



AULA 2 – WORD VECTORS

RENAN AVILA E ALEXANDRE A. CADAVAL

UNIVERSIDADE DE SÃO PAULO

ESCOLA POLITÉCNICA

PROGRAMA DE PÓS-GRADUAÇÃO ENGENHARIA DE COMPUTAÇÃO

24 DE SETEMBRO DE 2025

AGENDA

- Vocabulário
- Representação Palavra
- Significado de Palavras
- Representação de Palavras
- Word Vector Embedding
- Aprendendo Word Vector Embedding
- Word2Vector
- Word2Vector Skip Gram
- Gradient Descend

VOCABULÁRIO

- Conjunto de Palavras/Vocábulos
- Léxico – Todas as palavras de uma língua
 - Inglês: > 1mi ; 170 mil (dicionário Oxford) ; 20 a 35 mil fluente
 - Português: 356 mil (Vocabulário Ortográfico da Lingua Portuguesa)
- Símbolos
 - Sinais de pontuação.
 - EOS – Final da seguencia
 - UNK – Desconhecido
 - Outros...
- Representação
 - $V = [a, ábaco, \dots, glacial, \dots, zebra, \dots, UNK] \rightarrow N$ (número de palavras do dicionário) + 1 (UNK)

REPRESENTAÇÃO DAS PALAVRAS

- Índice no vocabulário
 - A -> 1
 - Ábaco -> 2
 - Zebra -> 123.345
- One-Hot Vector
 - Valor 1 para a posição da palavra no vocabulário e zero nas demais...
 - A = [1,0,0,0,..., 0]
 - Ábaco = [0,1,0,0,0,...,0]
 - Zebra = [0,0,0,0,..., 1, ..., 0]
- Não representa relação entre palavras.
- Porque Matrizes ?

SIGNIFICADO DAS PALAVRAS

- Ideia representada por palavras
- Símbolo <-> Ideia
 - Semântica Denotacional
- WordNET 
 - Dicionário léxico para uso em computação
 - Sem contexto
 - Subjetivo
 - Requer ação humana
 - Dificuldade para calcular similaridade

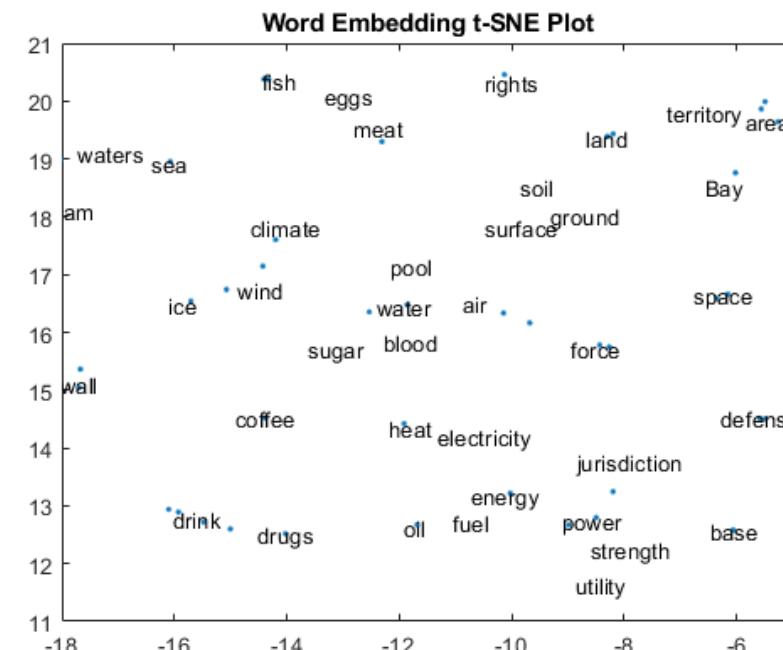
Hamburger

- Hamburger (an inhabitant of Hamburg)
- direct hypernym:
 - German (a person of German nationality)
- sister term
 - German (a person of German nationality)
 - East German (a native/inhabitant of the former GDR)
 - Bavarian (a native/inhabitant of Bavaria)
- derivationally related form
 - Hamburg (a port city in northern Germany on the Elbe River that was founded by Chalemagne in the...)

WORD VECTOR EMBEDDING

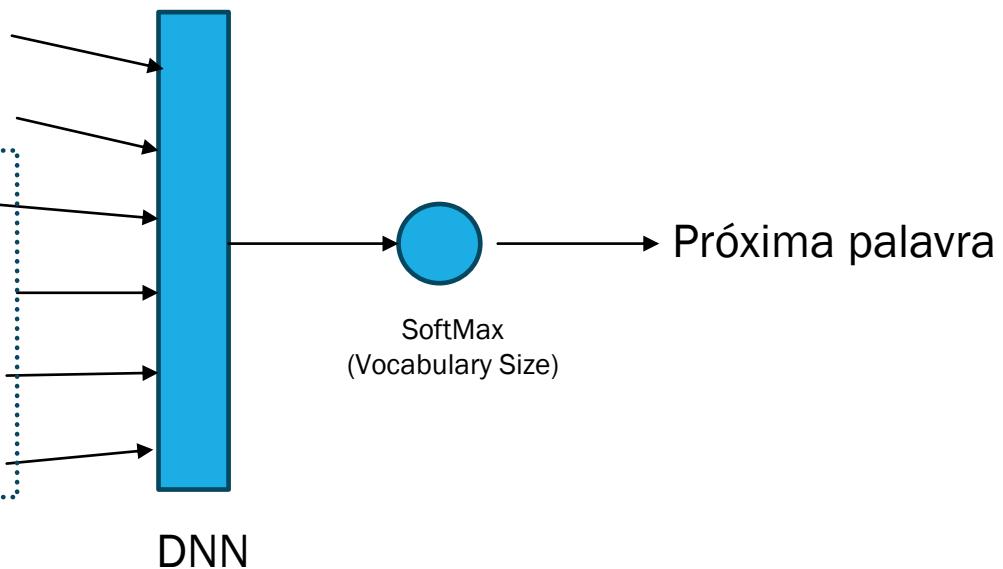
- Representação de atributos (features)
- Número de atributos ao redor de 1000 dá bom resultado
- Normalizar (Propriedades numéricas boas)
- Diferença dos Vetores $E_{\text{laranja}} - E_{\text{verde}} \approx 0$
- Similaridade do Cosseno: $\text{sim}(u,v) = u.v / \|u\|.\|v\|$
 - Muito utilizada!
- Semantica Distribuida
- Mapa t-SNE

Atributo	Laranja	Verde	Homem	Mulher	Rei	Rainha
Cor	0,97	0,98	0,01	0,01	0,02	0,03
Pessoa	0,1	0,02	0,99	0,99	0,98	0,97
Masculino	0,01	0,2	1,0	0,0	1,0	0,0
Feminino	0,01	-0,01	0,01	0,99	-0,01	0,98
Realeza	0,0	0,0	0,01	-0,02	0,97	0,98



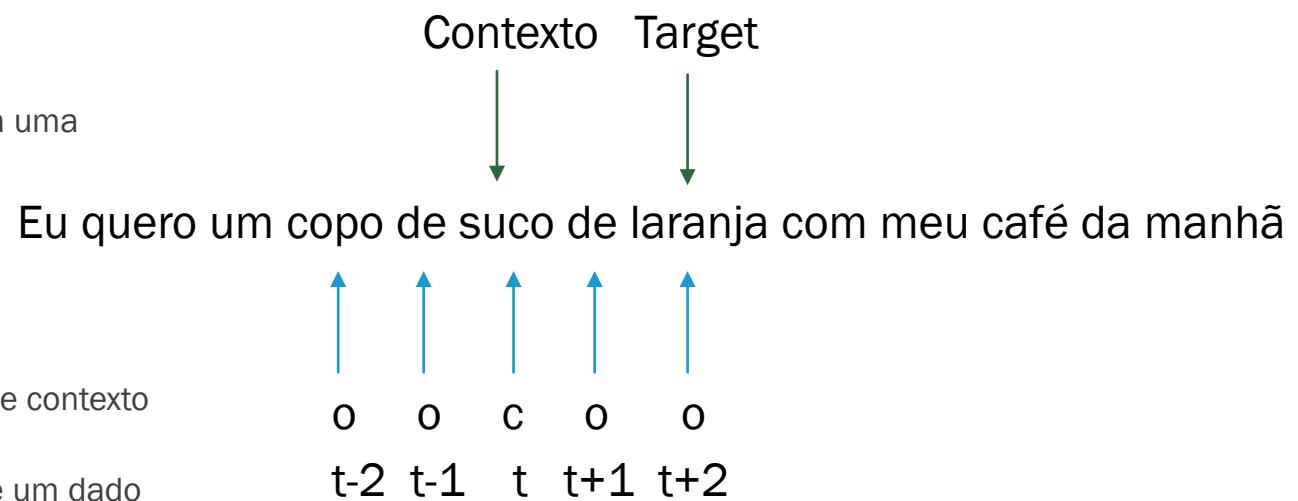
APRENDENDO WORD VECTOR EMBEDDING

- Eu quero um copo de suco de __?__
- Eu -----> O_{4343} -----> E -----> e_{4343}
- quero-----> O_{9665} -----> E -----> e_{9665}
- um -----> O_{8678} -----> E -----> e_{8678}
- copo-----> O_{3852} -----> E -----> e_{3852}
- de-----> O_{6163} -----> E -----> e_{6163}
- suco-----> O_{6257} -----> E -----> e_{6257}



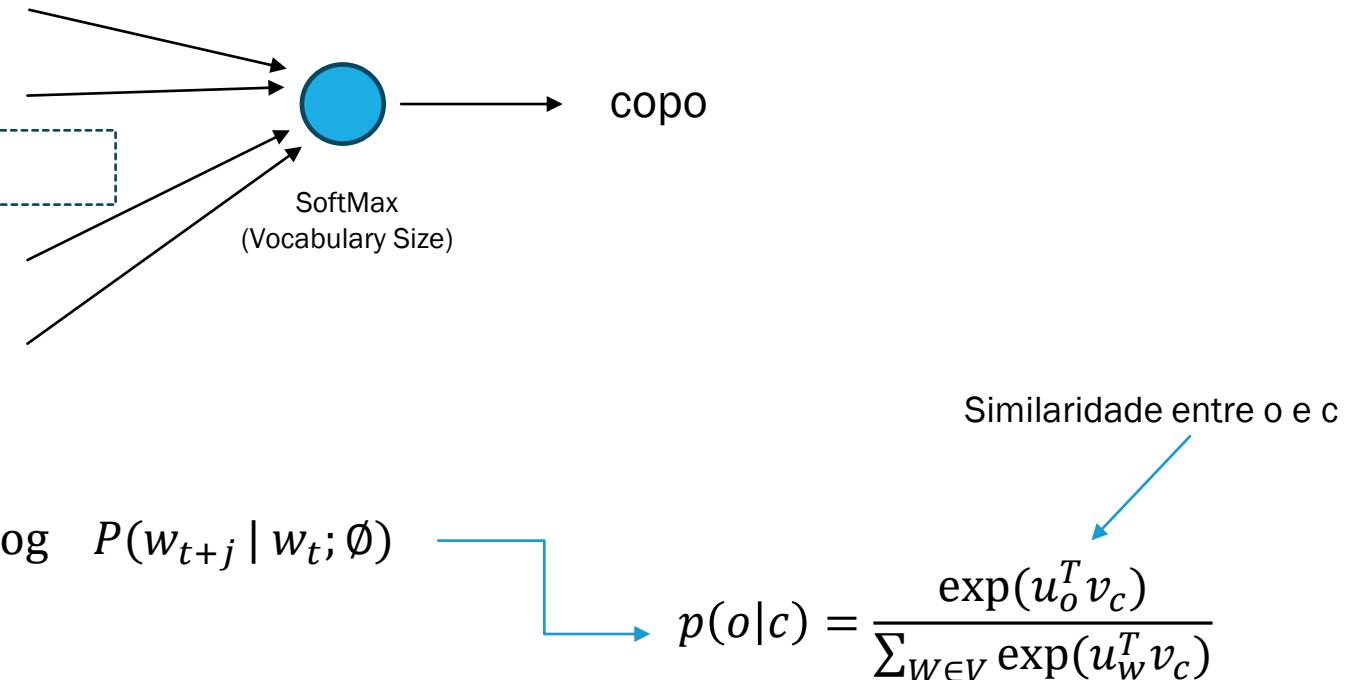
WORD2VEC

- Word2vec é um framework para aprender WordVectors (Mikolov et al. 2013)
- CBOW - Continuous Bag-of-Words:
 - Seleciona uma palavra contexto “c” e prevê uma palavra target “t” a uma distância fixa positiva ou negativa.
- Skip Gram:
 - Ter um grande corpus de texto : uma longa lista de palavras
 - Cada palavra de um vocabulário é representada por um vetor
 - Passe por cada posição t no texto, que tenha uma palavra central c e contexto (outside) o
 - Use a similaridade de cada c and o para calcular a probabilidade de um dado c (ou vice-versa)
 - Vá ajustando os vetores para maximizar as probabilidades



WORD2VEC – SKIP GRAM

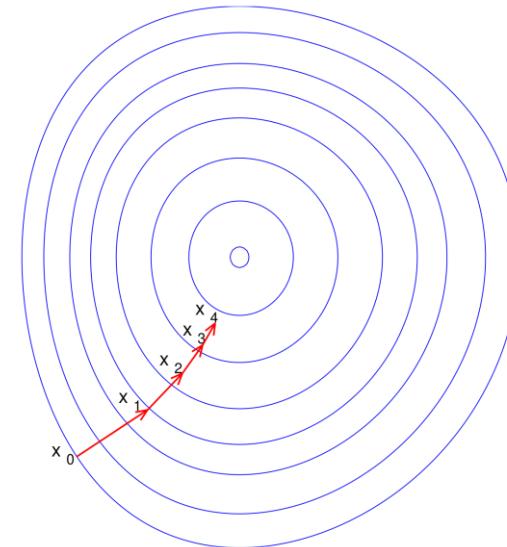
- Eu -----> O_{4343} -----> E -----> e_{4343}
- quero-----> O_{9665} -----> E -----> e_{9665}
- um -----> O_{8678} -----> E -----> e_{8678}
- copo-----> O_{3852} -----> E -----> e_{3852}
- de-----> O_{6163} -----> E -----> e_{6163}
- SUCO-----> O_{6257} -----> E -----> e_{6257}



Usar: Softmax Hieráquico

GRADIENT DESCENT

- Algoritmo para encontrar o mínimo valor de um função (se possível convexa)
- Usar uma estratégia incremental baseada na inclinação negativa num ponto (derivada)
- Alfa é a taxa de aprendizado – learning rate
- Bastante pesado
- Stochastic Gradient Descend
 - Mais leve
 - Amostra janelas no corpus



$$\theta^{New} = \theta^{Old} - \alpha \nabla_{\theta} J(\theta)$$

$$\theta_J^{new} = \theta_J^{old} - \alpha \frac{\partial}{\partial \theta_J^{old}} J(\theta)$$