

PROPOR 2018

Processamento de Linguagem Natural por meio de Redes Neurais Profundas

Teoria e aplicações



Cronograma

Parte 1 - Processamento de Linguagem com redes profundas:

- Introdução a Aprendizado de Máquina
- Redes Neurais Artificiais
- Melhorando a memória das Redes Recorrentes
- Aprendizado de máquina para PLN

Introdução a Aprendizado de Máquina

Algoritmos de **aprendizado de máquina** constroem modelos a partir de dados a fim de fazer previsões ou tomar decisões.

Regressão

$\mathbf{x}_1, \dots, \mathbf{x}_N$



y_1, \dots, y_N

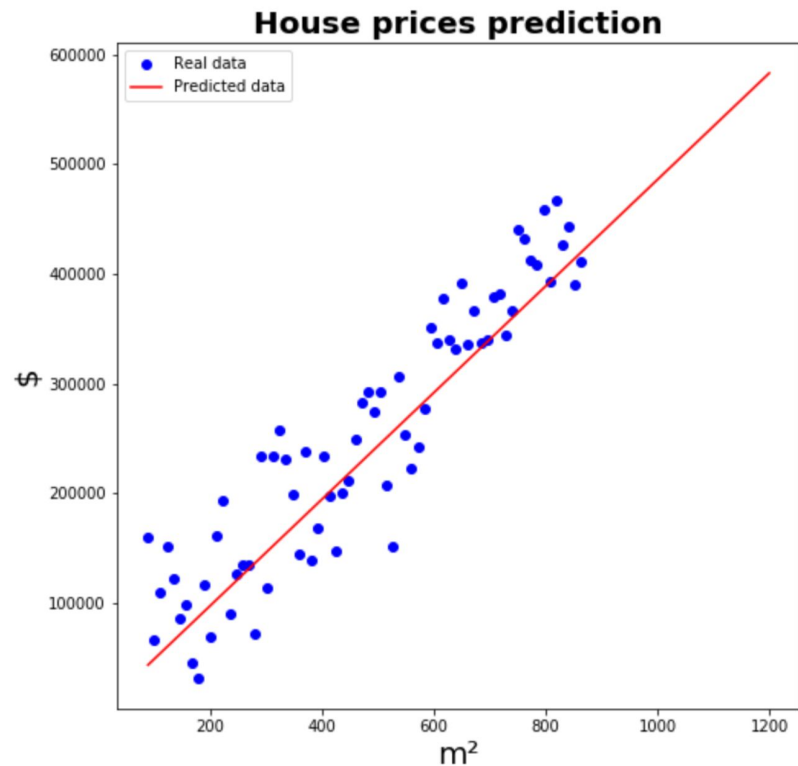
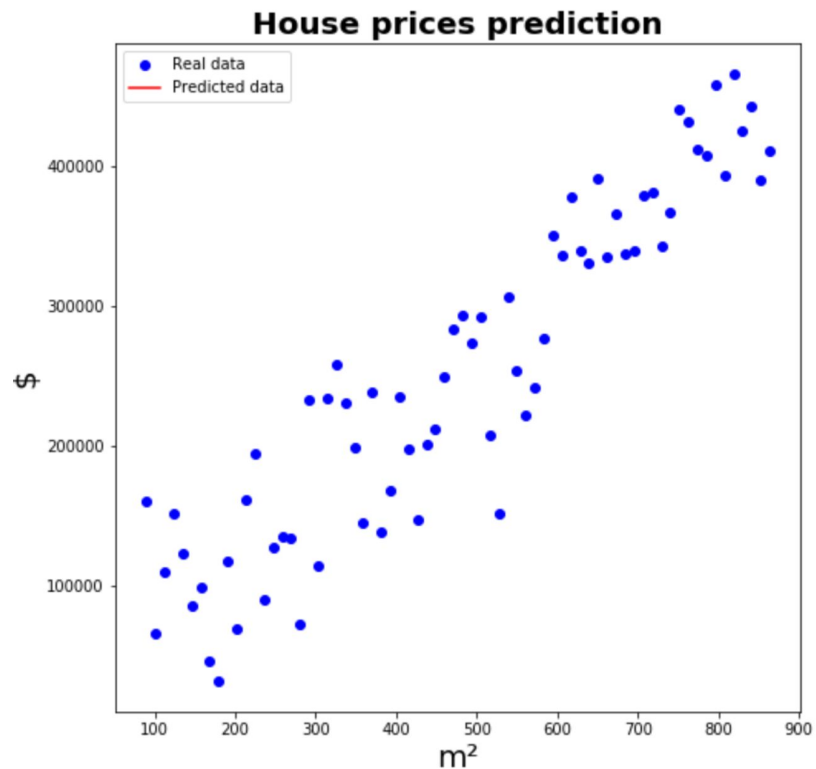
Classificação

$\mathbf{x}_1, \dots, \mathbf{x}_N$



L_1, \dots, L_N

Regressão



Classificação: Fashion MNIST



$y = \text{camisa}$



$y = \text{vestido}$

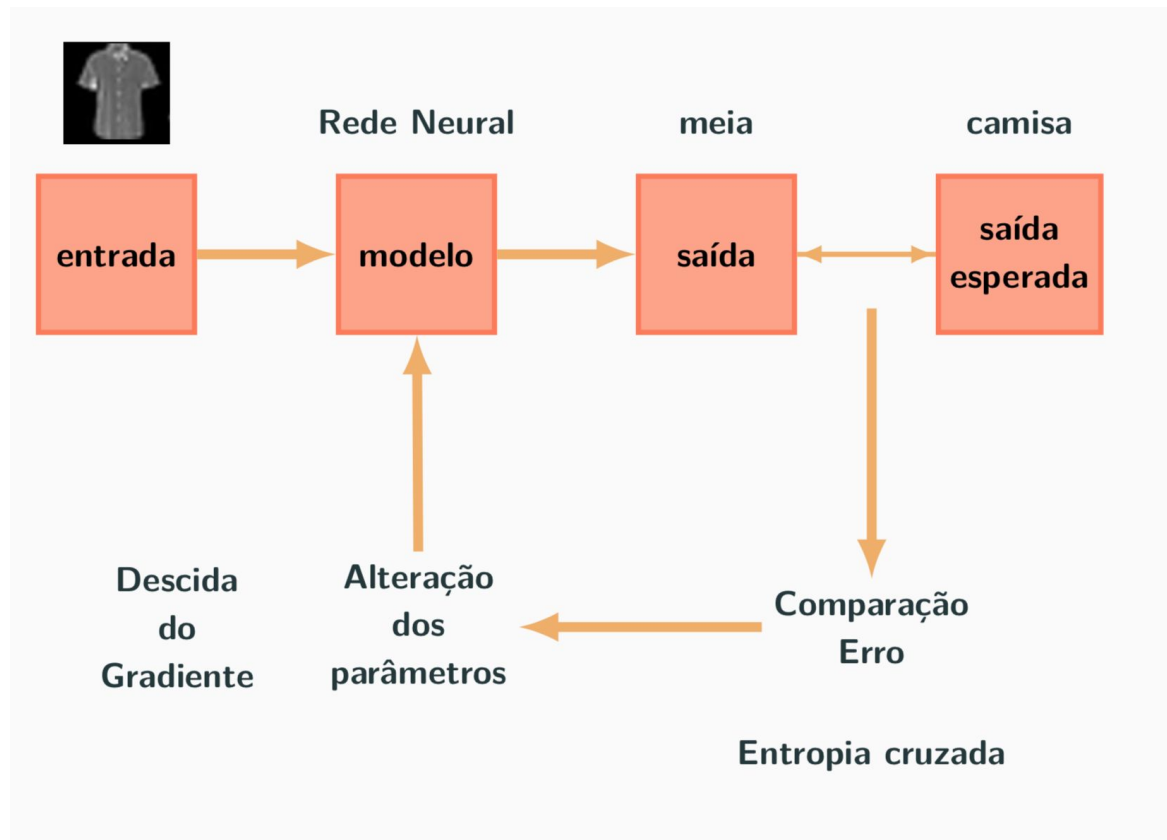


$y = \text{bolsa}$



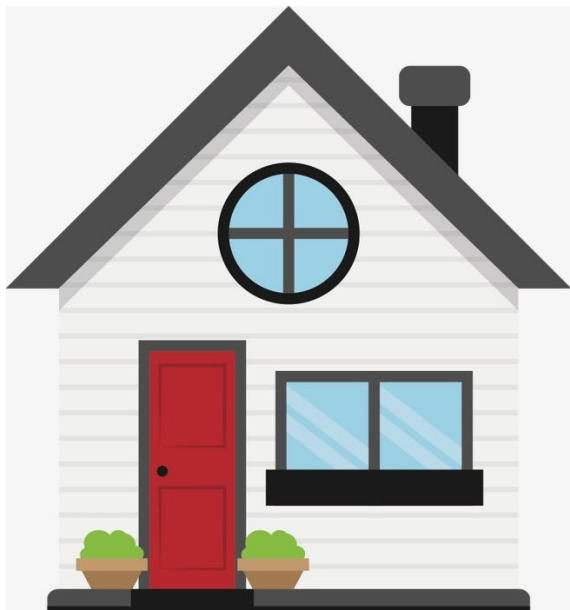
$y = \text{tênis}$

Aprendizado de Máquina: Pipeline



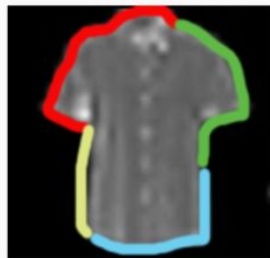
Aprendizado de Máquina: Features

- São características ou traços do objeto do aprendizado



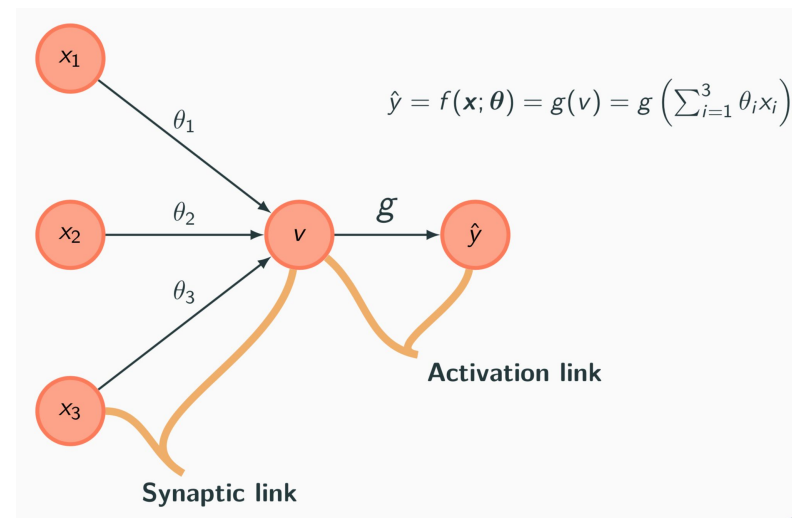
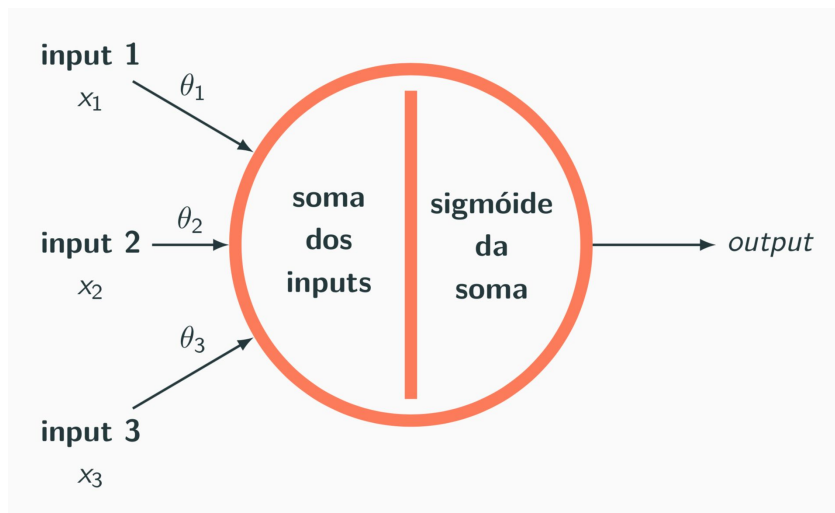
Exemplo 1: Na regressão para encontrar uma função de preço de casas, features podem ser dados como: área construída, distância do centro, tamanho do terreno, preço, etc.

Aprendizado de Máquina: Features

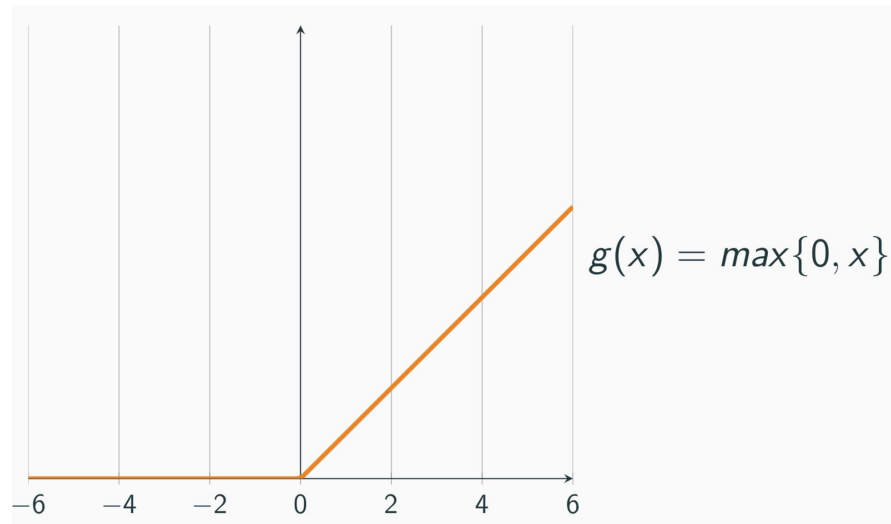
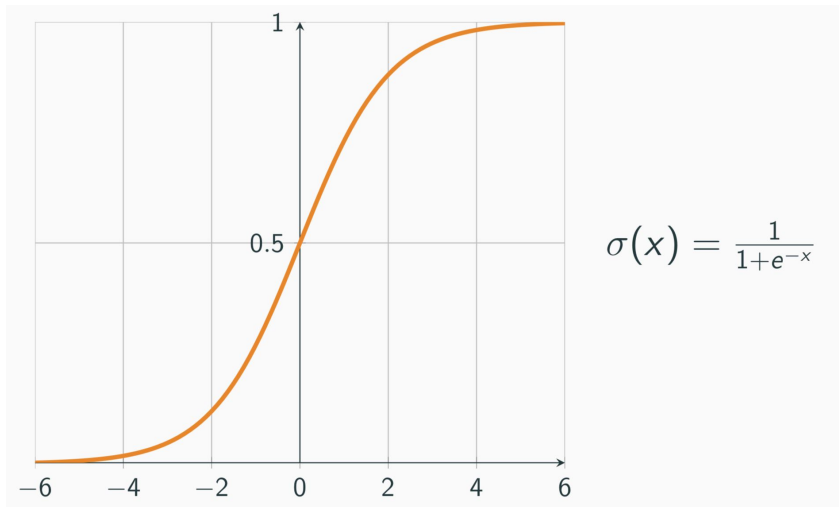


Exemplo 2: Em reconhecimento de imagem, feature pode ser um padrão de contorno, cor, textura, etc.

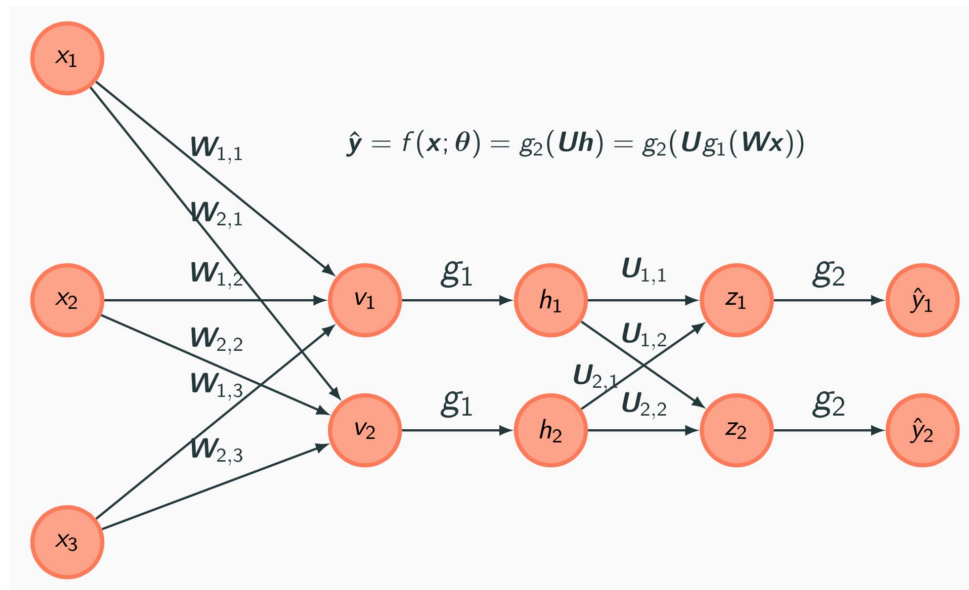
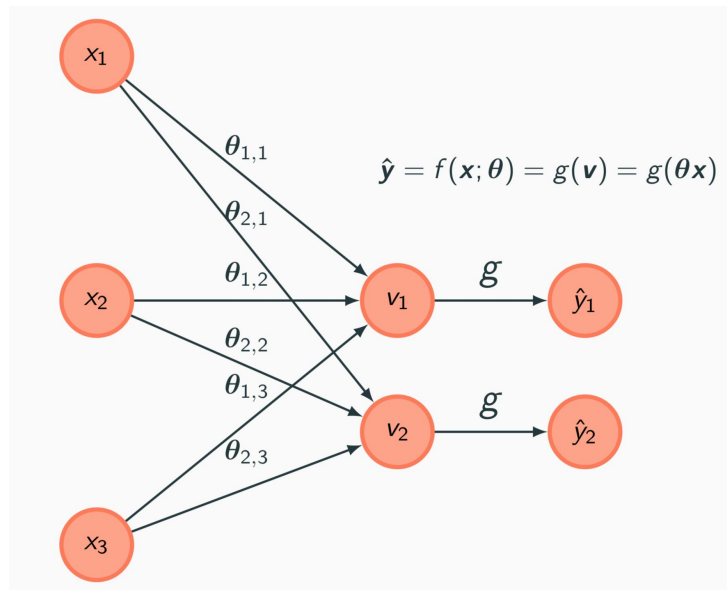
Redes Neurais Artificiais: Perceptron



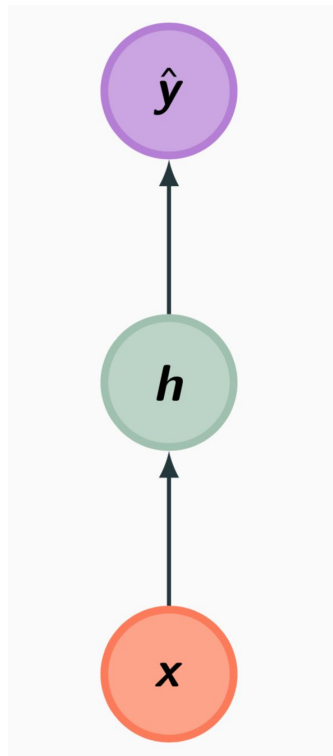
Redes Neurais Artificiais: Ativação



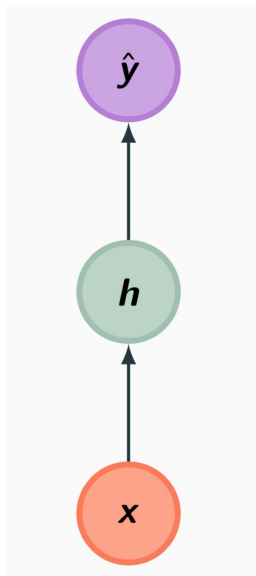
Redes Neurais Artificiais: Feedforward Multi Layer



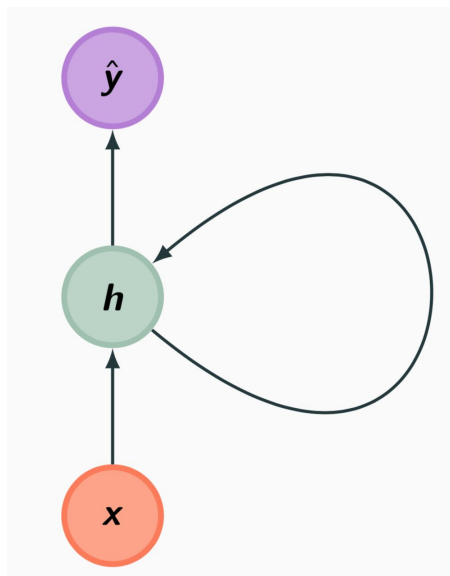
Versão resumida de uma Feedforward



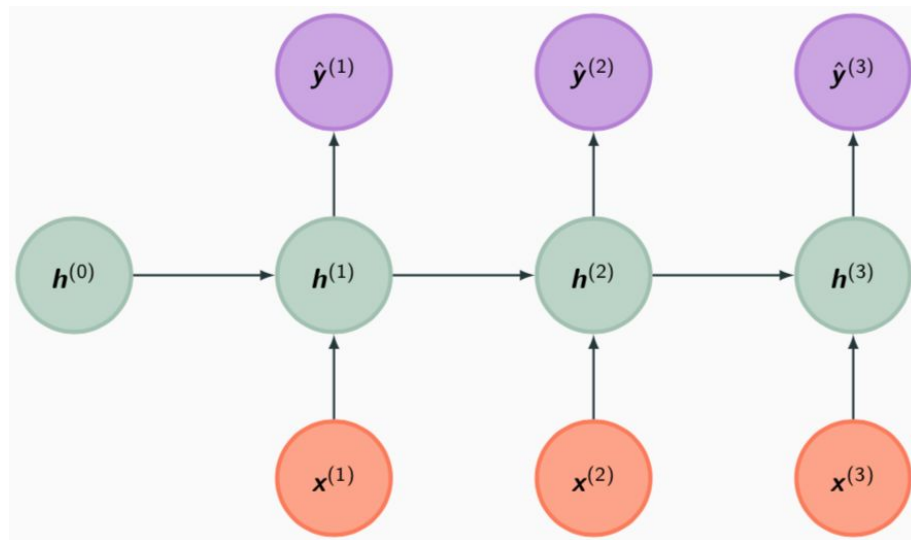
Rede Neural Recorrente



feedforward



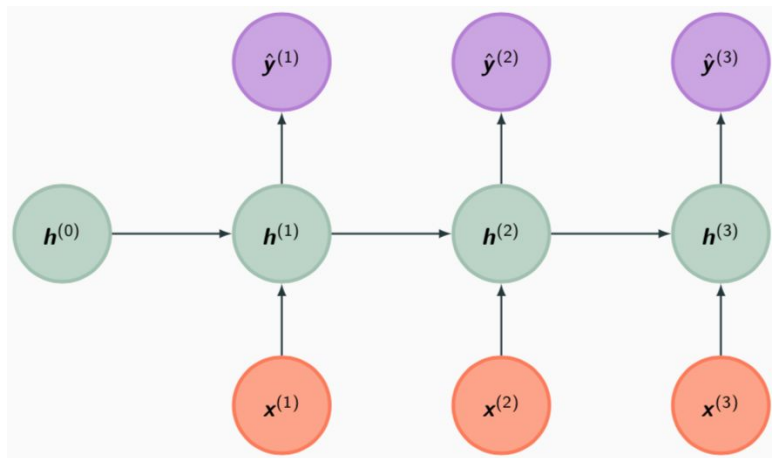
RNN - versão compacta



RNN - versão expandida

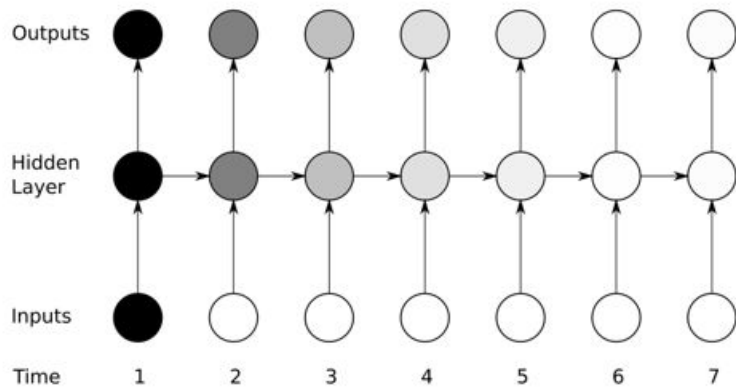
Rede Neural Recorrente

- Redes neurais recorrentes são capazes de representar uma função de distribuição de probabilidade, que pode ser associada a muitos problemas do mundo real.

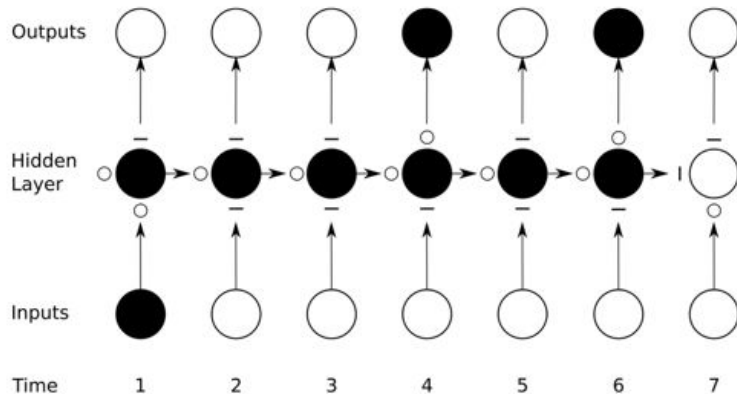


$$p(x_t = 1 | x_{t-1}, \dots, x_1)$$

Rede Neural Recorrente: Vanishing gradient

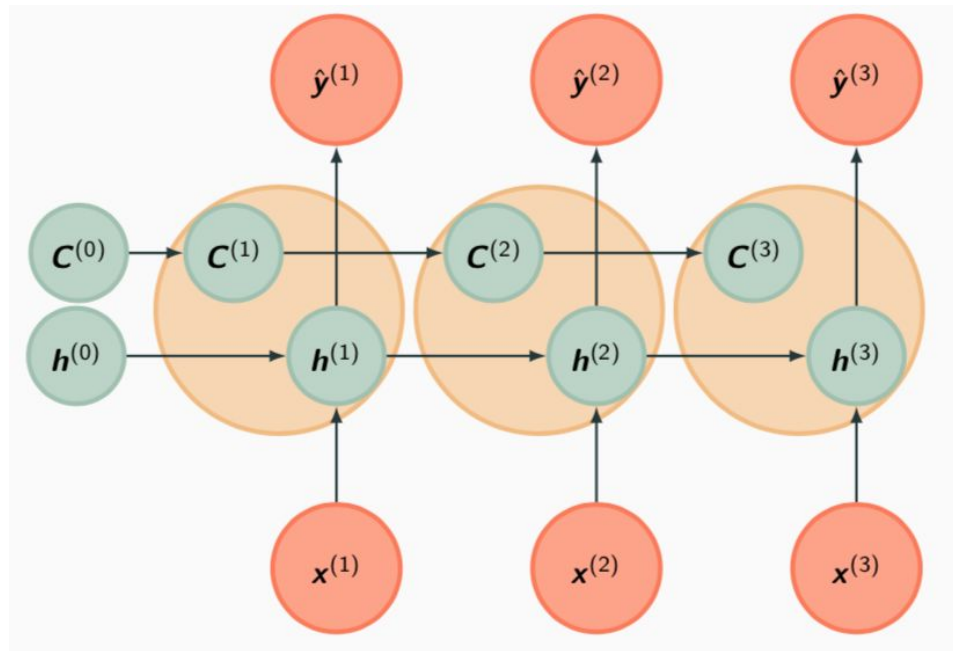
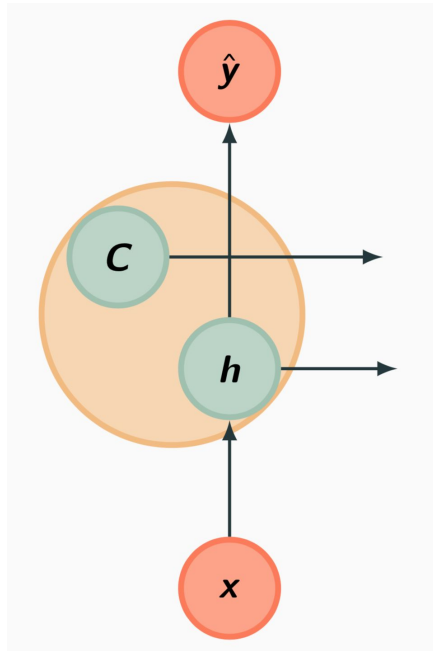


Influência do 1º nó da rede nos seguintes



Influência do 1º nó da rede nos seguintes com LSTM

Rede Recorrente: LSTM



Aprendizado de Máquina em PLN

Algoritmos de aprendizado de máquina vem sendo utilizado largamente como proposta de solução de diversos problemas de processamento de linguagem natural:

- Detecção de Spam
- Modelo de linguagem
- Traduções automáticas
- Agentes Virtuais (chatbots)
- Análise de sentimento
- Reconhecimento de fala
- Legenda automática
- Extração de Entidades

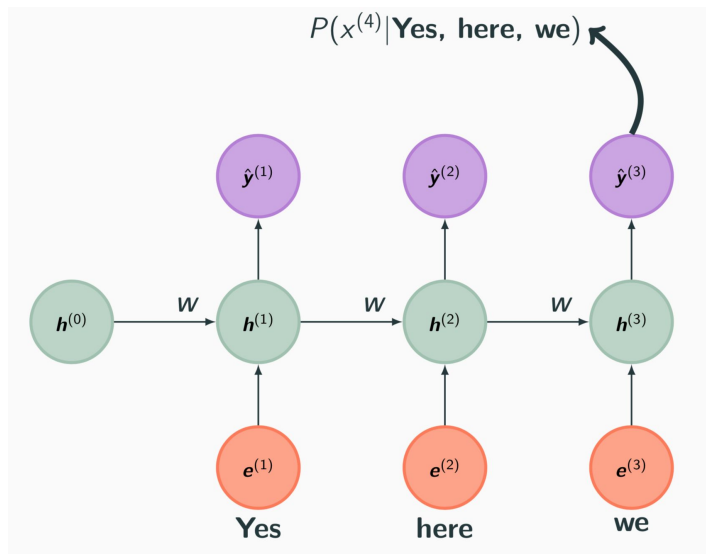
Features em Processamento de Linguagem

Feature Engineering - Conhecimento de especialista para estabelecer features, essas features geralmente são baseadas em **n-gramas** que é uma subsequência de **n** elementos de uma sequência dada.

Word embeddings - Conjunto de técnicas para mapear sentenças para vetores de números reais, as mais conhecidas são: redes neurais recorrentes, redução de dimensionalidade (SVD, PCA, etc), matriz de co-ocorrência.

De Palavras para Vetores com Rede Recorrente

- Mikolov propôs em 2012 um modelo para geração de vetores (word2vec), onde o treinamento era realizado a partir de texto somente.



- Um dos resultados desse treinamento é uma representação vetorial para cada palavra, onde palavras que co-ocorrem no mesmo “contexto” possuem a norma vetorial pequena entre si.

Projeção Bidimensional (PCA)



Classificação de Texto: Spam



Olha o DESCONTO! De São Paulo para Uberlândia por apenas 100.81, CLIQUE AQUI!

Classificação de Texto: Detecção de Spam

Documento 1 (Categoria SPAM):

Olha o DESCONTO! De São Paulo para Uberlândia por apenas 100.81, CLIQUE AQUI!

Documento 2 (Categoria HAM):

Oi João, preciso alinhar com você apenas alguns pontos para a reunião de amanhã

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

**Olha
DESCONTO
para
preciso
você
apenas
para
reunião
CLIQUE**

Documentos para o
treinamento

Documentos
vetorizados
(1-grama)

Dicionário

Classificação de Texto: Detecção de Spam

Documento 1 (Categoria SPAM):

Olha o DESCONTO! De São Paulo para Uberlândia por apenas 100.81, CLIQUE AQUI!

caso geral:

$$f(x) = y$$

Documento 2 (Categoria HAM):

Oi João, preciso alinhar com você apenas alguns pontos para a reunião de amanhã

caso Spam:

$$f(\text{Documento}) = \text{Categoria}$$

Modelo de Linguagem com Redes Neurais

Chamamos modelo de linguagem uma distribuição de probabilidade sobre uma sequência de tokens em uma língua natural.

$$P(x_1, x_2, x_3, x_4) = p$$

Modelos de linguagem podem ser utilizados para:

- Reconhecimento de fala
- Tradução automática
- Autocompletar texto
- Correção de texto
- Sumarização de texto
- Resposta automatizada

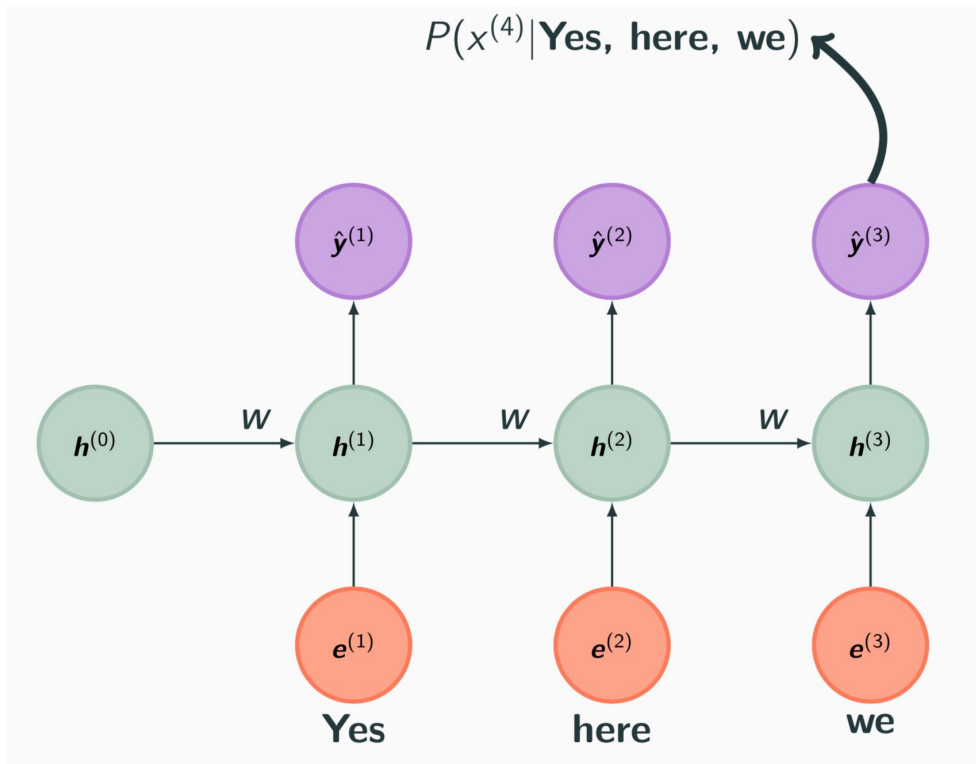
Modelo de Linguagem com Redes Neurais

Podemos usar um modelo para predição de dados sequencias: uma rede recorrente (RNN).
Nossa tarefa de aprendizado é estimar a distribuição de probabilidade:

$$P(x_n = \text{palavra}_{j^*} | x_1, \dots, x_{n-1})$$

Para qualquer $(n - 1)$ sequência de palavras x_1, \dots, x_{n-1} .

Modelo de Linguagem: Exemplo

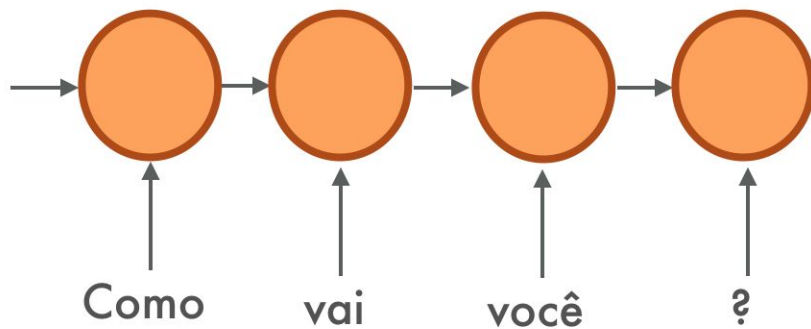


Cronograma

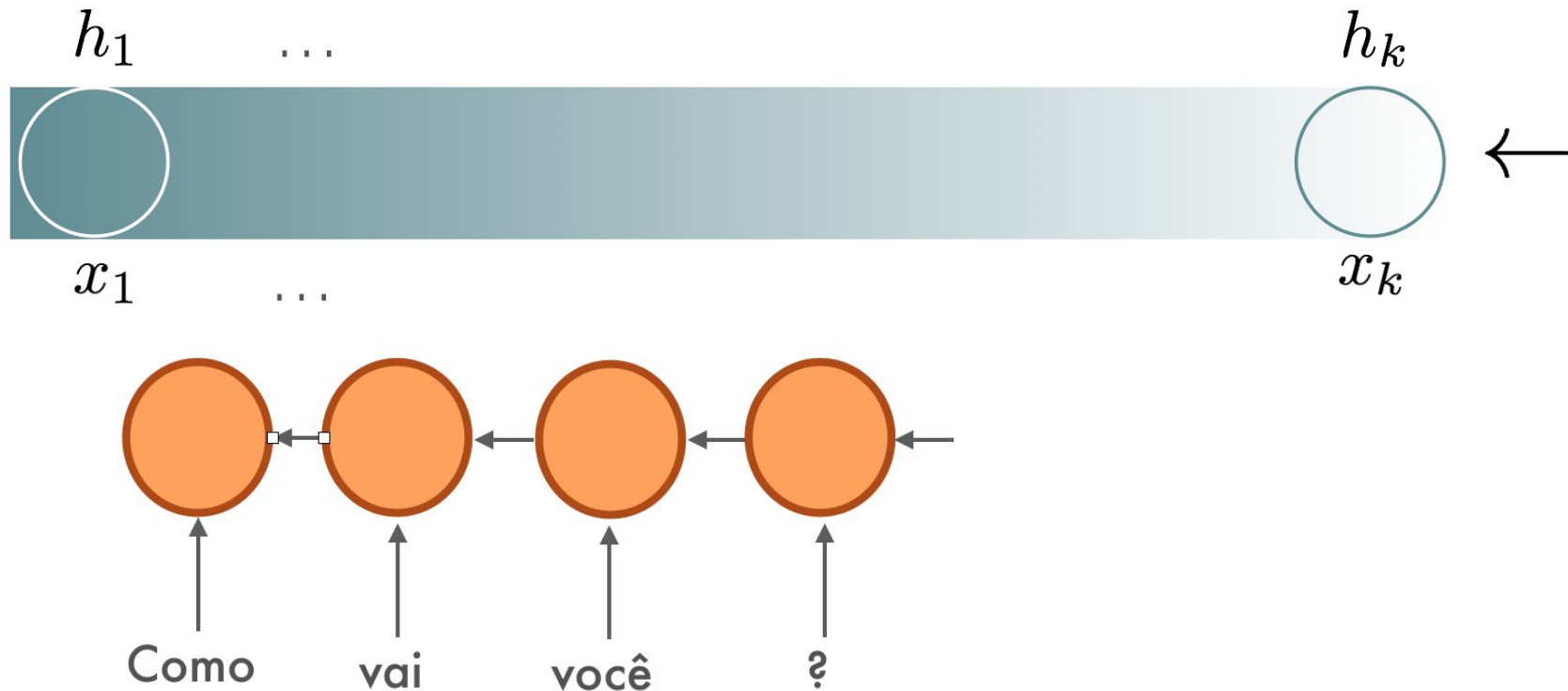
Parte 2 - Arquitetura de codificação decodificação com Atenção Neural:

- Rede Recorrente Bidirecional
- Codificação-Decodificação
- Exemplo: Aprendizado de diálogo
- Exemplo: Tradução automática
- Mecanismo Atenção Neural

Redes Recorrentes: Pra frente



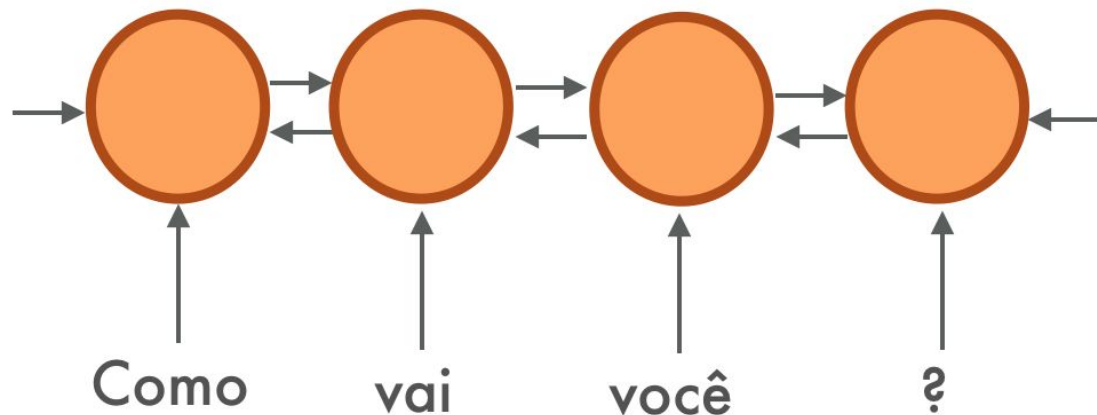
Redes Recorrentes: Para trás



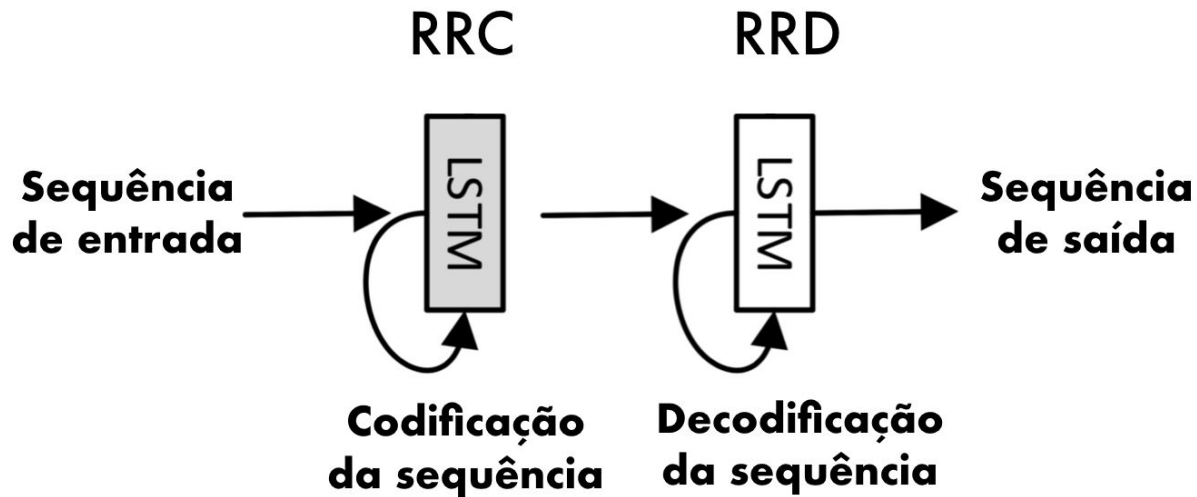
Redes Recorrentes: Bidirecionais

$$h_1 = \vec{h}_1 \cdot \overleftarrow{h}_1$$

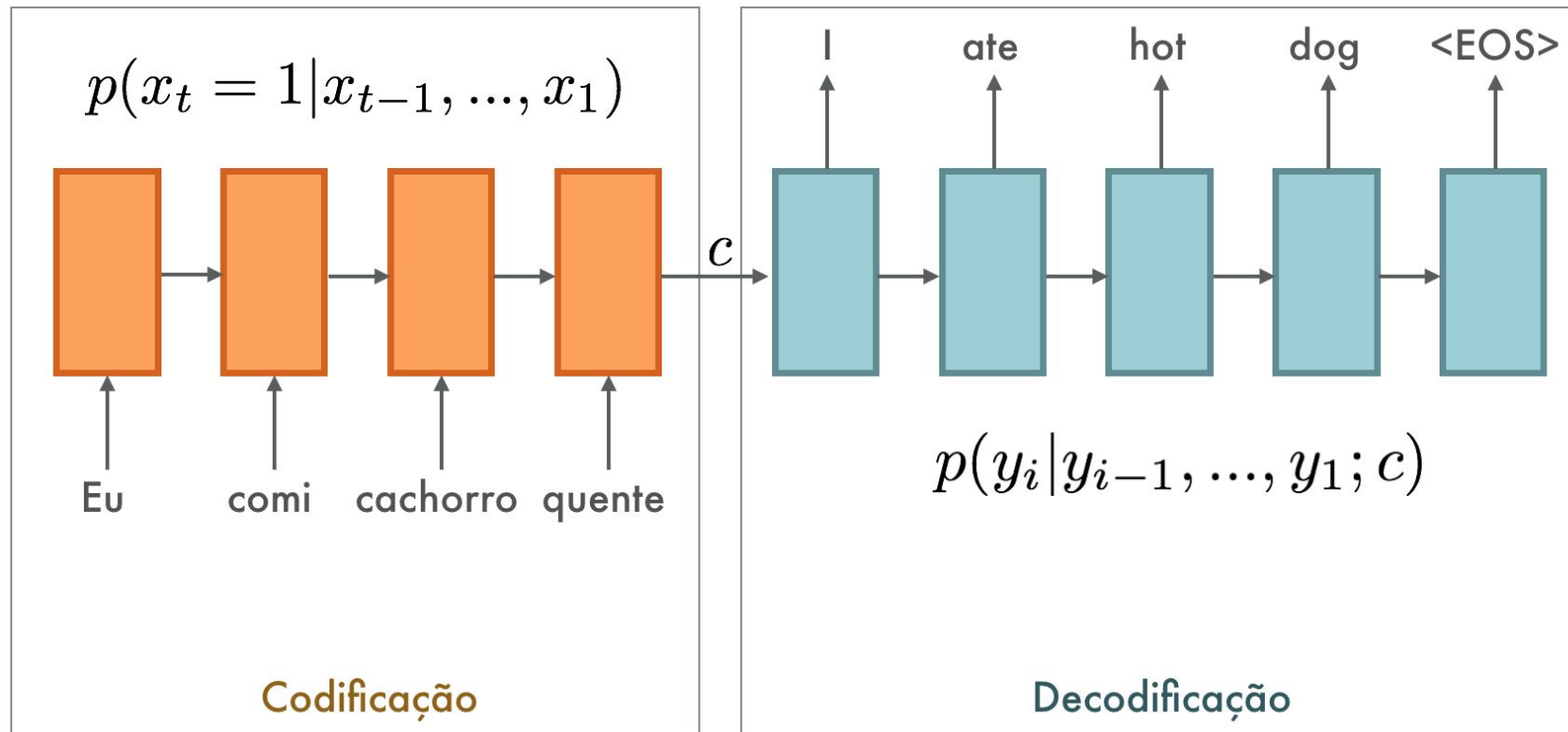
$$h_k = \vec{h}_k \cdot \overleftarrow{h}_k$$



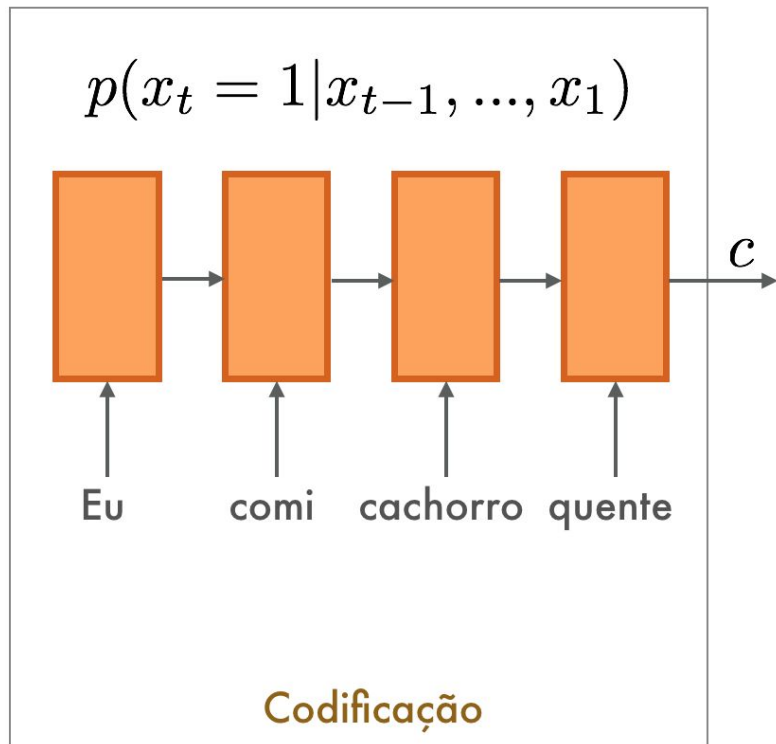
Codificador-Decoder (LSTM)



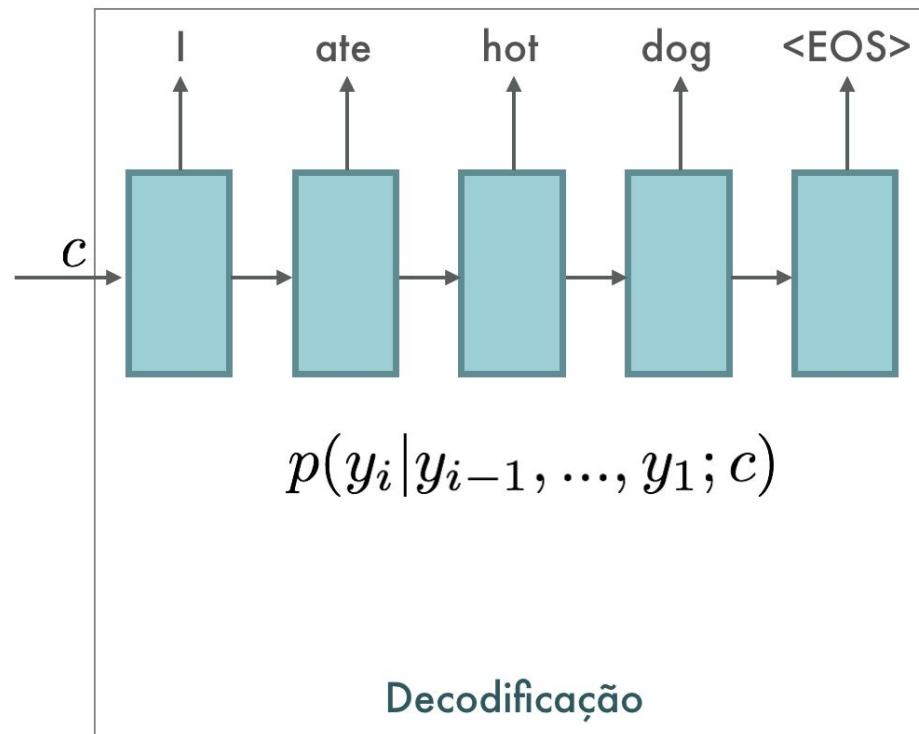
Codificador-Decoder (LSTM)



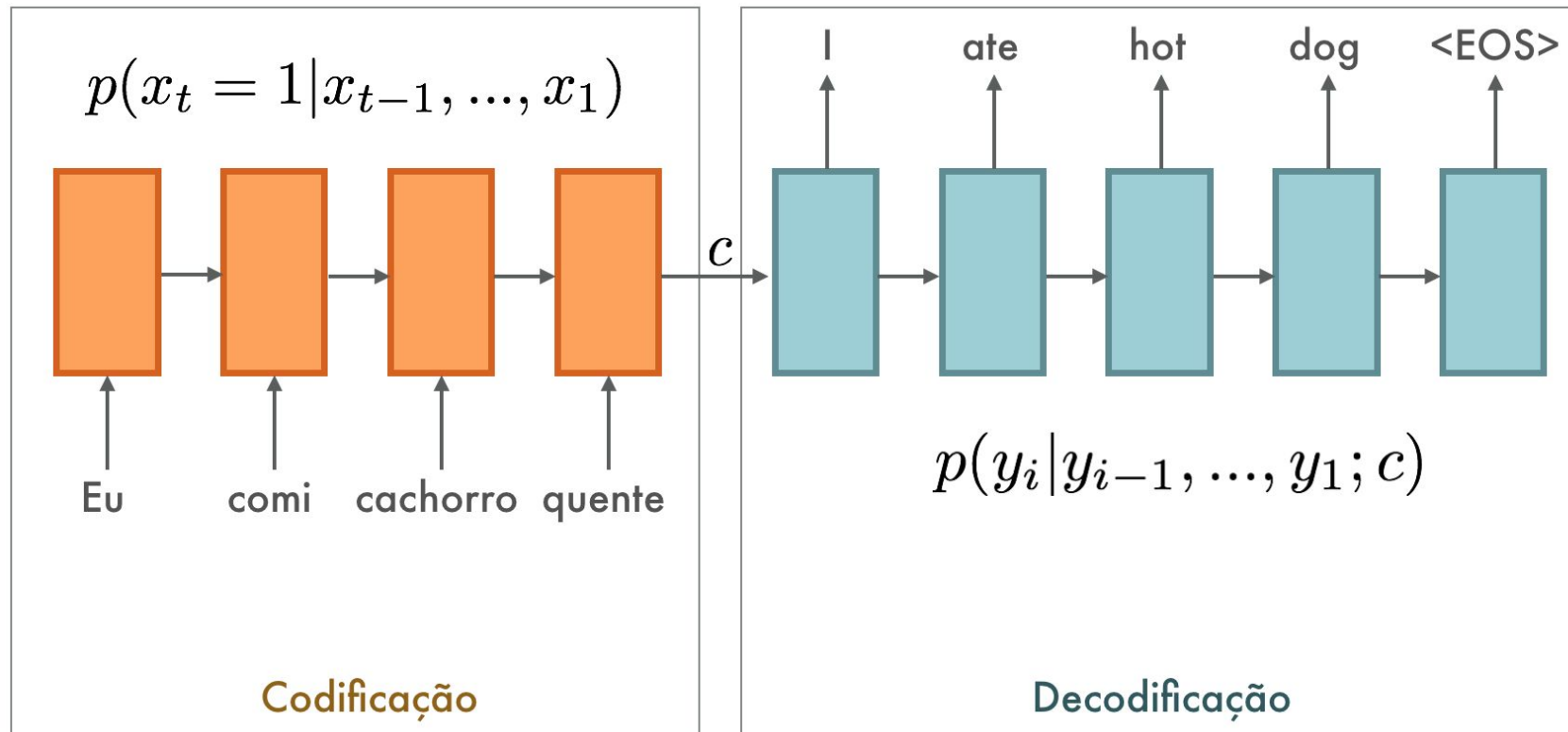
Codificador-Decodificador (LSTM)



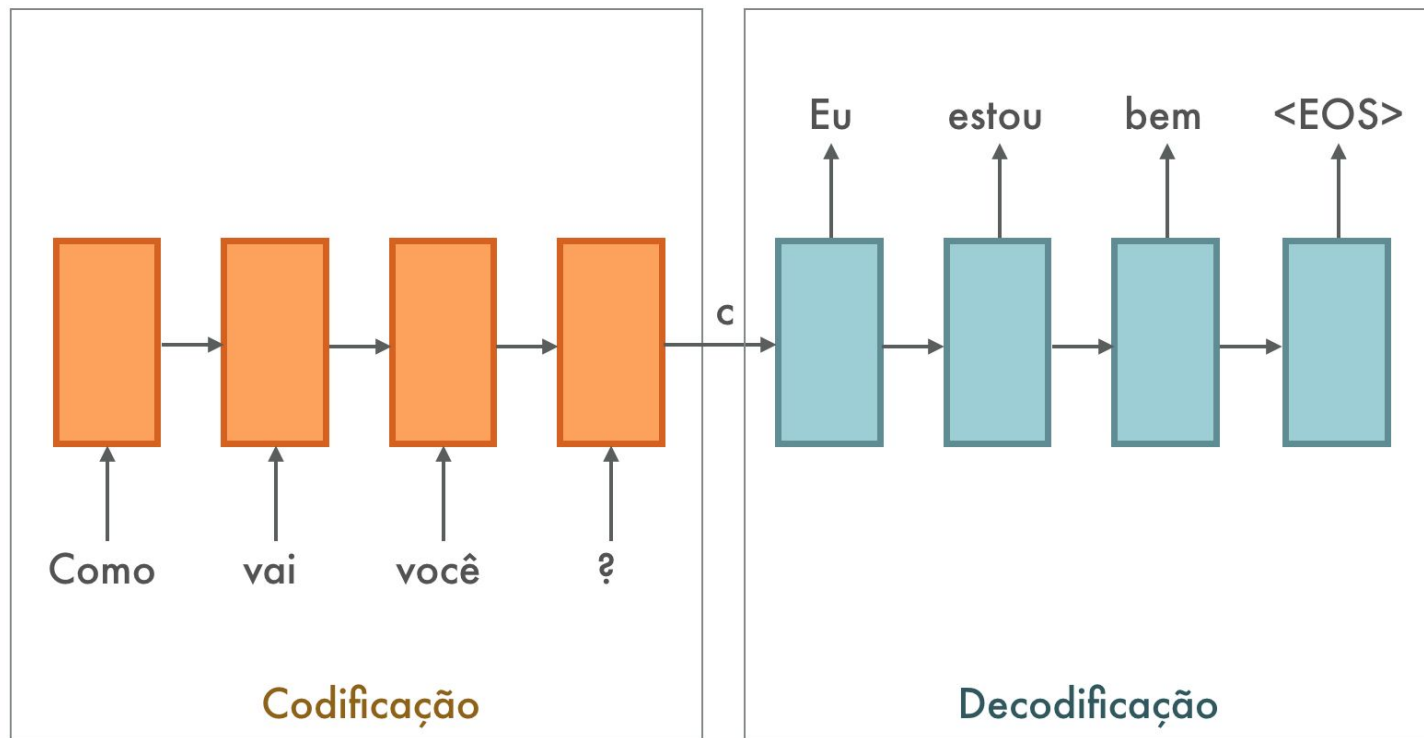
Codificador-Decoder (LSTM)



Codificador-Decoder (LSTM)



Exemplos: (Diálogo)



Algumas considerações sobre o modelo

- O que é esperado de um bom modelo de tradução ?
- É correto associar um problema de tradução a um problema de aprendizagem de sequência ?
- Associação direta entre objetos inteiros ?
- Aprendizado composicional ?

Alinhamento entre traduções

- É possível capturar a ideia de alinhamento entre subsequências dentro no contexto de uma tradução?
- Composicionalidade “semântica” melhora o aprendizado de maneira geral.

Eu comi **cachorro** quente



I ate hot **dog**

Atenção Neural

- Inspirado pela atenção visual dos humanos;
- Humanos estabelecem uma região focal no campo visual;



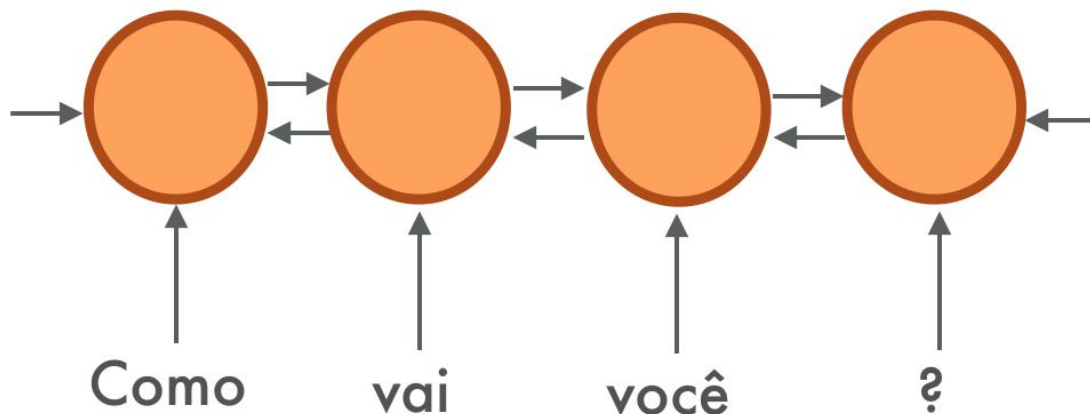
A woman is throwing a frisbee in a park.



A dog is standing on a hardwood floor.

Codificador Decodificador com Atenção

Mudança no codificador (Bidirecional)



Codificador Decodificador com Atenção

Mudança no decodificador

$$p(y_i | y_1, \dots, y_{i-1}, c) = g(y_{i-1}, h_t, c) \quad \text{antigo}$$

$$p(y_i | y_1, \dots, y_{i-1}, c_i) = g(y_{i-1}, s_i, c_i) \quad \text{novo}$$

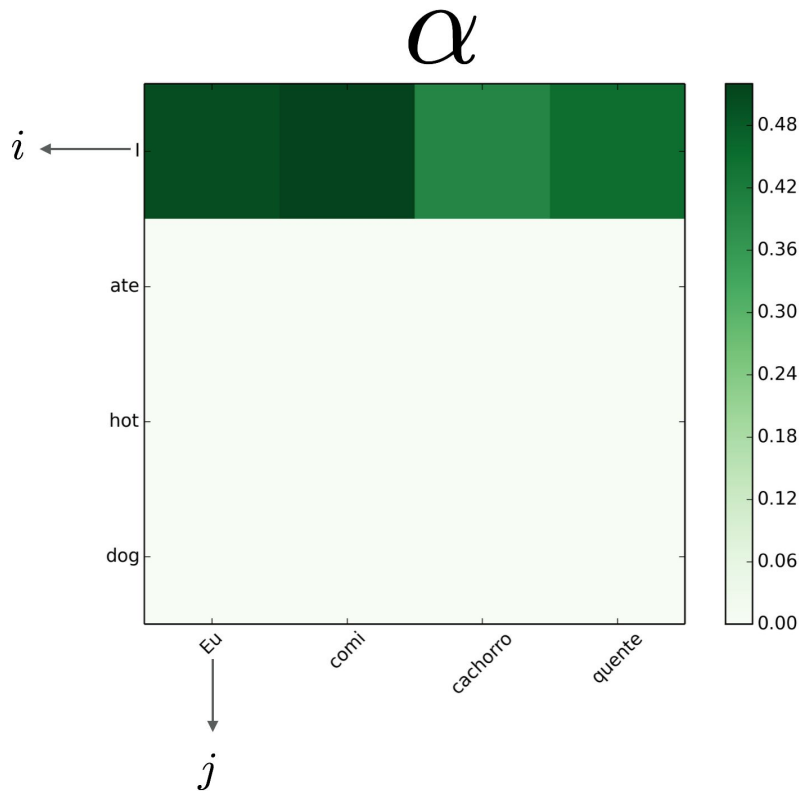
$$s_i = f(s_{i-1}, y_{i-1}, c_i)$$

Atenção Neural: Quem é o c_i ? (Exemplo)

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$$

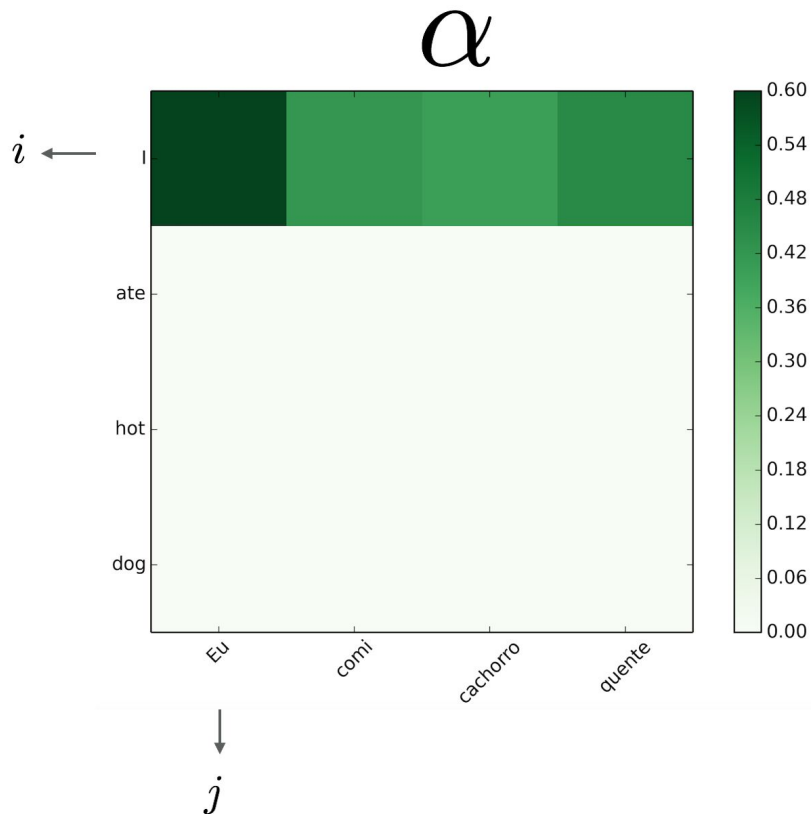
$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})}$$

$$e_{ij} = a(s_{i-1}, h_j)$$

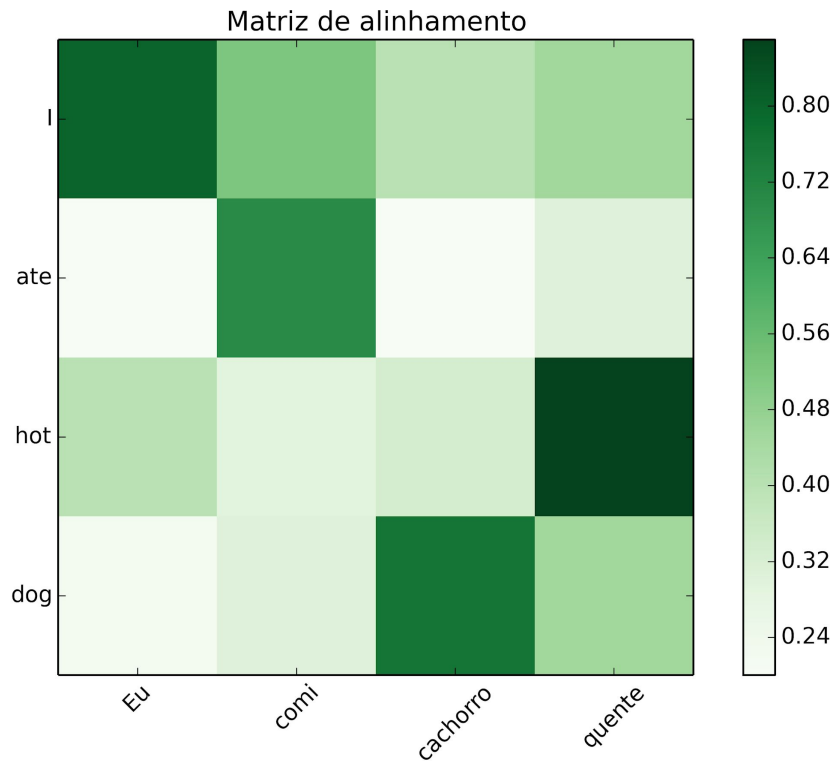


Atenção Neural: Quem é o c_i ? (Exemplo)

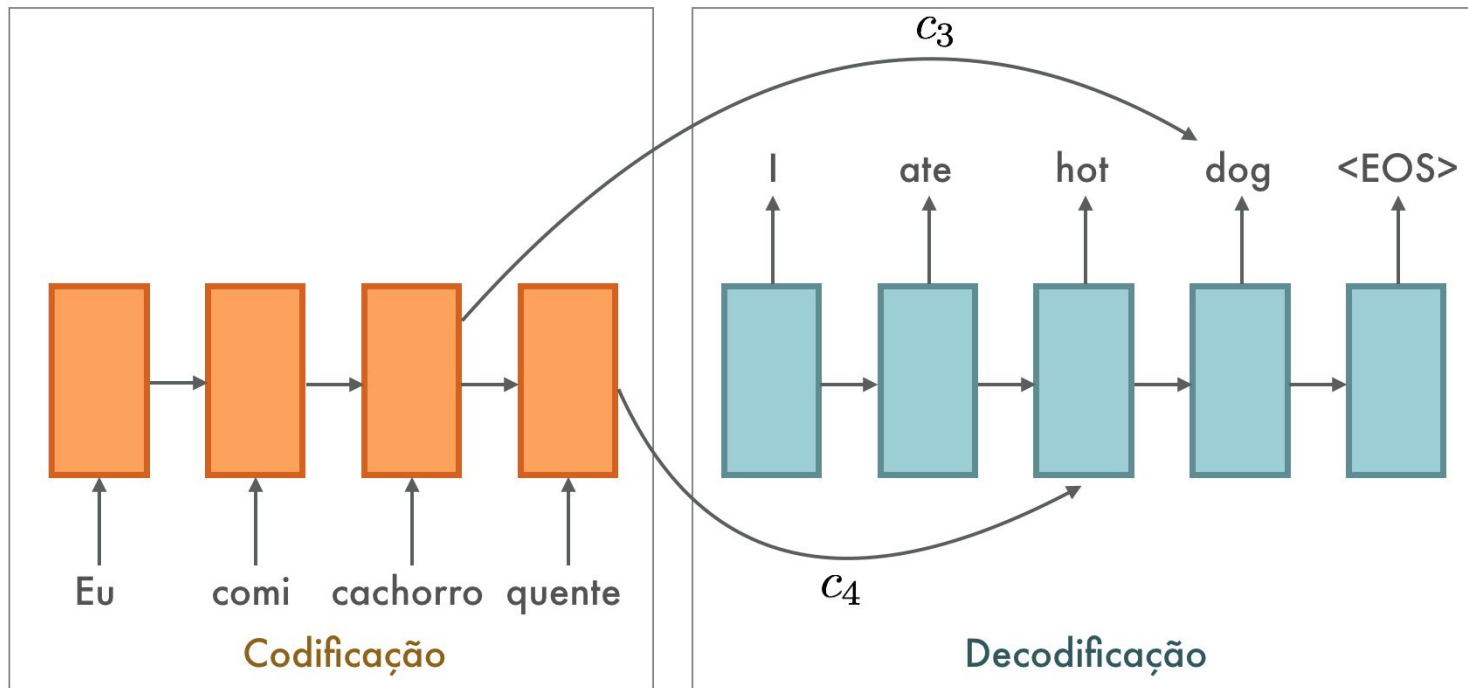
- Na medida em que o treinamento vai passando por exemplos em que a palavra “I” se está associada a palavra “Eu”, a posição $\alpha(I, \text{Eu})$ tende a aumentar de valor, consequência da softmax.



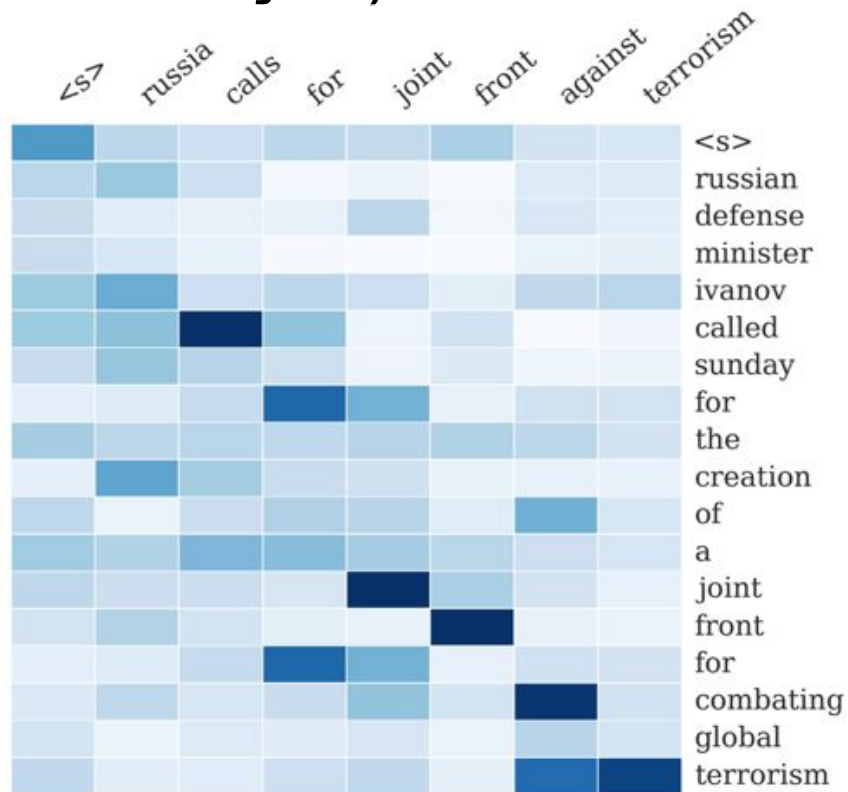
Atenção Neural (Exemplo)



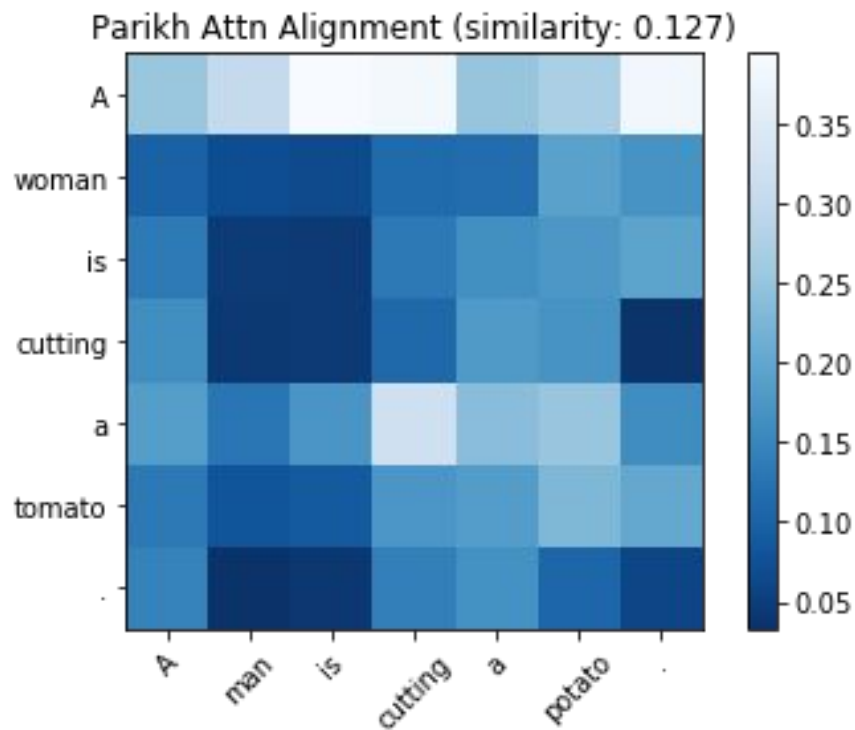
Atenção Neural (Exemplo)



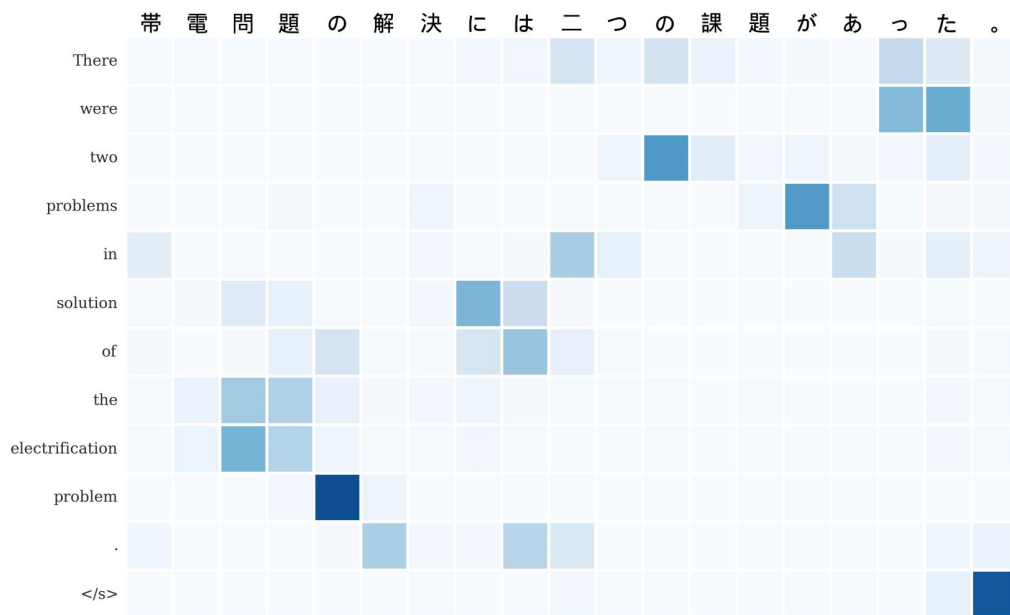
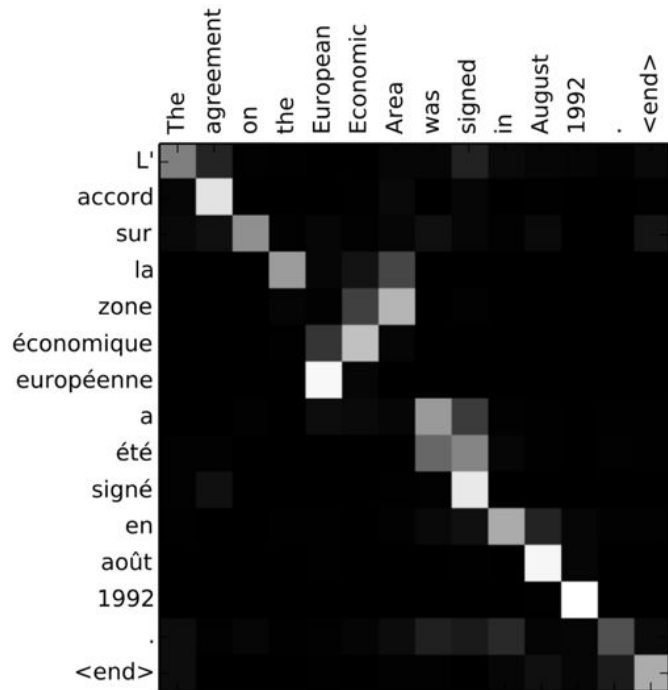
Exemplos: (Sumarização)



Exemplos: (Similaridade semântica entre tarefas)

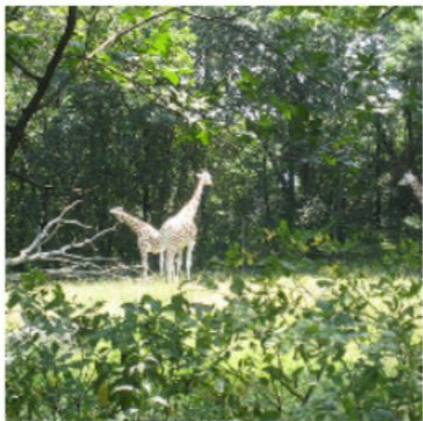


Exemplos: (Tradução)

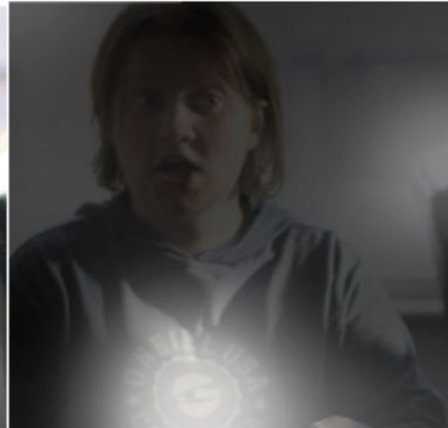


Atenção neural (Erros)

Assim como a atenção visual humana, o mecanismo de atenção neural artificial também pode errar.



A large white bird standing in a forest.



A woman holding a clock in her hand.

Referências Bibliográficas

HAYAT, Aadil; SUNIL, Masare Akshay. A Neural Conversational Model.

Mikolov, Tomas - Efficient Estimation of Word Representations in Vector Space, 2013.

BAHDANAU, Dzmitry; CHO, Kyunghyun; BENGIO, Yoshua. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473, 2014.

Bishop, Christopher M - Neural networks for pattern recognition, 1995.

Graves, Alex - Neural Networks: Supervised Sequence Labelling with Recurrent Neural Networks, 2012.

KIM, Yoon et al. Structured attention networks. arXiv preprint arXiv:1702.00887, 2017.

Obrigado!

Fabiano Luz

<https://www.linkedin.com/in/fabianoluzbr/>

fluz@ime.usp.br

Felipe Salvatore

<https://www.linkedin.com/in/felipe-salvatore-99522576/>

felsal@ime.usp.br

Marcelo finger

<https://www.linkedin.com/in/marcelo-finger-89b258/>

mfinger@ime.usp.br