Mathematische Grundlagen der Computerlinguistik

Matrizen/Maschinelles Lernen

Michael Staniek

ICL, Universität Heidelberg, SoSe 2019

26.06.2019

Inhalte der heutigen Vorlesung

- Matrizen
- Maschinelles Lernen

Matrixgrößen

Matrix Shapes

The size of a matrix is the number of rows by the number of columns.

$$\begin{bmatrix} -1 & 5 & 0 & 1 \\ 2 & -3 & 8 & -7 \end{bmatrix} \rightarrow 2 \times 4$$

"2 by 4"

$$\begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & 2 \\ 2 & -1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \rightarrow 4 \times 3$$

"4 by 3"

$$\begin{bmatrix} 3 & 3 & 1 & 5 & 1 \end{bmatrix} \rightarrow 1 \times 5$$

"1 by 5"

$$\begin{bmatrix} 4 & 1 \\ 0 & 9 \end{bmatrix} \rightarrow 2 \times 2$$

"2 by 2"

Nullmatrix

Zero Matrix

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

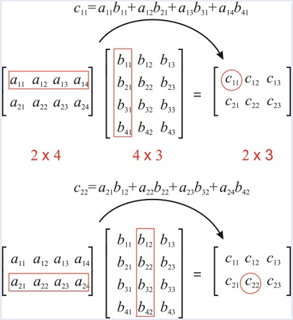
Einheitsmatrix

Identity Matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrixmultiplikation

Matrixmultiplication



Transponierte Matrix

Matrixtranspose

$$A \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \qquad A^T \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$

$$A \begin{bmatrix} 1 & 4 & 3 \\ 8 & 2 & 6 \\ 7 & 8 & 3 \\ 4 & 9 & 6 \\ 7 & 8 & 1 \end{bmatrix} \qquad A^T \begin{bmatrix} 1 & 8 & 7 & 4 & 7 \\ 4 & 2 & 8 & 9 & 8 \\ 3 & 6 & 3 & 6 & 1 \end{bmatrix}$$

Determinante

Determinante

Hauptdiagonale Nebendiagonale

$$D = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11} \cdot a_{22} - a_{12} \cdot a_{21}$$

Zahlenbeispiel

$$\det A = \begin{vmatrix} 2 & -3 \\ 4 & 5 \end{vmatrix} = 2 \cdot 5 - (-3 \cdot 4) = 10 + 12 = 22$$

Eigenwerte und Eigenvektoren

Eigenwert

- A eine Matrix
- x ein Vektor (Eigenvektor)
- λ ein Skalar (Eigenwert)
- $Ax = \lambda x$

$$Ax = \lambda x$$
$$Ax - \lambda x = 0$$
$$(A - \lambda E_n)x = 0$$
$$det(A - \lambda E_n) = 0$$

Beispiel

Eigenwert und Eigenvektor

$$\begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}$$

$$\begin{vmatrix} 5-\lambda & -2 \\ -2 & 2-\lambda \end{vmatrix} = 0 \rightarrow (5-\lambda)(2-\lambda)-4$$

$$10-7\lambda+\lambda^2-4 \rightarrow \lambda^2-7\lambda+6=(\lambda-6)(\lambda-1)$$

$$\lambda=6, \lambda=1$$

Klassifikation

- Wie sagen wir Kategorien Voraus?

x	y
(0.9, 0.8)	1
(0.9, -0.8)	-1

- Loss: max(0, -y*w*x)
- Intuition: Wenn w*x einen Wert über 0 und damit das Vorzeichen + hat, y aber den Wert -1 hat, kriegt man wenn man sie miteinander multipliziert eine negative Zahl.
- Von dieser nimmt man die Negation, da der Fehler ein Positiver Wert sein muss.
- Kein Fehler wenn Vorzeichen von (w*x) mit y übereinstimmt!
- Wenn Fehler größer als 0, Update. Ableitung?

Lossfunktion fürs Klassifizieren

max(0, -y*w*x)

Lossfunktion fürs Klassifizieren

max(0, -y*w*x)

$$\begin{cases} -y * w * x & \text{if } -y * w * x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Lossfunktion fürs Klassifizieren

max(0, -y*w*x)

$$\begin{cases} -y * w * x & \text{if } -y * w * x > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{cases} -y * w * x & \text{if } y * w * x \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

Lossfunktion fürs Klassifizieren

$$\begin{cases} -y * w * x & \text{if } y * w * x \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

Lossfunktion fürs Klassifizieren

$$\begin{cases} -y * w * x & \text{if } y * w * x \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{cases} -y * x & \text{if } y * w * x \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

Gradient Descent

Learningrate 0.1

x	y	w	$\max(0, -y * w * x)$	-y*x if y*w*x <= 0
(0.9,0.8)	1	(1,0.1)	0	(0, 0)
(0.9,-0.8)	-1	(1,0.1)	0.82	(0.9, -0.8)

Gradient Descent

Learningrate 0.1

x	y	w	$\max(0, -y * w * x)$	-y*x if y*w*x <= 0
(0.9,0.8)	1	(1,0.1)	0	(0, 0)
(0.9,-0.8)	-1	(1,0.1)	0.82	(0.9, -0.8)
(0.9,0.8)	1	(0.91,0.18)	0	(0, 0)
(0.9,-0.8)	-1	(0.91,0.18)	0.675	(0.9, -0.8)

Gradient Descent

Learningrate 0.1

x	y	w	$\max(0, -y * w * x)$	-y*x if y*w*x <= 0
(0.9,0.8)	1	(1,0.1)	0	(0, 0)
(0.9,-0.8)	-1	(1,0.1)	0.82	(0.9, -0.8)
(0.9,0.8)	1	(0.91,0.18)	0	(0, 0)
(0.9,-0.8)	-1	(0.91,0.18)	0.675	(0.9, -0.8)
(0.9,0.8)	1	(0.82,0.26)	0	(0, 0)
(0.9,-0.8)	-1	(0.82,0.26)	0.52	(0.9, -0.8)

Gradient Descent

Learningrate 0.1

x	y	w	$\max(0, -y * w * x)$	-y*x if y*w*x <= 0
(0.9,0.8)	1	(1,0.1)	0	(0, 0)
(0.9,-0.8)	-1	(1,0.1)	0.82	(0.9, -0.8)
(0.9,0.8)	1	(0.91,0.18)	0	(0, 0)
(0.9,-0.8)	-1	(0.91,0.18)	0.675	(0.9, -0.8)
(0.9,0.8)	1	(0.82,0.26)	0	(0, 0)
(0.9,-0.8)	-1	(0.82,0.26)	0.52	(0.9, -0.8)
(0.9,0.8)	1	(0.73,0.34)	0	(0, 0)
(0.9,-0.8)	-1	(0.73,0.34)	0.385	(0.9, -0.8)

Vielen Dank für die Aufmerksamkeit!