

Vektoren 6

- Wir betrachten im folgenden einige Anwendungsbeispiele zur Vektorrechnung, die aus verschiedenen Gebieten der Informatik stammen, speziell aus der Computergraphik und dem Information Retrieval.
- Beispiel: Spiegelungen und Brechungen von Lichtstrahlen an den Oberflächen von virtuellen dreidimensionalen Objekten in der Computergraphik.
- Beispiel: Berechnung der farblichen Darstellung von Oberflächen in der dreidimensionalen Computergraphik mit Hilfe von Normalenvektoren. Grundideen beim Flat-Shading, Gouraud-Shading und Phong-Shading.
- Beispiel: Beschreibung von Objekten (Textdateien, Audiodateien, digitale Bilder u.s.w.) mit „**Feature-Vektoren**“.
- Beispiel: Einsatz von Hyperebenen in höherdimensionalen Vektorräumen bei der Klassifizierung von Daten durch eine „Support Vector Machine“.
- Problemstellung: Wie groß ist die „Entfernung“ zwischen zwei Textdateien? Wie kann ein Computerprogramm erkennen, daß ein Zeitungsartikel über Fußball einem Bericht über Tennis ähnlicher ist, als einem Essay über Liebeslyrik?

- **Ähnlichkeitssuche**

Die letzte Frage führt uns zu den Methode der Ähnlichkeitssuche. Dabei sollen Objekte, beispielsweise Text-Dokumente oder Bilder, die einer Anfrage am ehesten entsprechen, gefunden werden. Die Objekte werden durch Vektoren beschrieben, die Anfrage ebenfalls.

Die Grundidee der Ähnlichkeitssuche besteht nun darin, als Antwort auf die Anfrage diejenigen Objekte (Text-Dokumente . . .) zu liefern, deren Vektoren „die größte Ähnlichkeit“ mit dem Anfragevektor haben.

Im folgenden beschäftigen wir uns damit, die Ähnlichkeit von Vektoren mathematisch zu erfassen. Wir werden sehen, daß es dafür mehrere Möglichkeiten gibt.

- Definition (Abstand von Vektoren)

Es seien $\vec{a}, \vec{b} \in \mathbb{R}^n$ zwei n -dimensionale Vektoren. Die reelle Zahl

$$d(\vec{a}, \vec{b}) = |\vec{a} - \vec{b}|$$

heißt der **Abstand** oder die **Distanz** zwischen \vec{a} und \vec{b} .

- Veranschaulichung, Skizze.

- Satz

Es seien $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^n$ beliebige n -dimensionale Vektoren. Es gilt

1. $d(\vec{a}, \vec{b}) \geq 0$, wobei $d(\vec{a}, \vec{b}) = 0 \Leftrightarrow \vec{a} = \vec{b}$,
2. $d(\vec{a}, \vec{b}) = d(\vec{b}, \vec{a})$ (Symmetrie),
3. $d(\vec{a}, \vec{b}) \leq d(\vec{a}, \vec{c}) + d(\vec{c}, \vec{b})$ (Dreiecksungleichung).

- Anmerkungen zum Beweis.

- Definition (Metrik, metrischer Raum)

Es sei M eine Menge und $d : M \times M \rightarrow \mathbb{R}$ eine Funktion, die für alle $x, y, z \in M$ folgende Eigenschaften hat:

1. $d(x, y) \geq 0$, wobei $d(x, y) = 0 \Leftrightarrow x = y$,
2. $d(x, y) = d(y, x)$ (Symmetrie),
3. $d(x, y) \leq d(x, z) + d(z, y)$ (Dreiecksungleichung).

Dann heißt die Funktion d eine **Metrik** auf der Menge M , und M zusammen mit d wird ein **metrischer Raum** genannt.

- Anmerkung: Durch die Funktion d wird also zwei Elementen x und y der Menge M eine positive reelle Zahl $d(x, y)$ zugeordnet. Diese Zahl heißt der **Abstand** oder die **Distanz** zwischen x und y .

Mit metrischen Räumen kann man den anschaulichen Begriff der Entfernung zwischen zwei Punkten verallgemeinern.

- Beispiele:

- der euklidische Abstand (Luftlinienentfernung);
- die Manhattan-Metrik;
- die Metrik der französischen Eisenbahn.

- Beispiel: Ähnlichkeitssuche mit Hilfe der euklidischen Metrik.

- Anmerkung: Neben dem Abstand können wir auch den Winkel zwischen zwei Vektoren als Maß für die Ähnlichkeit der Vektoren verwenden. Allerdings steht der Winkelbegriff zunächst nur im Anschauungsraum zur Verfügung.

Also versuchen wir, den Begriff des Winkels im n -dimensionalen Vektorraum einzuführen. Das ist wegen der Beziehung

$$\varphi = \arccos \left(\frac{\vec{a} \cdot \vec{b}}{|\vec{a}| \cdot |\vec{b}|} \right)$$

mit Hilfe des Skalarprodukts machbar. Aber wir haben das Skalarprodukt bisher nur im zwei- und dreidimensionalen Raum. Dies liegt eben gerade daran, daß bei der Definition ein Winkel verwendet wurde. Wir drehen uns offenbar im Kreis. Aber es gibt einen Ausweg.

Um das Skalarprodukt im n -dimensionalen Raum zu bekommen, brauchen wir eine Definition, die ohne den Winkelbegriff auskommt. Wir verwenden die Formel, die wir im Zwei- und Dreidimensionalen hergeleitet haben, bei der das Skalarprodukt aus den Koordinaten der Vektoren berechnet wird.

- Definition (Skalarprodukt und Zwischenwinkel im n -dimensionalen Raum)

Es seien $\vec{a}, \vec{b} \in \mathbb{R}^n$ mit $\vec{a} = (a_1, a_2, \dots, a_n)$ und $\vec{b} = (b_1, b_2, \dots, b_n)$.

(a) Die reelle Zahl

$$\vec{a} \cdot \vec{b} = a_1 b_1 + a_2 b_2 + \dots + a_n b_n = \sum_{i=1}^n a_i b_i$$

heißt das **Skalarprodukt** von \vec{a} und \vec{b} .

(b) Wir bezeichnen

$$\arccos \left(\frac{\vec{a} \cdot \vec{b}}{|\vec{a}| \cdot |\vec{b}|} \right)$$

als den **Winkel** zwischen \vec{a} und \vec{b} .

- Definition (Orthogonalität im n -dimensionalen Raum)

Zwei Vektoren aus dem \mathbb{R}^n heißen **orthogonal**, wenn ihr Skalarprodukt gleich Null ist.

- Anmerkung: Damit verallgemeinern wir anschauliche Begriffe des zwei- und dreidimensionalen Raumes auf den anschaulich nicht mehr vorstellbaren n -dimensionalen Raum. Das Beispiel der Ähnlichkeitssuche zeigt, wie nützlich das sein kann. Es gibt noch viele weitere Anwendungen.

- Beispiel: Ähnlichkeitssuche, bei der als Abstandsbegriff der Winkel zwischen zwei Vektoren verwendet wird. (In der Praxis wird auch der Cosinus des Winkels verwendet.)
- Anmerkung: Im n -dimensionalen Raum gelten für das Skalarprodukt dieselben Rechenregeln, die wir bereits vom zwei- und dreidimensionalen Raum kennen.

- Satz (Rechenregeln für das Skalarprodukt)

Es seien $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}$.

- (a) Kommutativität: $\vec{a} \cdot \vec{b} = \vec{b} \cdot \vec{a}$,
- (b) Distributivität: $\vec{a} \cdot (\vec{b} + \vec{c}) = \vec{a} \cdot \vec{b} + \vec{a} \cdot \vec{c}$,
- (c) $\lambda(\vec{a} \cdot \vec{b}) = (\lambda \vec{a}) \cdot \vec{b} = \vec{a} \cdot (\lambda \vec{b})$,
- (d) $\vec{a} \cdot \vec{a} \geq 0$, $\vec{a} \cdot \vec{a} = 0$ genau dann, wenn $\vec{a} = \vec{0}$.

- Beweis: Folgt sofort aus der Definition des Skalarprodukts im n -dimensionalen Raum und den Rechenregeln für die reellen Zahlen.

- Anmerkung: Die Menge der n -dimensionalen Einheitsvektoren bildet zusammen mit der Abstandsfunktion, die den Zwischenwinkel als Abstand zweier Vektoren verwendet, einen metrischen Raum.

- Anmerkung: Als kleine Auswahl („small sample“) an Beispielen „where the problem of finding near neighbors in a large metric space occurs“ nennt Sergey Brin in einer Veröffentlichung die Gebiete Genetik, Spracherkennung, Bilderkennung, Video-Kompression, Data Mining, Information Retrieval und Copy Detection.

(S. Brin: Near Neighbor Search in Large Metric Spaces. Proceedings of the 21st VLDB Conference, 1995.)

- Anmerkung: Es sei noch einmal ausdrücklich betont, daß metrische Räume nicht nur auf Mengen von Vektoren gebildet werden können. Zum Beispiel wird die Menge aller stetigen Funktionen auf dem Intervall $[a, b]$ mit der Metrik

$$d(f, g) = \max_{x \in [a, b]} |f(x) - g(x)|$$

zu einem metrischen Raum.