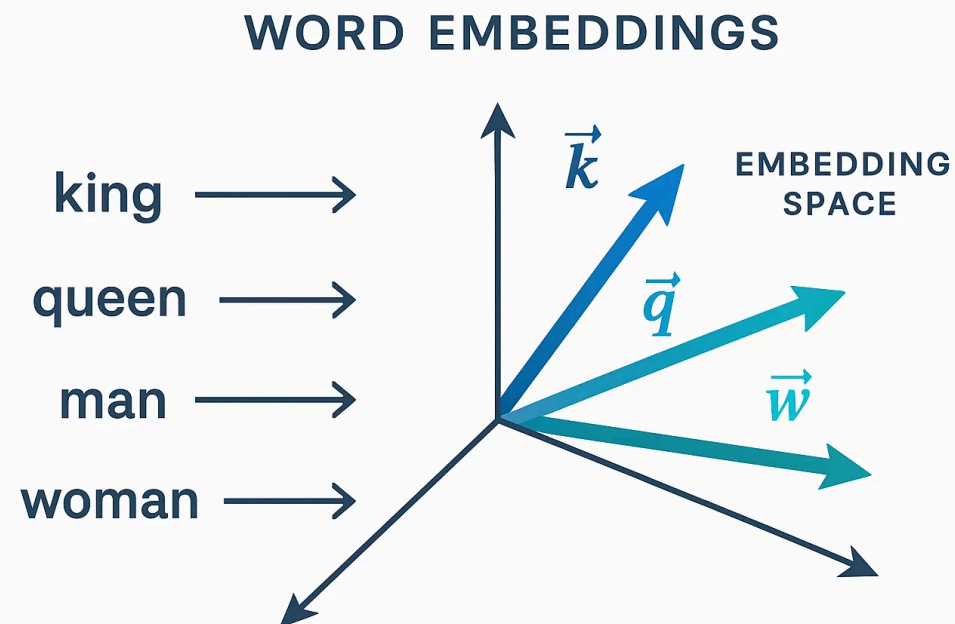


Introdução aos Word Embeddings

Word Embeddings são representações vetoriais densas de palavras que capturam **relações semânticas** em um espaço multidimensional.

- Transformam palavras em vetores numéricos
- Capturam significado e contexto
- Permitem operações matemáticas com palavras
- Base para modelos de NLP modernos



Limitações das Representações One-Hot

⚠ Problemas do One-Hot

Alta dimensionalidade: Tamanho = vocabulário

Esparsidade extrema: Apenas 1 posição ativa

Sem relações semânticas: Todas palavras igualmente distantes

Ineficiência computacional: Muito espaço desperdiçado

💡 Vantagens dos Embeddings Densos

Baixa dimensionalidade: 50-300 dimensões

Representação densa: Todos valores significativos

Relações semânticas: Palavras similares próximas

Eficiência: Menos parâmetros, mais informação

ONE-HOT ENCODING

0	0	0	0
0	1	1	0
0	0	0	1
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	1
0	0	0	0
0	1	0	0
0	0	0	0



DENSE EMBEDDINGS

0,14	-1,32	0,45	-0,67
-0,88	0,25	1,10	0,61
0,79	0,46	-0,33	1,27
-0,27	2,03	-0,12	-1,45

Hipótese Distribucional

*"Palavras que aparecem em **contextos similares** tendem a ter **significados similares**"*

— J.R. Firth (1957)

💡 Princípio Fundamental

Contexto define significado: Palavras vizinhas

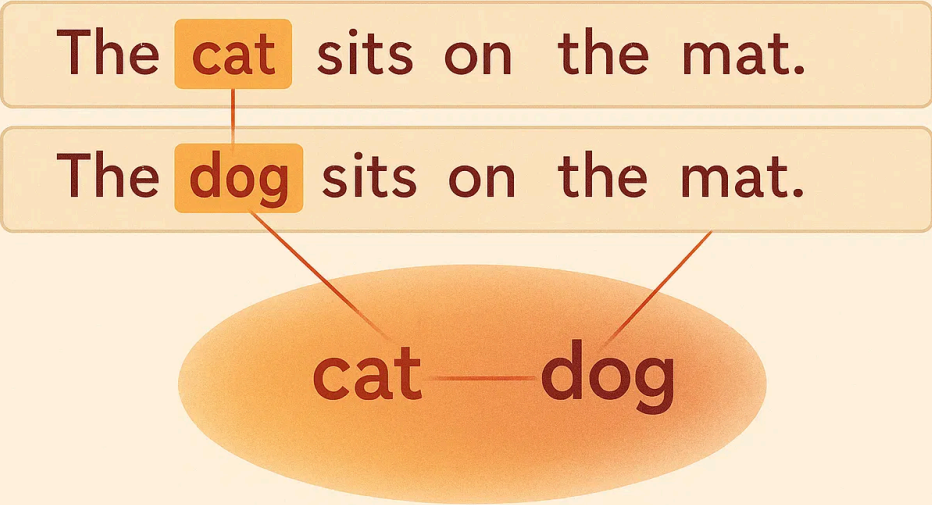
Padrões de co-ocorrência: Frequência conjunta

Base para: Word2Vec, GloVe e FastText

DISTRIBUTIONAL HYPOTHESIS

The **cat** sits on the mat.

The **dog** sits on the mat.



cat — dog

Distributional hypothesis: word appearing in similar contexts tends to cluster in a conceptual space

Operações Vetoriais e Analogias

Word embeddings permitem realizar **operações algébricas** com palavras, capturando relações semânticas complexas.

A famosa analogia:

rei - **homem** + **mulher** \approx **rainha**

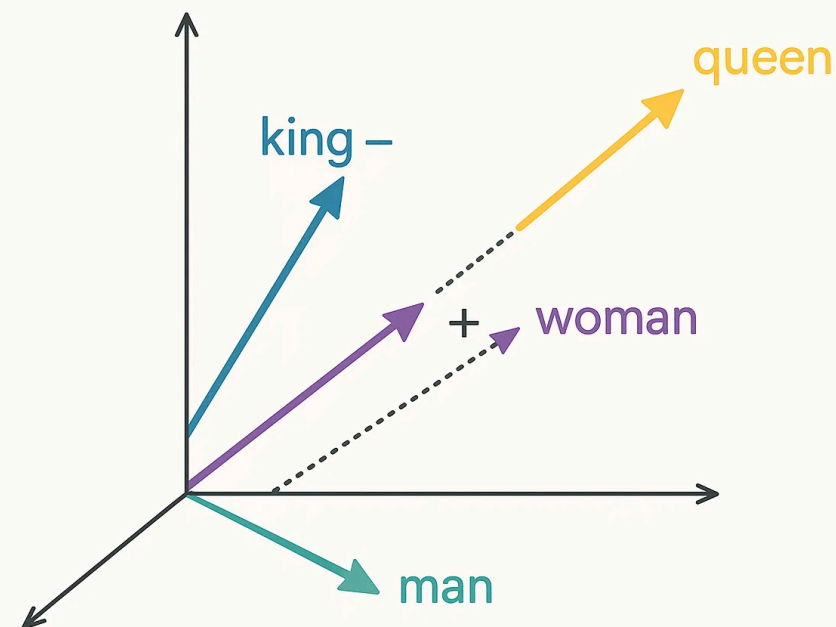
💡 Outras analogias possíveis:

Paris - França + Itália \approx Roma (capitais)

Carro - roda + asa \approx Avião (transportes)

Médico - homem + mulher \approx Médica (profissões)

Estas operações funcionam porque os embeddings **codificam relações** entre conceitos no espaço vetorial, permitindo **raciocínio analógico**.



Visualização de Embeddings

Embeddings existem em espaços de **alta dimensionalidade** (50-300 dimensões), mas podemos visualizá-los em 2D/3D usando técnicas de **redução de dimensionalidade**.

Técnicas de Visualização

PCA: Preserva variância global

t-SNE: Preserva estrutura local

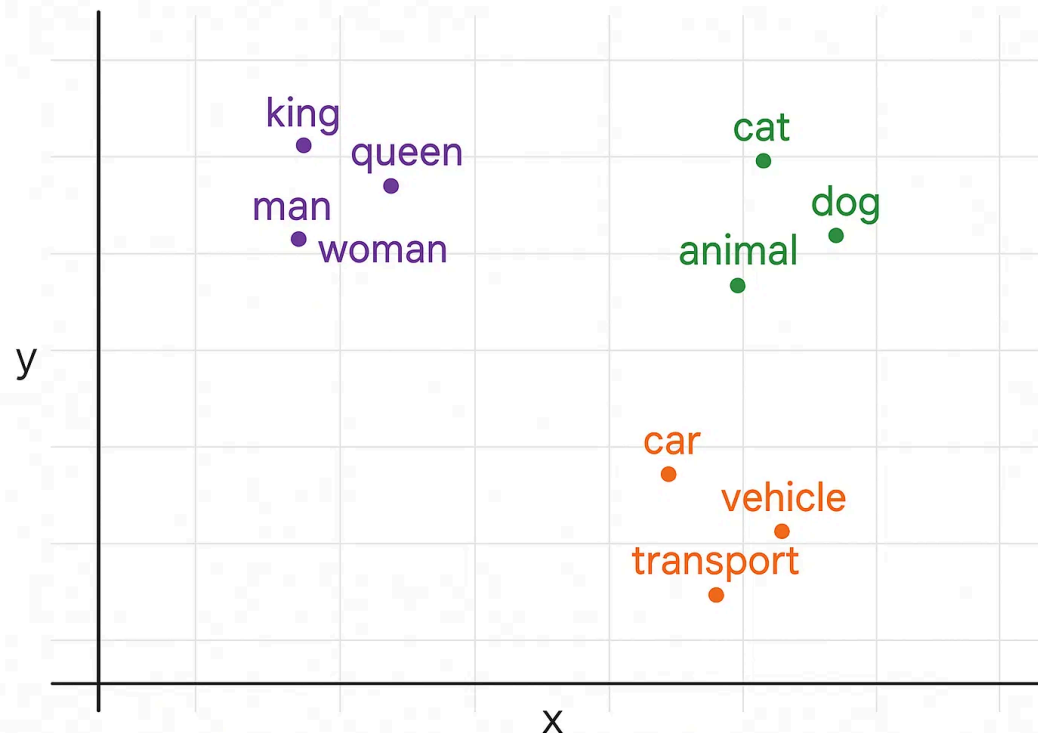
UMAP: Equilíbrio entre estrutura local e global

💡 Insights da Visualização

Palavras semanticamente similares formam **clusters**

Relações análogas aparecem como **vetores paralelos**

Distância entre palavras indica **similaridade semântica**



Aplicações Práticas

Word embeddings são fundamentais para diversas aplicações de **Processamento de Linguagem Natural** e **Inteligência Artificial**.

Tradução Automática

Mapeamento entre espaços vetoriais de diferentes idiomas

Análise de Sentimento

Detecção de emoções e opiniões em textos

Classificação de Documentos

Categorização automática de textos por tópico

Chatbots e Assistentes Virtuais

Compreensão de linguagem natural e geração de respostas

APPLICATIONS OF WORD EMBEDDINGS



MACHINE
TRANSLATION



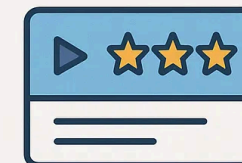
SENTIMENT
ANALYSIS



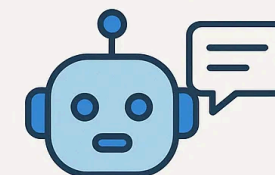
DOCUMENT
CLASSIFICATION



SEARCH
ENGINES



RECOMMENDATION
SYSTEMS



CHATBOTS