Mathematische Grundlagen der Computerlinguistik

Lineare Algebra

Dozentin: Wiebke Petersen

10. Foliensatz

Einleitung

Die lineare Algebra beschäftigt sich mit

- Vektorräumen
- linearen Abbildungen zwischen Vektorräumen
- linearen Gleichungssystemen
- Matrizen

Gruppen (Wiederholung)

Eine Gruppe $\mathbf{G} = (G, \otimes)$ ist eine algebraische Struktur mit:

G1 $\otimes$ ist assoziativ

G2 G enthält ein neutrales Element

G3 jedes Element aus G hat ein inverses Element in G.

G heißt abelsche Gruppe, wenn die Verknüpfung $\otimes$ kommutativ ist.

Körper

Ein Körper $\mathbf{K} = (K, +, \cdot)$ ist eine algebraische Struktur mit zwei Verknüpfungen $+ : K \times K \to K$ (Addition) und $\cdot : K \times K \to K$ (Multiplikation):

K1 $(K, +)$ ist eine abelsche Gruppe.
Das neutrale Element der Addition wird mit 0 bezeichnet und das zu a inverse Element mit $-a$.

K2 Die Multiplikation lässt sich auf $K \setminus \{0\}$ beschränken (für $a, b \in K \setminus \{0\}$ gilt $a \cdot b \in K \setminus \{0\}$), und $K \setminus \{0\}$ zusammen mit der Multiplikation bildet eine abelsche Gruppe.
Das neutrale Element der Multiplikation wird mit 1 bezeichnet und das zu $a \in K \setminus \{0\}$ inverse Element mit a^{-1}.

K3 Es gelten die folgenden Distributivgesetze:
$a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ und $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$.

Körper: Anmerkungen und Beispiele

- Statt $a \cdot b$ schreibt man zumeist ab.
- Per Konvention bindet die Multiplikation stärker als die Addition, es gilt also
  $(ab) + c = ab + c$.
- Statt $a + (-b)$ schreibt man zumeist $a - b$.
- Es gelten folgende Zusammenhänge (versuchen sie die Aussagen zu beweisen):
  - $0 \cdot a = a \cdot 0 = 0$ für alle $a \in K$.
  - Ist $ab = 0$ so gilt $a = 0$ oder $b = 0$.
  - Für alle $a, b \in K$ gilt $a(-b) = (-a)b = -(ab)$. Außerdem gilt $(-a)(-b) = ab$.

Beispiele:

- $(\mathbf{Q}, +, \cdot)$ ist ein Körper.
- $(\mathbb{R}, +, \cdot)$ ist ein Körper.
- Aber $(\mathbf{Z}, +, \cdot)$ ist kein Körper.

Vektorraum

Sei K ein Körper. Ein K-Vektorraum ist ein Tripel $(V, +, \cdot)$ bestehend

- aus einer Menge V
- einer Verknüpfung $+ : V \times V \to V$ (Vektoraddition) und einer
- einer Verknüpfung $\cdot : K \times V \to V$ (skalare Multiplikation)

So dass folgendes gilt:

V1 $(V, +)$ ist eine kommutative Gruppe (auch abelsche Gruppe). Das neutrale Element 0 heißt Nullvektor und das zu $v \in V$ inverse Element $-v$ heißt der zu v negative Vektor

V2 für alle $v, w \in V$ und $\lambda, \mu \in K$ gilt:

- $(\lambda + \mu) \cdot v = (\lambda \cdot v) + (\mu \cdot v)$
- $\lambda \cdot (v + w) = (\lambda \cdot v) + (\lambda \cdot w)$
- $(\lambda\mu) \cdot v = \lambda \cdot (\mu \cdot v)$
- $1 \cdot v = v$.

Die Elemente von V nennt man Vektoren, die von K Skalare.

Vektorraum: Anmerkungen

- Statt $\lambda \cdot v$ mit $\lambda \in K$ und $v \in V$ schreibt man zumeist λv
- Per Konvention bindet die skalare Multiplikation stärker als die Addition in V und die Addition in K, also $\lambda v + w = (\lambda v) + w$
- In jedem K-Vektorraum gilt:
  - $0 \cdot v = 0$ für alle $v \in V$
  - $\lambda \cdot 0 = 0$ für alle $\lambda \in K$
  - Wenn $\lambda \in K$ und $v \in V$ und $\lambda v = 0$, dann gilt $\lambda = 0$ oder $v = 0$.
  - $(-1) \cdot v = -v$ für alle $v \in V$.

  Können sie diese Aussagen beweisen?

Beweise einiger Eigenschaften von K-Vektorräumen

Sei $(V, +, \cdot)$ ein K-Vektorraum.

- Aussage: $0 \cdot v = 0$ für alle $v \in V$
  Beweis: es gilt $0 \cdot v = (0 + 0) \cdot v = 0 \cdot v + 0 \cdot v$ und $0 \cdot v + 0 = 0 \cdot v$. Daraus folgt $0 \cdot v = 0$. $\square$
- Aussage: $\lambda \cdot 0 = 0$ für alle $\lambda \in K$
  Beweis: es gilt $\lambda \cdot 0 = \lambda \cdot (0 + 0) = \lambda \cdot 0 + \lambda \cdot 0$ und $\lambda \cdot 0 + 0 = \lambda \cdot 0$. Daraus folgt $\lambda \cdot 0 = 0$. $\square$
- Aussage: Wenn $\lambda \in K$ und $v \in V$ und $\lambda \cdot v = 0$, dann gilt $\lambda = 0$ oder $v = 0$.
  Beweis durch Fallunterscheidung: 1. Fall: Wenn $\lambda = 0$ und $\lambda \cdot v = 0$ dann folgt die Aussage direkt. 2. Fall: Wenn $\lambda \neq 0$ und $\lambda \cdot v = 0$, dann gilt $v = 1 \cdot v = (\lambda^{-1}\lambda) \cdot v = \lambda^{-1}(\lambda \cdot v) = \lambda^{-1} \cdot 0 = 0$. $\square$
- Aussage: $(-1) \cdot v = -v$ für alle $v \in V$.
  Beweis: aus $v + (-1) \cdot v = 1 \cdot v + (-1) \cdot v = (1 - 1) \cdot v = 0 \cdot v = 0$ folgt $(-1) \cdot v = -v$. $\square$

Vektorraum Beispiel: Vektorraum der Funktionen

Setzt man $K = \mathbb{R}$ und $V = \{f \mid f \text{ ist eine Funktion mit } f : \mathbb{R} \to \mathbb{R}\}$, dann bildet $(V, +, \cdot)$ einen $\mathbb{R}$-Vektorraum mit

- Vektoraddition $V \times V \to V$:
  $(f + g) : \mathbb{R} \to \mathbb{R}$ mit $(f + g)(x) = f(x) + g(x)$
- skalare Multiplikation $\mathbb{R} \times V \to V$:
  $(\lambda \cdot f) : \mathbb{R} \to \mathbb{R}$ mit $(\lambda \cdot f)(x) = \lambda f(x)$

Vektorraum Beispiel: Koordinatenraum oder Raum der geordneten n-Tupel

Ist K ein Körper und $n \in \mathbf{N}$, so bildet das n-fache kartesische Produkt $K^n = \{(x_1, \ldots, x_n) \mid x_1, \ldots, x_n \in K\}$ die Menge aller n-Tupel $(x_1, \ldots, x_n)$ mit Komponenten aus K. $(K^n, +, \cdot)$ bildet einen K-Vektorraum mit

- Vektoraddition $+ : K^n \times K^n \to K^n$ definiert durch

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$

- und skalare Multiplikation $\cdot : K \times K^n \to K^n$ definiert durch

$$a \cdot (x_1, \ldots, x_n) = (a \cdot x_1, \ldots, a \cdot x_n)$$

Der Vektorraum $(K^n, +, \cdot)$ wird auch als Koordinatenraum der Dimension n oder als Raum der geordneten n-Tupel bezeichnet.

Vektoren in Koordinatenräumen

Koordinatenvektoren notiert man häufig als Spaltenvektoren. Für das Tupel $(x_1, \ldots x_n)$ schreibt man dann

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Für die Vektoraddition gilt dann:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

und für die skalare Multiplikation gilt

$$a \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a \cdot x_1 \\ \vdots \\ a \cdot x_n \end{pmatrix}$$

Skalarprodukt in $\mathbb{R}^n$

Für $v = (v_1, \ldots v_n)$ und $w = (w_1, \ldots w_n)$ mit $v, w \in \mathbb{R}^n$ ist das Skalarprodukt $\langle v, w \rangle$ definiert als

$$\langle v, w \rangle = v_1 w_1 + v_2 w_2 + \ldots v_n w_n.$$

Das Skalarprodukt zweier Vektoren ist ein Skalar:

$$\langle\ ,\ \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

Die Länge oder Norm eines Vektors $v \in \mathbb{R}^n$ ist

$$\|v\| = \sqrt{\langle v, v \rangle}$$

Die Distanz zweier Vektoren $v, w \in \mathbb{R}^n$ ist

$$d(v, w) = \|w - v\|$$

Zusammenhang zwischen Skalarprodukt und Winkel

Seien $v, w \in \mathbb{R}^n$ und $\theta = \sphericalangle(v, w)$ der Winkel zwischen v und w, dann gilt

$$\langle v, w \rangle = \|v\| \|w\| \cos(\theta)$$

Für den Beweis benötigt man den Kosinussatz (`https://de.wikipedia.org/wiki/Kosinussatz`): Seien a, b, c die Seiten eines Winkels und γ der der Seite c gegenüberliegende Winkel, dann gilt

$$c^2 = a^2 + b^2 - 2ab\cos(\gamma)$$

Vektorraum Beispiel: Matrizenraum (1)

Ein rechteckiges Schema

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$$

von Elementen a_{ij} aus einem Körper K heißt Matrix. Die Elemente a_{ij} heißen Komponenten der Matrix. Man spricht von den Zeilen und Spalten einer Matrix. Eine Matrix mit m Zeilen und n Spalten heißt $m \times n$-Matrix (lese: m Kreuz n Matrix).

Vektorraum Beispiel: Matrizenraum (2)

Sei K ein Körper und $m, n \in \mathbb{N}$. Dann ist

$$K^{m \times n} = \{(a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n} \mid a_{ij} \in K\}$$

die Menge der Matrizen der Größe $m \times n$ mit Komponenten aus K. $K^{m \times n}$ bildet mit der folgenden Vektoraddition und skalaren Multiplikation einen K-Vektorraum:

- Vektor- oder besser Matrizenaddition:
  $+ : K^{m \times n} \times K^{m \times n} \to K^{m \times n}$ mit $A + B = (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})$.
- Multiplikation mit einem Skalar:
  $\cdot : K \times K^{m \times n} \to K^{m \times n}$ mit $\lambda \cdot A = \lambda \cdot (a_{ij}) = (\lambda \cdot a_{ij})$.

Der Vektorraum $(K^{m \times n}, +, \cdot)$ wird auch der Matrizenraum oder Raum der Matrizen der Größe $m \times n$ über dem Körper K genannt.

Die Vektorraumeigenschaften von $(K^{m \times n}, +, \cdot)$ folgen unmittelbar aus den Vektorraumeigenschaften von $(K^n, +, \cdot)$. Das neutrale Element der Matrizenaddition ist die Matrix (a_{ij}) mit $a_{ij} = 0$ für alle i, j. Die zu $(a_{ij}$ additiv inverse Matrix ist $(-a_{ij})$.

Beispiel: Matrizenaddition und Multiplikation mit Skalar

Sei im folgenden $K = \mathbb{R}$.

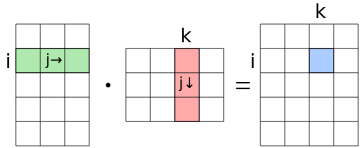
$$\begin{pmatrix} 3 & 2 & 4 \\ 0 & 1 & -1 \end{pmatrix} + \begin{pmatrix} 1 & 3 & -2 \\ 2 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 3+1 & 2+3 & 4-2 \\ 0+2 & 1+0 & -1+1 \end{pmatrix} = \begin{pmatrix} 4 & 5 & 2 \\ 2 & 1 & 0 \end{pmatrix}$$

$$3 \cdot \begin{pmatrix} 3 & 2 & 4 \\ 0 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 3\cdot 3 & 3\cdot 2 & 3\cdot 4 \\ 3\cdot 0 & 3\cdot 1 & 3\cdot -1 \end{pmatrix} = \begin{pmatrix} 9 & 6 & 12 \\ 0 & 3 & -3 \end{pmatrix}$$

Matrizenmultiplikation

Die Matrizenmultiplikation ist die Operation $\cdot K^{m\times p} \times K^{p\times n} \rightarrow K^{m\times n}$ mit $A \cdot B = C$ und

$$c_{ik} = \sum_{j=1}^{p} a_{ij} \cdot b_{jk}$$



Beispiel:

$$\begin{pmatrix} 3 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 4 & 0 \end{pmatrix} = \begin{pmatrix} 7 & 8 \\ 9 & 2 \end{pmatrix}$$

transponierte Matrix

Die Transponierte einer $m \times n$-Matrix $A = (a_{ij})$ ist die $n \times m$-Matrix $A^T = (a_{ji})$. Also

$$A^T = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}^T = \begin{pmatrix} a_{11} & \dots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \dots & a_{mn} \end{pmatrix}$$

Es gelten folgende Aussagen:

- $(A + B)^T = A^T + B^T$
- $(c \cdot A)^T = c \cdot A^T$
- $(A^T)^T = A$
- $(A \cdot B)^T = B^T \cdot A^T$

Beispiel: $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$

Beispiele: Matrizenmultiplikation

Zusammenhang zwischen Skalarprodukt von Vektoren und Matrizenmultiplikation:

$$\left\langle \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\rangle = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}^T \cdot \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = 2$$

Multiplikation einer Matrix und eines Vektors

$$\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 9 \\ 12 \end{pmatrix}$$

Matrizen als lineare Abbildungen

Seien V und W Vektorräume. Eine Abbildung $f : V \to W$ heißt linear oder ein Homomorphismus, wenn folgendes gilt:

1. $f(x + y) = f(x) + f(y)$ und
2. $f(\lambda x) = \lambda f(x)$.

Bemerkungen:

- Jede Matrix $A \in K^{m \times n}$ definiert eine lineare Abbildung $A : K^n \to K^m$ mit $x \mapsto Ax$.
- Die Matrizenmultiplikation entspricht der Komposition der linearen Abbildungen: Sei $A \in K^{m \times n}$ und $B \in K^{r \times m}$, dann ist $B \circ A : K^n \to K^r$ mit $B \circ A(x) = (B \cdot A)x$.

Beispiel:

$$\left(\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \cdot \begin{pmatrix} 1 & 4 & 1 & 0 \\ 2 & 5 & 0 & 1 \\ 3 & 6 & 2 & 1 \end{pmatrix}\right) \begin{pmatrix} 1 \\ 2 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 14 & 32 & 7 & 5 \\ 32 & 77 & 16 & 11 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 97 \\ 229 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \circ \begin{pmatrix} 1 & 4 & 1 & 0 \\ 2 & 5 & 0 & 1 \\ 3 & 6 & 2 & 1 \end{pmatrix} \left(\begin{pmatrix} 1 \\ 2 \\ 2 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \left(\begin{pmatrix} 11 \\ 13 \\ 20 \end{pmatrix}\right) = \begin{pmatrix} 97 \\ 229 \end{pmatrix}$$

Matrizen zur Beschreibung linearer Gleichungssysteme

Ein Gleichungssystem mit n Variablen kann man auch als Matrixproblem beschreiben:
Beispiel: Das Gleichungssystems

$$\begin{array}{rcrcrcr} -x & + & 2y & + & z & = & -2 \\ 3x & - & 8y & - & 2z & = & 4 \\ x & + & & & 4z & = & -2 \end{array}$$

lässt sich beschreiben als

$$\begin{pmatrix} -1 & 2 & 1 \\ 3 & -8 & -2 \\ 1 & 0 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -2 \\ 4 \\ -2 \end{pmatrix}$$

Zur Lösung solcher Systeme kann z.B. des Gaußsche Eliminationsverfahren eingesetzt werden.
(Lösung für das Beispiel $x = 2, y = 0.5, z = -1$).

Distributionelle Semantik

Die folgenden Folien entstammen dem Foliensatz “Distributional Semantics” von Marco Baroni und Gemma Boleda zu dem Kurs “CS 388: Natural Language Processing”.

`https://www.cs.utexas.edu/~mooney/cs388/slides/dist-sem-intro-NLP-class-UT.pdf`

The distributional hypothesis

- The meaning of a word is the set of contexts in which it occurs in texts
- *Important aspects of* the meaning of a word *are a function of* (*can be approximated by*) the set of contexts in which it occurs in texts

The distributional hypothesis in real life

McDonald & Ramscar 2001

He filled the wampimuk, passed it around and we all drunk some

We found a little, hairy wampimuk sleeping behind the tree

Collecting context counts for target word dog

The dog barked in the park.
The owner of the dog put him
on the leash since he barked.

bark	++
park	+
owner	+
leash	+

Collecting context counts for target word dog

The dog barked in the park.
The owner of the dog put him
on the leash since he barked.

bark	++
park	+
owner	+
leash	+

Collecting context counts for target word dog

The dog barked in the park.
The owner of the dog put him
on the leash since he barked.

bark	++
park	+
owner	+
leash	+

Collecting context counts for target word dog

The dog barked in the park.
The owner of the dog put him
on the leash since he barked.

bark	++
park	+
owner	+
leash	+

Collecting context counts for target word dog

The dog barked in the park.
The owner of the dog put him
on the leash since he barked.

bark	++
park	+
owner	+
leash	+

Collecting context counts for target word dog

The dog barked in the park.
The owner of the dog put him
on the leash since he barked.

bark	++
park	+
owner	+
leash	+

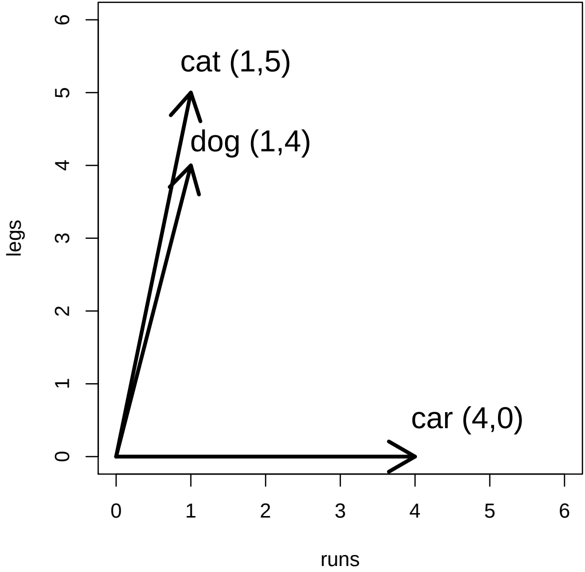
The co-occurrence matrix

	leash	walk	run	owner	pet	bark
dog	3	5	2	5	3	2
cat	0	3	3	2	3	0
lion	0	3	2	0	1	0
light	0	0	0	0	0	0
bark	1	0	0	2	1	0
car	0	0	1	3	0	0

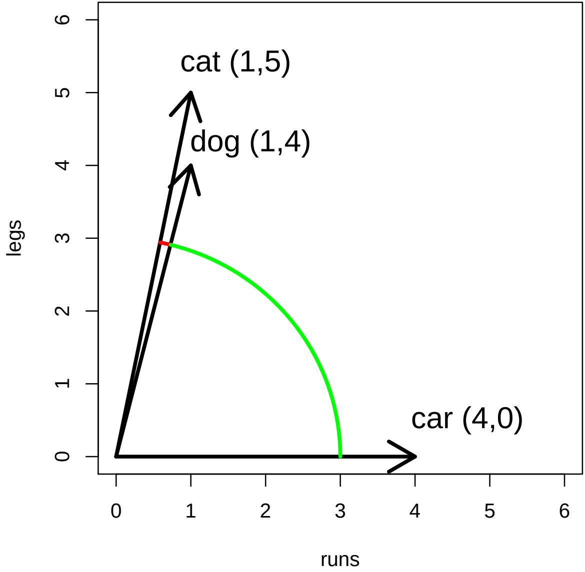
Contexts as vectors

	runs	legs
dog	1	4
cat	1	5
car	4	0

Semantic space



Semantic similarity as angle between vectors



Nearest neighbour examples

BNC, 2-content-word-window context

rhino	fall	rock
woodpecker	rise	lava
rhinoceros	increase	sand
swan	fluctuation	boulder
whale	drop	ice
ivory	decrease	jazz
plover	reduction	slab
elephant	logarithm	cliff
bear	decline	pop
satin	cut	basalt
sweatshirt	hike	crevice

Nearest neighbour examples

BNC, 2-content-word-window context

green	good	sing
blue	bad	dance
yellow	excellent	whistle
brown	superb	mime
bright	poor	shout
emerald	improved	sound
grey	perfect	listen
speckled	clever	recite
greenish	terrific	play
purple	lucky	hear
gleaming	smashing	hiss