

# Musterlösung: Einführung in Vektordatenbanken

## Aufgabe 1: Bag-of-Words-Tabelle

Wort	Satz 1	Satz 2	Satz 3
der	1	0	1
die	0	1	0
hund	1	0	1
katze	0	1	0
spielt	1	1	0
schläft	0	0	1
im	1	1	1
garten	1	0	0
haus	0	1	1

## Aufgabe 2: Vektoren

Reihenfolge wie oben:

- Satz 1: (1, 0, 1, 0, 1, 0, 1, 1, 0)
- Satz 2: (0, 1, 0, 1, 1, 0, 1, 0, 1)
- Satz 3: (1, 0, 1, 0, 0, 1, 1, 0, 1)

## Aufgabe 3: Skalarprodukt

- Satz 1 · Satz 2:  
$$= (1 \cdot 0) + (0 \cdot 1) + (1 \cdot 0) + (0 \cdot 1) + (1 \cdot 1) + (0 \cdot 0) + (1 \cdot 1) + (1 \cdot 0) + (0 \cdot 1) = 2$$
- Satz 1 · Satz 3:  
$$= (1 \cdot 1) + (0 \cdot 0) + (1 \cdot 1) + (0 \cdot 0) + (1 \cdot 0) + (0 \cdot 1) + (1 \cdot 1) + (1 \cdot 0) + (0 \cdot 1) = 3$$

## Aufgabe 4: Vektorlängen

- $\|\text{Satz 1}\| = \sqrt{5}$
- $\|\text{Satz 2}\| = \sqrt{5}$
- $\|\text{Satz 3}\| = \sqrt{5}$

## Aufgabe 5: Cosinus-Ähnlichkeit

- $\text{sim}(\text{Satz 1}, \text{Satz 2})$ :

$$\frac{2}{\sqrt{5} \cdot \sqrt{5}} = \frac{2}{5} = 0,4$$

- $\text{sim}(\text{Satz 1}, \text{Satz 3})$ :

$$\frac{3}{5} = 0,6$$

**Antwort:** Satz 3 ist ähnlicher zu Satz 1.

## Aufgabe 6

„Hund“ ist informativer, weil es ein inhaltstragendes Wort ist, während „der“ ein häufiges Funktionswort ist.

## Aufgabe 7

Das Wort erscheint nicht im Vektor (alle Werte 0), es trägt nichts zur Ähnlichkeit bei.

## Aufgabe 8

Bag of Words berücksichtigt keine Wortreihenfolge und keinen Kontext.

## Zusatzaufgabe

Suchanfrage: „Hund im Haus“

Vektor:

$$(0, 0, 1, 0, 0, 0, 1, 0, 1)$$

Ähnlichkeiten:

- mit Satz 1: 2
- mit Satz 2: 2
- mit Satz 3: 3

**Ergebnis:** Satz 3 ist am ähnlichsten.