

Gegeben sind drei Sätze:

1. Der Hund spielt im Garten
2. Die Katze spielt im Haus
3. Der Hund schläft im Haus

Aufgabe 1: Bag-of-Words-Tabelle

Wort	Satz 1	Satz 2	Satz 3
der	1	0	1
die	0	1	0
hund			

Aufgabe 2: Vektoren aufschreiben

- Satz 1: _____
- Satz 2: _____
- Satz 3: _____

Aufgabe 3: Skalarprodukt

- Satz 1 und Satz 2: _____
- Satz 1 und Satz 3: _____

Aufgabe 4: Vektorlängen

$$\|A\| = \sqrt{a_1^2 + a_2^2 + \dots}$$

- $\|\text{Satz 1}\| =$ _____
- $\|\text{Satz 2}\| =$ _____
- $\|\text{Satz 3}\| =$ _____

Aufgabe 5: Cosinus-Ähnlichkeit

Die Cosinus-Ähnlichkeit ist definiert als:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

- $\text{sim}(\text{Satz 1}, \text{Satz 2}) =$ _____
- $\text{sim}(\text{Satz 1}, \text{Satz 3}) =$ _____

Welcher Satz ist ähnlicher zu Satz 1?

Aufgabe 6

Warum ist „Hund“ informativer als „der“?

Aufgabe 7

Was passiert, wenn ein Wort in keinem Satz vorkommt?

Aufgabe 8

Warum sind Bag-of-Words-Vektoren oft nicht ideal für Bedeutung?

Zusatzaufgabe für Zuhause:

Gegeben: Suchanfrage „Hund im Haus“ Welcher Satz ist am ähnlichsten? Begründe mathematisch.
