

Distributionelle Semantik

Schritt 1: Zählen Sie, wie häufig jedes Wort (*Fahrrad*, *Motorrad*, *Auto*) in einem Kasten zusammen mit den Wörtern *Sport* bzw. *Benzin* vorkommt. (Das gemeinsame Vorkommen nennt man auch *Kookkurrenz*.)

	Benzin	Sport
Fahrrad		
Motorrad		
Auto		

Gesunder **Sport**: Nimm dein **Fahrrad!**

Fahrrad zu fahren: sportlich sein
und **Benzin** sparen.

Mein **Motorrad** braucht **Benzin**.

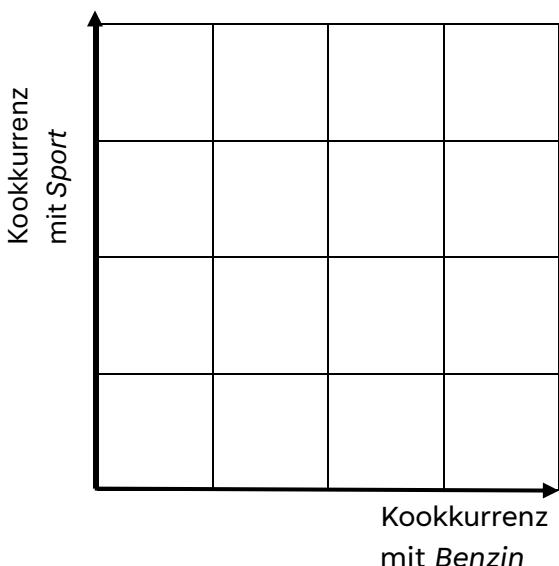
Motorrad fahren ist für mich ein Sport.
Mein **Auto** tankt **Benzin**.
E-Autos brauchen kein **Benzin**.

Schritt 2: Aus den Häufigkeiten können wir Vektoren erstellen.

$$v_{Wort} = \begin{pmatrix} \#Benzin \\ \#Sport \end{pmatrix} \quad v_{Fahrrad} =$$

$$v_{Motorrad} = \quad v_{Auto} =$$

Schritt 3: Wir zeichnen die Vektoren in ein Koordinatensystem ein.



Der **Winkel** θ zwischen zwei Vektoren a und b entspricht der aus den Textdaten berechneten **Ähnlichkeit** der Wörter.

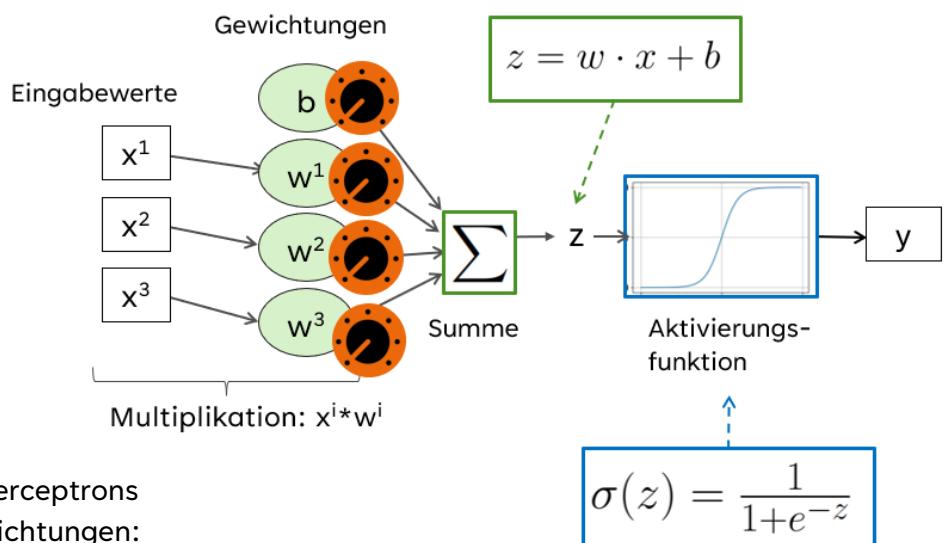
Computerlinguisten sprechen hier auch von der Cosinus-Ähnlichkeit, da:

$$\cos(\theta) = \frac{a \cdot b}{\|a\| \cdot \|b\|}$$

Lesen Sie aus dem Koordinatensystem ab, ob das Wort *Fahrrad* ähnlicher zu *Motorrad* oder zu *Auto* ist (ohne zu rechnen 😊).

Das Perceptron

Das Perceptron ist die kleinste Einheit eines künstlichen neuronalen Netzes.



$$x = \begin{pmatrix} 1 \\ 2 \\ 0.5 \end{pmatrix} \quad w = \begin{pmatrix} 0.25 \\ 0.25 \\ 0.5 \end{pmatrix} \quad b = 0.1$$

Berechnen Sie die Aktivierung des Perceptrons für folgende Eingabewerte und Gewichtungen: