Matrizen

Katja Markert

Institut für Computerlinguistik
Uni Heidelberg
markert@cl.uni-heidelberg.de

May 21, 2019

Bisher und heute

1. Bisher: Sparse Embeddings
2. Probleme: Zuviel Rauschen, versteckte Ähnlichkeiten
3. Eine Möglichkeit zum Erreichen von dichten Vektoren ist mittels Matrizenzerlegungen (Singular Value Decomposition)
4. Jetzt: Hintergrund zu den mathematischen Begriffen und Methoden, die für Matrizenzerlegung notwendig ist
5. Insbesondere: Matrizen

Übersicht

Übersicht

Matrixdefinition (reelle Matrizen)

Eine reelle $m \times n$ **Matrix** A ist ein rechteckiges Schema aus m Zeilen und n Spalten mit reellen Zahlen als **Einträgen/Elementen**.

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

mit $a_{ij} \in \mathbb{R}$ *fuer* $i \in \{1, \dots m\}, j \in \{1, \dots, n\}$

Wir nennen die i-te Zeile auch den i-ten Zeilenvektor:

$$A_{(i)} = (a_{i1}, \dots a_{in})$$

Ebenso die j-te Spalte den j-ten Spaltenvektor:

$$A^{(j)} = (a_{1j}, \dots a_{mj})$$

Man schreibt auch $A = (a_{ij})$

:

Begriffe: Zeilenindex, Spaltenindex, Hauptdiagonale

Einige Sonderfälle

- Jeder n-dimensionale Vektor kann als $n \times 1$ bzw $1 \times n$ Matrix aufgefasst werden.
- Die **Nullmatrix** hat nur Nullen als Einträge.
- Eine Matrix heisst **quadratisch**, wenn $m = n$ (gleiche Zeilen und Spaltenanzahl)

Diagonalmatrizen, Einheitsmatrix

- Eine $n \times n$ Matrix $D = (d_{ij})$ heisst **Diagonalmatrix**, wenn gilt:

$$\forall i, j \in \{1, \ldots, n\}, i \neq j \implies d_{i,j} = 0$$

$$\begin{pmatrix} d_{11} & 0 & \cdots & \cdots & 0 \\ 0 & d_{22} & \vdots & 0 & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ \vdots & \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & d_{nn} \end{pmatrix}$$

- Die $n \times n$ **Einheitsmatrix** ist die Diagonalmatrix $E = (e_{ij})$ mit $e_{ii} = 1 \forall i \in \{1, \ldots, n\}$

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \vdots & 0 & \vdots \\ \vdots & \vdots & \ddots & 0 & \vdots \\ \vdots & \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}$$

Zwei kurze Fragen

- Ist die Nullmatrix nach Definition eine Diagonalmatrix?
- Wie kann man die Definition einer Diagonalmatrix auf nicht-quadratische Matrizen ausdehnen?

Matrizen als Vektorraum

Gegeben seien zwei reelle $m \times n$ Matrizen A und B. Dann definiert man die Addition der beiden Matrizen komponentenweise mit:

$$(A+B)_{ij} := a_{ij} + b_{ij}$$

sowie die Skalarmultiplikation ebenfalls komponentenweise für jedes $\alpha \in \mathbb{R}$

$$(\alpha \cdot A)_{ij} := \alpha \cdot a_{ij}$$

Kurzübung

- Zeigen Sie: Gegeben zwei natürliche Zahlen m und n. Dann ist die Menge der reellen $m \times n$-Matrizen

$$\mathbb{R}^{m \times n} := \{(a_{ij}) | a_{ij} \in \mathbb{R}\}$$

  ein reeller Vektorraum (**Matrizenraum**) unter der komponentenweise Addition und komponentenweiser Skalarmultiplikation.
  - Warum gelten Kommutativ, Assoziativ und Distributivgesetze?
  - Was ist das neutrale Element (bzgl Matrixaddition)?
  - Was ist das additiv inverse Element zu einer gegebenen Matrix?
- Was könnte eine Basis des Matrizenraums sein?
- Es gilt: Die Menge der Diagonalmatrizen ist ein Unterraum des Matrizenraums $\mathbb{R}^{n \times n}$.

Normen im Matrizenraum: Frobeniusnorm

Auf diesem Vektorraum können wir nun auch Normen definieren. Zum Beispiel das “Äquivalent” zur euklidischen Norm, die sogenannte **Frobeniusnorm**:

Frobeniusnorm

Gegeben sei eine Matrix $A = (a_{ij}) \in \mathbb{R}^{m \times n}$. Dann ist die Frobeniusnorm definiert als:

$$\|A\|_F := \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$$

Normen im Matrizenraum: Frobeniusnorm

Auf diesem Vektorraum können wir nun auch Normen definieren. Zum Beispiel das "Äquivalent" zur euklidischen Norm, die sogenannte **Frobeniusnorm**:

Frobeniusnorm

Gegeben sei eine Matrix $A = (a_{ij}) \in \mathbb{R}^{m \times n}$. Dann ist die Frobeniusnorm definiert als:

$$\|A\|_F := \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$$

Übungsüberlegungen für zu Hause

- Warum kann man einfach sehen, dass die Frobeniusnorm eine Norm ist?
- Berechnen Sie die Frobeniusnorm der Matrix

$$\begin{pmatrix} 1 & 2 & 1 \\ -1 & 2 & -3 \\ 0 & 1 & 2 \end{pmatrix}$$

- Wie ist dann die Metrik, die von der Frobeniusnorm induziert wird?

Übersicht

1 Matrizen

2 Matrizen als lineare Abbildung

3 Transposition, Rang, Orthonormalität

Multiplikation Matrix mit Vektor

Wir definieren nun die Multiplikation einer $m \times n$ Matrix $A = (a_{ij})$ mit einem (Spalten)Vektor $\vec{v}$ der Dimension n. Heraus kommt ein Vektor $\vec{v}'$ der Dimension m.

$$A\vec{v} = \begin{pmatrix} a_{11} & \cdots & \cdots & a_{1n} \\ a_{21} & \cdots & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & \cdots & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ \cdots \\ v_n \end{pmatrix} := \begin{pmatrix} a_{11}v_1 + a_{12}v_2 + \cdots + a_{1n}v_n \\ a_{21}v|1 + a_{22}v_2 + \cdots + a_{2n}v_n \\ \cdots \\ a_{m1}v_1 + a_{m2}v_2 + \cdots + a_{mn}v_n \end{pmatrix}$$

bzw

$$A\vec{v} = \begin{pmatrix} \langle A_{(1)}, v \rangle \\ \langle A_{(2)}, v \rangle \\ \cdots \\ \langle A_m), v \rangle \end{pmatrix} =: \vec{v}' \in \mathbb{R}^m$$

Merke als: Reihenvektor mal Spaltenvektor.
Dimensionsübereinstimmungen!

Beispiel

$$\begin{pmatrix} 3 & 0 & 4 \\ 1 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -5 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \cdot 1 + 0 \cdot (-5) + 4 \cdot 4 \\ 1 \cdot 1 + 3 \cdot (-5) + 1 \cdot 4 \end{pmatrix} = \begin{pmatrix} 19 \\ -10 \end{pmatrix}$$

Mit der Multiplikation durch eine $m \times n$-Matrix A kann ein Vektor v aus dem $\mathbb{R}^n$ in einen im Raum $\mathbb{R}^m$ überführt werden mittels einer linearen Kombination aus Reihenvektoren in A und v. Die Matrix definiert eine **lineare Abbildung**.

Bei quadratischen Matrizen bleiben wir im gleichen Raum.

Beispiel:
Was macht die Matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Drehung um 90 Grad

Matrizen als lineare Abbildung

Mit der Multiplikation durch eine $m \times n$-Matrix A kann ein Vektor v aus dem $\mathbb{R}^n$ in einen im Raum $\mathbb{R}^m$ überführt werden mittels einer linearen Kombination aus Reihenvektoren in A und v. Die Matrix definiert eine **lineare Abbildung**.

Bei quadratischen Matrizen bleiben wir im gleichen Raum.

Beispiel:
Was macht die Matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Drehung um 90 Grad

Idee: Bei manchen Vektoren wird durch manche quadratische (!) Matrizen die Richtung nicht geändert, sie werden "schlimmstenfalls" skaliert

Eigenvektor

Gegeben sei eine quadratische Matrix $A \in \mathbb{R}^{n \times n}$. Dann heisst ein Spaltenvektor $\vec{v} \in \mathbb{R}^n$ mit $\vec{v} \neq \vec{0}$ ein **(rechter) Eigenvektor** von A, wenn es ein $\lambda \in \mathbb{R}$ gibt mit

$$A \cdot \vec{v} = \lambda \cdot \vec{v}$$

λ heisst dann ein **Eigenwert**.

Beispiel zur Eigenwertbestimmung

Gegeben sei die Matrix A

$$\begin{pmatrix} 6 & -1 \\ 4 & 0 \end{pmatrix}$$

Für einen Eigenvektor $\begin{pmatrix} x \\ y \end{pmatrix}$ muss gelten:

$$\begin{pmatrix} 6 & -2 \\ 4 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix}$$

für ein $\lambda \in \mathbb{R}$.
Also
(i) $6x - 2y = \lambda x$ sowie (ii) $4x = \lambda y$.

Beispiel continued

(i) $6x - 2y = \lambda x$ sowie (ii) $4x = \lambda y$.

Aus (ii) in (i) folgt damit $\frac{6}{4}\lambda y - 2y = \frac{\lambda^2 y}{4}$ und damit $6\lambda - 8 = \lambda^2$ bzw $\lambda^2 - 6\lambda + 8 = 0$.

Nullstellenbestimmung ergibt, es gibt zwei Eigenwerte 2 und 4.

Eigenvektoren zu $\lambda = 2$:
$4x = 2y$ sowie $6x - 2y = 2x \rightarrow$ alle Vektoren mit $2x = y$

Eigenvektoren zu $\lambda = 4$:
$4x = 4y$ sowie $6x - 2y = 4x \rightarrow$ alle Vektoren mit $x = y$

Beispiel continued

(i) $6x - 2y = \lambda x$ sowie (ii) $4x = \lambda y$.

Aus (ii) in (i) folgt damit $\frac{6}{4}\lambda y - 2y = \frac{\lambda^2 y}{4}$ und damit $6\lambda - 8 = \lambda^2$ bzw $\lambda^2 - 6\lambda + 8 = 0$.

Nullstellenbestimmung ergibt, es gibt zwei Eigenwerte 2 und 4.

Eigenvektoren zu $\lambda = 2$:
$4x = 2y$ sowie $6x - 2y = 2x \rightarrow$ alle Vektoren mit $2x = y$

Eigenvektoren zu $\lambda = 4$:
$4x = 4y$ sowie $6x - 2y = 4x \rightarrow$ alle Vektoren mit $x = y$

Beispiel continued

(i) $6x - 2y = \lambda x$ sowie (ii) $4x = \lambda y$.

Aus (ii) in (i) folgt damit $\frac{6}{4}\lambda y - 2y = \frac{\lambda^2 y}{4}$ und damit $6\lambda - 8 = \lambda^2$ bzw $\lambda^2 - 6\lambda + 8 = 0$.

Nullstellenbestimmung ergibt, es gibt zwei Eigenwerte 2 und 4.

Eigenvektoren zu $\lambda = 2$:
$4x = 2y$ sowie $6x - 2y = 2x \rightarrow$ alle Vektoren mit $2x = y$

Eigenvektoren zu $\lambda = 4$:
$4x = 4y$ sowie $6x - 2y = 4x \rightarrow$ alle Vektoren mit $x = y$

Beispiel continued

(i) $6x - 2y = \lambda x$ sowie (ii) $4x = \lambda y$.

Aus (ii) in (i) folgt damit $\frac{6}{4}\lambda y - 2y = \frac{\lambda^2 y}{4}$ und damit $6\lambda - 8 = \lambda^2$ bzw $\lambda^2 - 6\lambda + 8 = 0$.

Nullstellenbestimmung ergibt, es gibt zwei Eigenwerte 2 und 4.

Eigenvektoren zu $\lambda = 2$:
$4x = 2y$ sowie $6x - 2y = 2x \rightarrow$ alle Vektoren mit $2x = y$

Eigenvektoren zu $\lambda = 4$:
$4x = 4y$ sowie $6x - 2y = 4x \rightarrow$ alle Vektoren mit $x = y$

Fragen und Beispiele

- Was sind Eigenwerte und Eigenvektoren zur Nullmatrix?
- Was sind Eigenwerte und Eigenvektoren der Einheitsmatrix?
- Was sind die Eigenvektoren einer beliebigen Diagonalmatrix?
- Was sind die Eigenwerte und Eigenvektoren unser Drehungsmatrix?

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Eigenschaften

- Eine Matrix muss keine Eigenwerte haben (siehe auch Nullstellenbestimmung von Polynomen)
- Eine $n \times n$ Matrix kann höchstens n Eigenwerte haben.
- Eine Matrix kann eine beliebige Anzahl von Eigenvektoren haben.
- Sind $\lambda_1, \lambda_2, \ldots, \lambda_r$ verschiedene Eigenwerte der $n \times n$ Matrix A mit zugehorigen Eigenvektoren $x_1, x_2, \ldots, x_r$, so sind die Vektoren $x_1, x_2, \ldots, x_r$ linear unabhangig
- Verfahren zur Eigenwertbestimmung: Gausssche Elimination, Determinanten ...

Die **Matrizenmultiplikation** ist eine binäre Verknüpfung zweier Matrizen

$$\cdot : \mathbb{R}^{l\times m} \times R^{m\times n} \to R^{l\times n}, \quad (A, B) \mapsto C = A \cdot B$$

die zwei Matrizen $A = (a_{ij})$ und $B = (b_{jk})$ eine weitere Matrix $C = (c_{ik})$ zuordnet. Hierbei ist definiert:

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

i-te Zeile mal *j*-te Spalte!

Beispiel

$$\begin{pmatrix} 3 & 0 & 4 \\ 1 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & -4 \\ -5 & 2 & 3 & -1 \\ 4 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 19 & 3 & 0 & -8 \\ -10 & 7 & 9 & -6 \end{pmatrix}$$

Bemerkungen

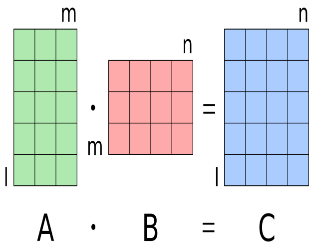

Bild von Quartl - Eigenes Werk, CC BY-SA 3.0,

https://commons.wikimedia.org/w/index.php?curid=27634996

- Intuition: Wir führen zwei Abbildungen hintereinander aus! Beispiel: Drehung plus Spiegelung = Drehspiegelung
- Besteht die erste Matrix aus nur einer Zeile und die zweite nur aus einer Spalte, so haben wir im Endeffekt das Skalarprodukt zweier Vektoren
- Matrix mal Vektor ist ein Spezialfall der Matrixmultiplikation

Matrixmultiplikation: Eigenschaften

- Nicht kommutativ: Im Allgemeinen $A \cdot B \neq B \cdot A$
- Aber assoziativ: $(A \cdot B) \cdot C = A \cdot (B \cdot C)$
- Die Multiplikation der quadratischen Matrix A mit der Einheitsmatrix ergibt wieder A
- Diagonalmatrix mal Diagonalmatrix ergibt wieder Diagonalmatrix

Zwischenfazit

- Matrizenraum: Vektorraum mit komponentenweiser Addition und komponentenweiser Skalarmultiplikation
- Frobeniusnorm
- Eine $m \times n$ Matrix stellt eine lineare Abbildung von $\mathbb{R}^n \to \mathbb{R}^m$ dar.
- Matrix-Vektor-Multiplikation führt diese Abbildung aus
- Eigenvektoren werden von einer quadratischen Abbildung nur skaliert
- Matrizenmultiplikation ist die Verknüpfung zweier Abbildungen

Übersicht

1. Matrizen
2. Matrizen als lineare Abbildung
3. Transposition, Rang, Orthonormalität

Rang einer $m \times n$ Matrix

Die Anzahl der linear unabhängigen Spaltenvektoren einer Matrix A nennt man den **Spaltenrang** von A.

Die Anzahl der linear unabhängigen Zeilenvektoren einer Matrix A nennt man den **Zeilenrang** von A

Die Matrix

$$A = \begin{pmatrix} 1 & 1 & 0 & 2 \\ -1 & -1 & 0 & -2 \end{pmatrix}$$

hat Spaltenrang 1. Sie hat auch Zeilenrang 1.

Es gilt immer (ohne Beweis): Zeilenrang = Spaltenrang = **Rang** und damit, $Rang(A) \leq min\{m, n\}$

Rang einer $n \times n$ Matrix

Eine quadratische Matrix $n \times n$ A hat genau dann den vollen Rang n, wenn keine ihrer Eigenwerte Null ist (da dann alle Vektoren linear unabhängig!)

Folgt direkt aus der Definition von linearer Unabhängigkeit!

Transposition

Die **Transponierte** einer $m \times n$ Matrix A ist die $n \times m$ Matrix A^T, bei der die Zeilen und Spalten von A vertauscht wurden.

$$\begin{pmatrix} 1 & 4 \\ 8 & -2 \\ -3 & 5 \end{pmatrix}^T = \begin{pmatrix} 1 & 8 & -3 \\ 4 & -2 & 5 \end{pmatrix}$$

es gilt $(A^T)^T = A$

Eigenschaften

Gegeben sei die $m \times n$ Matrix A sowie B eine $n \times l$ Matrix.

- Es gilt $(A \cdot B)^T = B^T \cdot A^T$
- Der Rang von A^T ist der Rang von A
- Die Frobeniusnorm von A^T ist die Frobeniusnorm von A.
- Ohne Beweis: Ist A quadratisch, dann hat A^T die gleichen Eigenwerte wie A! Die Eigenvektoren müssen aber nicht gleich sein.
- Ohne Beweis: Die Matrix $A \cdot A^T$ ist eine quadratische $m \times m$ Matrix mit nicht-negativen Eigenwerten. Die Matrix $A^T \cdot A$ eine quadratische $n \times n$ Matrix mit nicht-negativen Eigenwerten. Die positiven (nicht-Null) Eigenwerte der beiden Matrizen sind gleich.

Orthonormalmatrizen

Eine **Orthonormalmatrix** ist eine quadratische Matrix, deren Zeilen (Spalten) paarweise orthogonale Einheitsvektoren sind.

Beispiel:

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

oder

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Orthonormalmatrizen: Alternativdefinition

Gegeben quadratische Orthonormalmatrix A.

1. Da damit zwei Spaltenvektoren aufeinander senkrecht stehen, gilt dass das Skalarpodukt zweier Spaltenvektoren $\langle a^{(i)}, a^{(j)} \rangle = 0$ wenn $i \neq j$ gleich Null ist. Das Gleiche für Zeilenvektoren.
2. Da die Spaltenvektoren Einheitsvektoren sind, gilt $\langle a^{(i)}, a^{(i)} \rangle = 1$ für alle i

Damit gilt

$$A^T \cdot A = A^T \cdot A = E$$

Eigenschaften

- Orthonormale Matrizen bewahren das Skalarpodukt

$$(Ax)^T(Ay) = x^T A^T Ay = x^T y$$

- Bewahren damit auch die Normen und die Winkel der Ausgangsvektoren

$$cos(Ax, Ay) = \frac{(Ax)^T(Ay)}{\|Ax\|\|Ay\|} = \frac{x^T y}{\|x\|\|y\|} = cos(x, y)$$

- Im zwei und dreidimensionalem Raum sind die Orthogonalmatrizen imemr Spiegelungen, Rotationen oder Drehspiegelungen.

Fazit

1. Wenn wir eine nicht-quadratische Matrix A haben, können wir durch Übergang zu $A^T \cdot A$ doch wieder Eigenwerte und Eigenektoren berechnen.
2. Orthonormale Matrizen sind besonders schöne Matrizen, die Skalarprodukte, Normen und Winkel zwischen zwei Ausgangsvektoren nach der Abbildung bewahren

Literatur

- Gerd Fischer: Lineare Algebra. Eine Einführung für Studienanfänger
- Zu Eigenwerten ist auch gut: `https://www.empiwifo.uni-freiburg.de/lehre-teaching-1/winter-term-11-12/materialien-wirtschaftsmathematik/eigenwert`
- SVD Tutorial (ohne vollständigen Hintergrund) : Kirk Baker (2005): Singular Value Decomposition Tutorial `https://datajobs.com/data-science-repo/SVD-Tutorial-[Kirk-Baker].pdf`
- Am Donnerstag: SVD praktisch im Pool