot b$

Axiome für $\mathbb{R}$

- $(\mathbb{R}, +, \cdot)$ ist ein Körper.
- ...und wird durch $\leq$ **geordnet**:
  - Wenn $a \leq b$, dann auch $a + c \leq b + c$
  - Wenn $0 \leq a$ und $0 \leq b$, dann auch $0 \leq a \cdot b$

Jeder geordnete Körper enthält $\mathbb{Q}$ als geordneter Teilkörper (mit $\leq_{\mathbb{Q}}$).

Axiome für $\mathbb{R}$

- $(\mathbb{R}, +, \cdot)$ ist ein Körper.
- …und wird durch $\leq$ **geordnet**:
  - Wenn $a \leq b$, dann auch $a + c \leq b + c$
  - Wenn $0 \leq a$ und $0 \leq b$, dann auch $0 \leq a \cdot b$

  Jeder geordnete Körper enthält $\mathbb{Q}$ als geordneter Teilkörper (mit $\leq_{\mathbb{Q}}$).
- Eine Menge A von Elementen hat eine **obere Schranke** m, wenn gilt: für alle $a \in A$ gilt $a \leq m$.

  Das **Supremum** ist die kleinste obere Schranke. In $\mathbb{Q}$ hat die Menge $\{x | x^2 < 2\}$ obere Schranken (sie ist beschränkt), aber hat kein Supremum.

Axiome für $\mathbb{R}$

- $(\mathbb{R}, +, \cdot)$ ist ein Körper.
- ...und wird durch $\leq$ **geordnet**:
  - Wenn $a \leq b$, dann auch $a + c \leq b + c$
  - Wenn $0 \leq a$ und $0 \leq b$, dann auch $0 \leq a \cdot b$

  Jeder geordnete Körper enthält $\mathbb{Q}$ als geordneter Teilkörper (mit $\leq_{\mathbb{Q}}$).
- Eine Menge A von Elementen hat eine **obere Schranke** m, wenn gilt: für alle $a \in A$ gilt $a \leq m$.

  Das **Supremum** ist die kleinste obere Schranke. In $\mathbb{Q}$ hat die Menge $\{x | x^2 < 2\}$ obere Schranken (sie ist beschränkt), aber hat kein Supremum.
  Jede nach oben beschränkte Teilmenge von $\mathbb{R}$ besitzt ein Supremum in $\mathbb{R}$ (**Vollständigkeit**).

Ein Tableaubeweis mit $\mathbb{R}$

Für welche $x \in \mathbb{R}$ ist $|x^2 - 1| < a$?

Definition von $|\cdot| : \mathbb{R} \to \mathbb{R}$:

- Für $x < 0$ ist $|x| = -x$
- Für $x \geq 0$ ist $|x| = x$

Spaß mit Monoiden

Sei $(M, \cdot)$ ein Monoid und $f : \Sigma \to M$.
Dann können wir eine Abbildung $f^* : \Sigma^* \to M$ konstruieren durch

$$\begin{array}{rcll} f^*(\varepsilon) & = & 0_M & \\ f^*(a \circ b) & = & f(a) \cdot f^*(b) & \text{f.a. } a \in \Sigma, b \in \Sigma^* \end{array}$$

(dann gilt automatisch: $f^*(a \circ b) = f^*(a) \cdot f^*(b)$ (für $a, b \in \Sigma^*$)

Beispiel: Mit $f(\mathrm{e}) = 1$, $f(s) = 0$ sonst können wir so (a) die Anzahl der e in einem Wort bestimmen, (b) schauen, ob es *nur* e in einem Wort gibt. *Welches Monoid?*

Was steckt dahinter? (1)

Ein **Monoid-Homomorphismus** von einem Monoid $(G, \oplus)$ zu einem Monoid $(H, \otimes)$ ist eine Abbildung $f : G \rightarrow H$ mit den folgenden Eigenschaften:

- für alle $a, b \in G$ gilt: $f(a \oplus b) = f(a) \otimes f(b)$
- Seien e_G das neutrale Element von $(G, \oplus)$ und e_H das neutrale Element von $(H, \otimes)$. Dann gilt $f(e_G) = e_H$

Anmerkung 1: mit $H = \mathbb{Z}, \quad \oplus = \cdot$ und $f(e_G) = 0$ (d.h. $f(a) = 0$ für alle $a \in G$) hätten wir eine Abbildung, die das erste, aber nicht das zweite Kriterium erfüllt.

Was steckt dahinter? (2)

Ein **Monoid-Homomorphismus** von einem Monoid $(G, \oplus)$ zu einem Monoid $(H, \otimes)$ ist eine Abbildung $f : G \rightarrow H$ mit den folgenden Eigenschaften:

- für alle $a, b \in G$ gilt: $f(a \oplus b) = f(a) \otimes f(b)$
- Seien e_G das neutrale Element von $(G, \oplus)$ und e_H das neutrale Element von $(H, \otimes)$. Dann gilt $f(e_G) = e_H$

Anmerkung 2: Für jede Abbildung f von einem Alphabet Σ in ein Monoid H gibt es genau einen Monoid-Homomorphismus $f^* : \Sigma^* \rightarrow H$ mit $f^*(a) = f(a)$ für alle $a \in \Sigma$.

Was steckt dahinter? (3)

Homomorphismen (strukturerhaltende Operationen) gibt es für viele mathematische Strukturen (z.B. Monoide und Halbringe).

Halbring-Homomorphismen sind z.B. nützlich, um endliche Sprachen (d.h. den Halbring $(\mathcal{L}_{\mathit{fin}}, \cup, +\!\!+)$) auf andere Halbringe (etwa: $(\mathbb{R}, +, \cdot)$, $(\mathbb{R}, +, \max)$) abzubilden.

iPython und numpy

(Beispiel mit iPython-Notebook, Arrays und loadtxt)

Zusammenfassung

Was wir gesehen haben:

- Grafische Formalismen, Axiome / Definitionen / Beweise
- Mathematische Strukturen:
  Halbgruppe / Monoid / Gruppe $(G, \oplus)$
  Halbring / Körper $(G, \oplus, \otimes)$
- Axiomatisierung von $\mathbb{R}$
- Monoid-Homomorphismen (über Σ^*)
- iPython notebook, numpy.array

Was Sie können sollten:

- Den Unterschied zwischen (i) grafisch/textuell und (ii) formal/informell erklären können
- Formale Beweise führen
- Ein Monoid/einen Halbring erkennen
- Monoid-Homomorphismen definieren und nutzen

... und übrigens

Danke an alle, die im SoSe 14 Korrekturen zu den Folien gefunden haben:

- Mareike Hartmann
- Angela Schneider
- Robert Schütz

Zum Nachlesen:

- Jänich, Kapitel 1 (*Mengen und Abbildungen*)