



Formação Inteligência Artificial

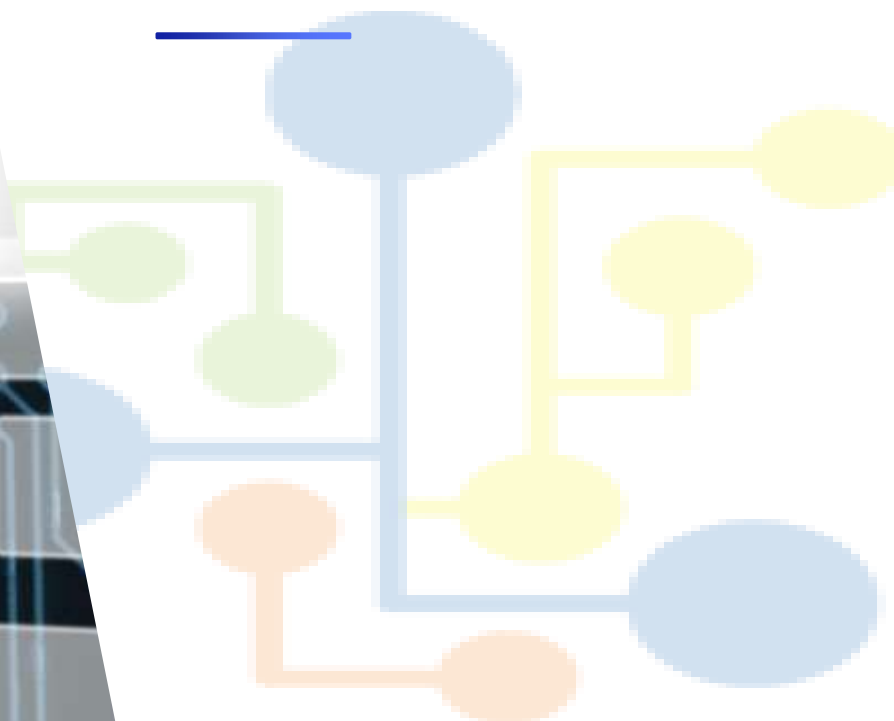


Processamento de Linguagem Natural



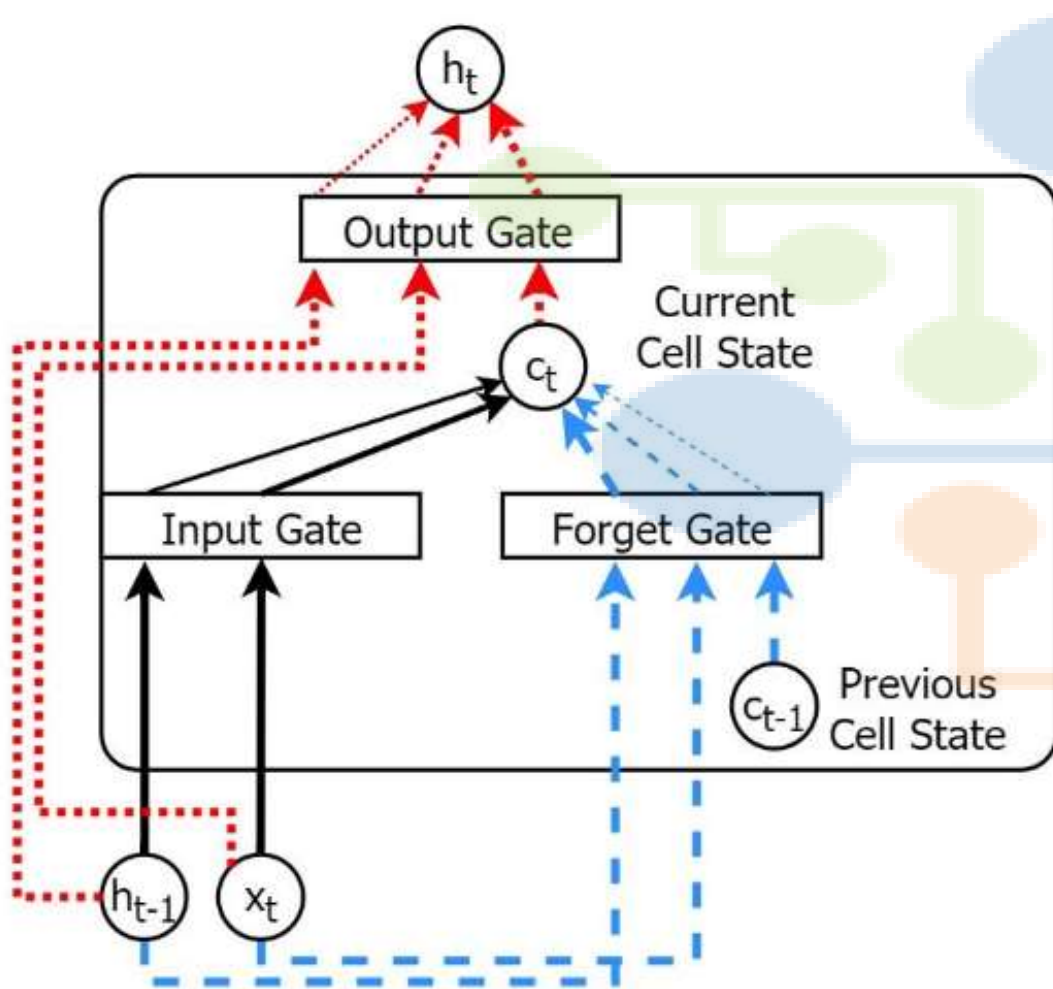


Deep Learning - Long-Short Term Memory





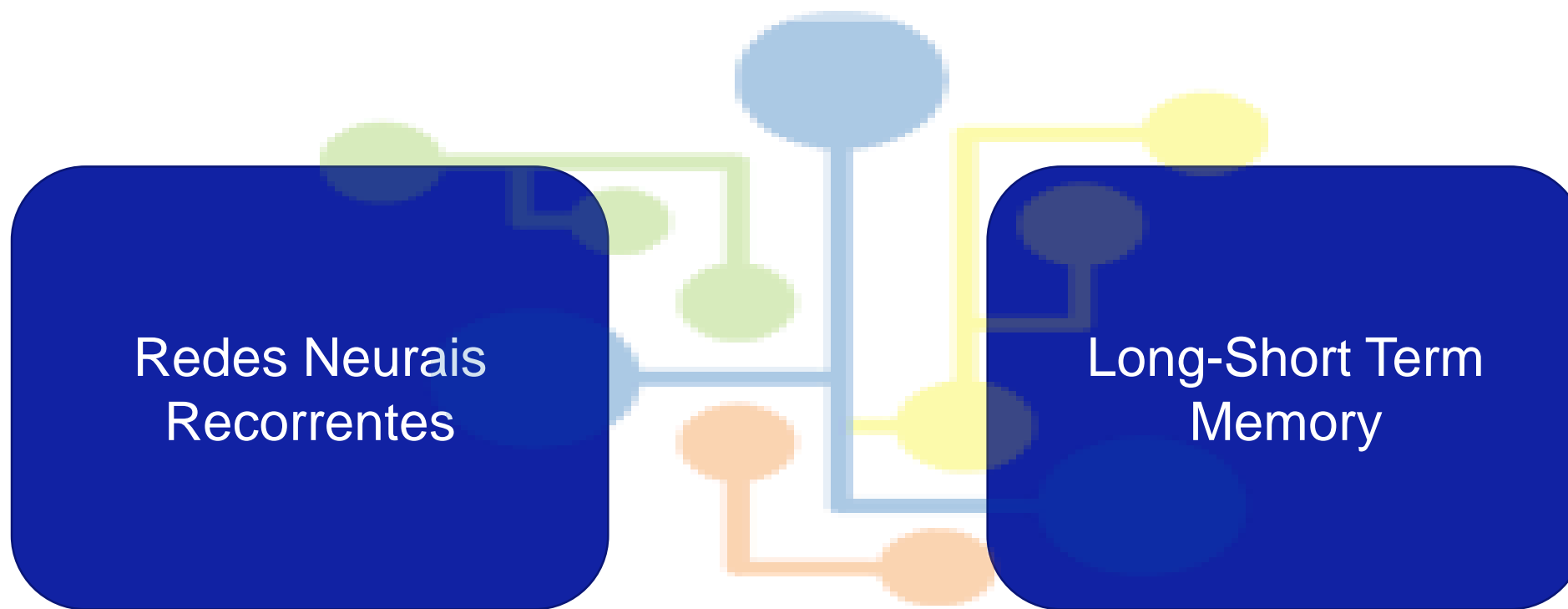
Deep Learning - Long-Short Term Memory



LSTMs são variantes das Redes Neurais Recorrentes.

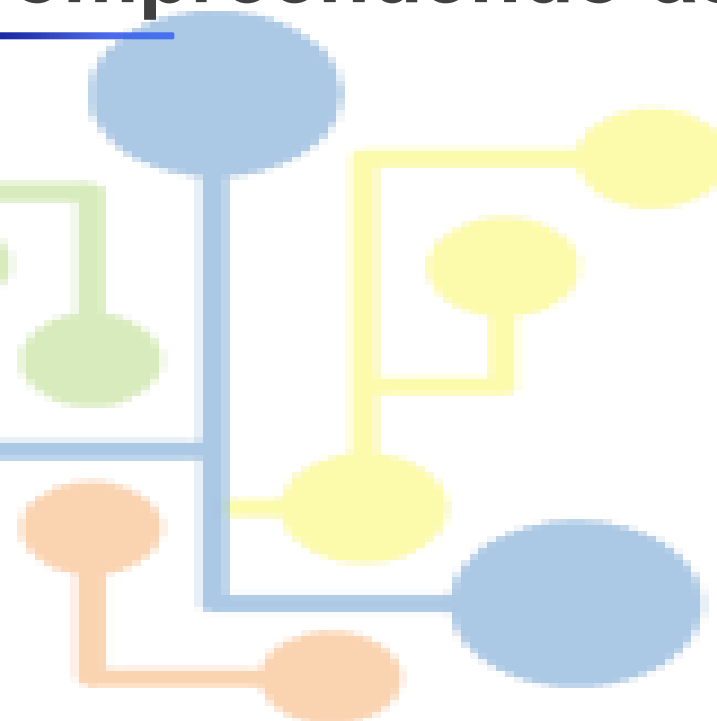


Deep Learning - Long-Short Term Memory



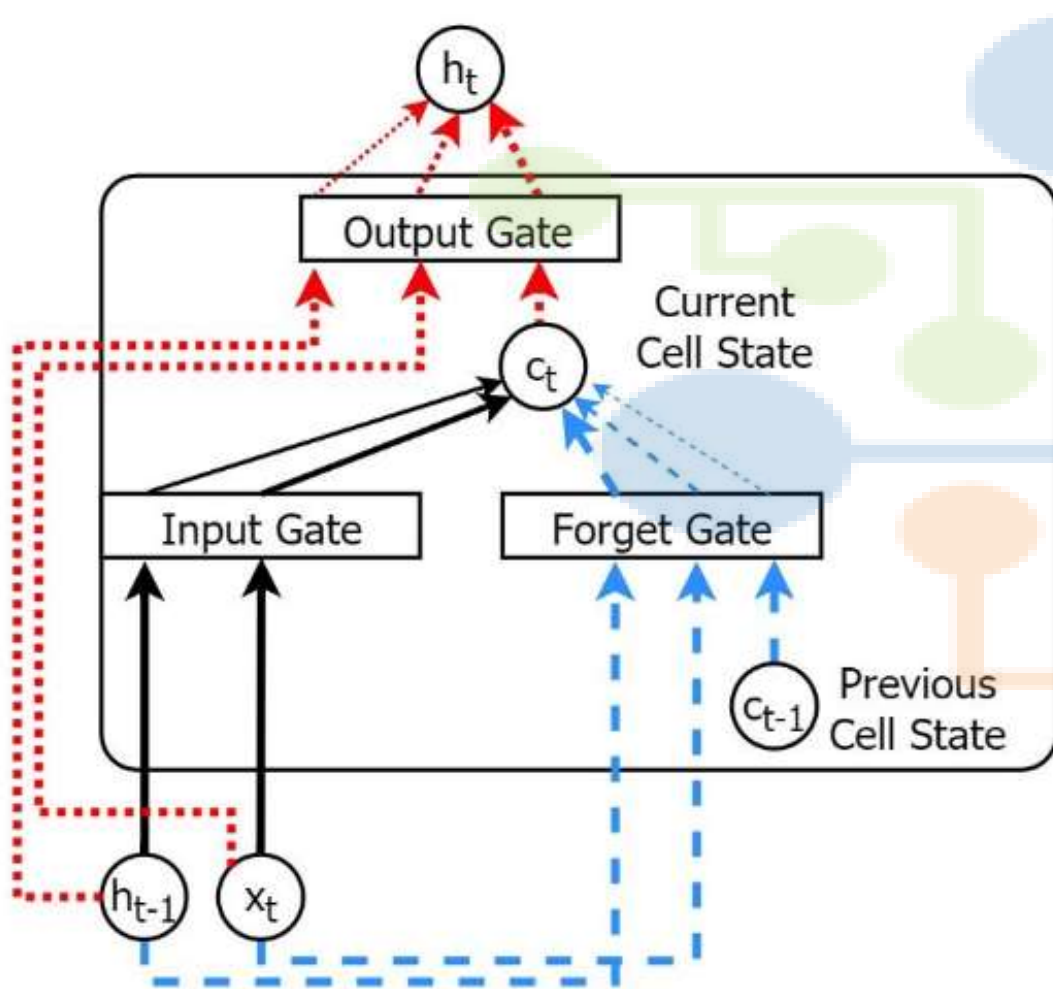


Compreendendo as LSTMs





Compreendendo as LSTMs



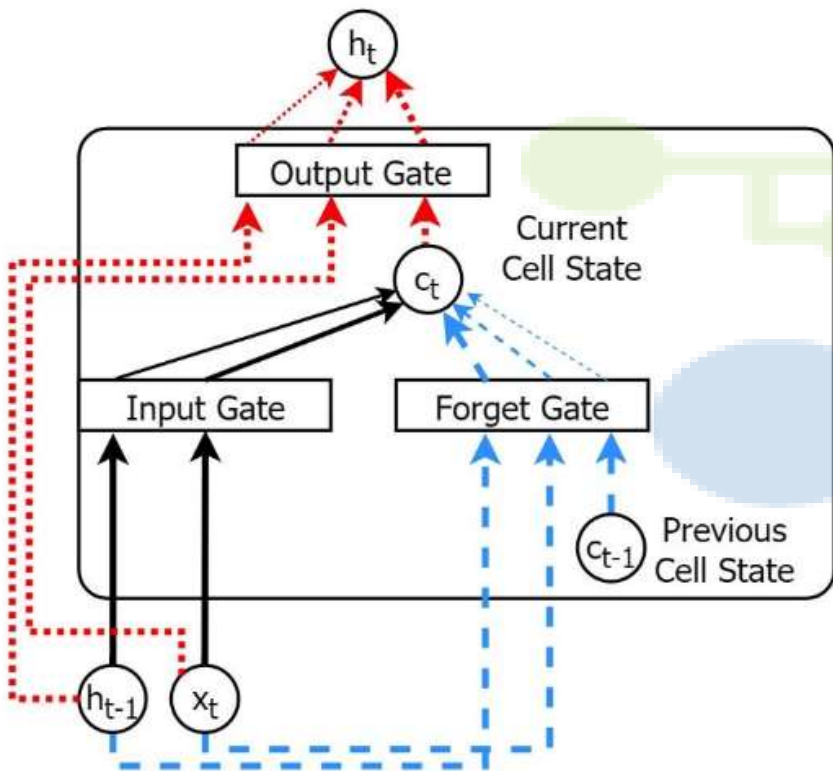
LSTMs são variantes das Redes Neurais Recorrentes.

Compreendendo as LSTMs



Data Science
Academy

Data Science Academy felipe.oliveiras2000@gmail.com 5f8a0b3ee32fc37d576ba60d



As LSTMs possuem 5 componentes:

- **Cell State** – célula interna de estado (a memória) da LSTM.
- **Hidden State** – estado externo oculto usado para calcular as previsões.
- **Input Gate** – determina quanto do input corrente é alimentado na cell state.
- **Forget Gate** – determina quanto da cell state anterior é enviado para a cell state corrente.
- **Output Gate** – determina quanto da cell state é output da hidden state.

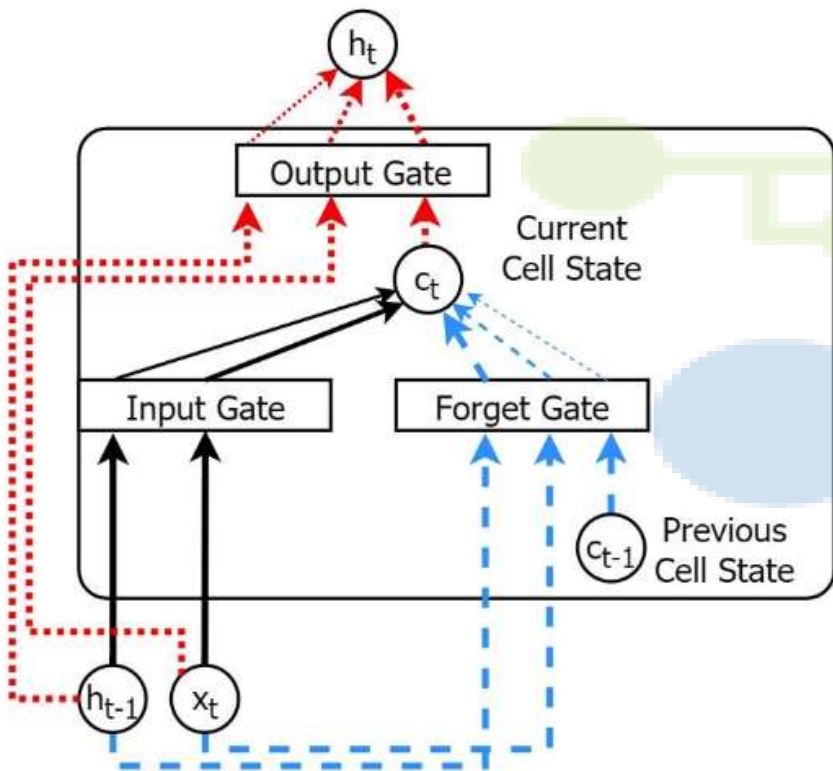
Compreendendo as LSTMs



Data Science
Academy

Data Science Academy felipe.oliveiras2000@gmail.com 5f8a0b3ee32fc37d576ba60d

As LSTMs possuem 3 Portões (Gates):



- **Input Gate** – gate com valores entre 0 (o input corrente não é gravado na cell state) e 1 (o input corrente é totalmente gravado na cell state). Ativação Sigmóide é usada para garantir que os valores fiquem entre 0 e 1.
- **Forget Gate** – gate sigmoidal com valores entre 0 (a cell state anterior é totalmente esquecida para o cálculo da cell state corrente) e 1 (a cell state anterior é totalmente usada para calcular a cell state corrente).
- **Output Gate** – gate sigmoidal que gera a saída com valores entre 0 (a cell state corrente é totalmente descartada para o cálculo do estado final) e 1 (a cell state corrente é totalmente usada no cálculo da hidden state).



Compreendendo as LSTMs

John gave Mary a puppy.

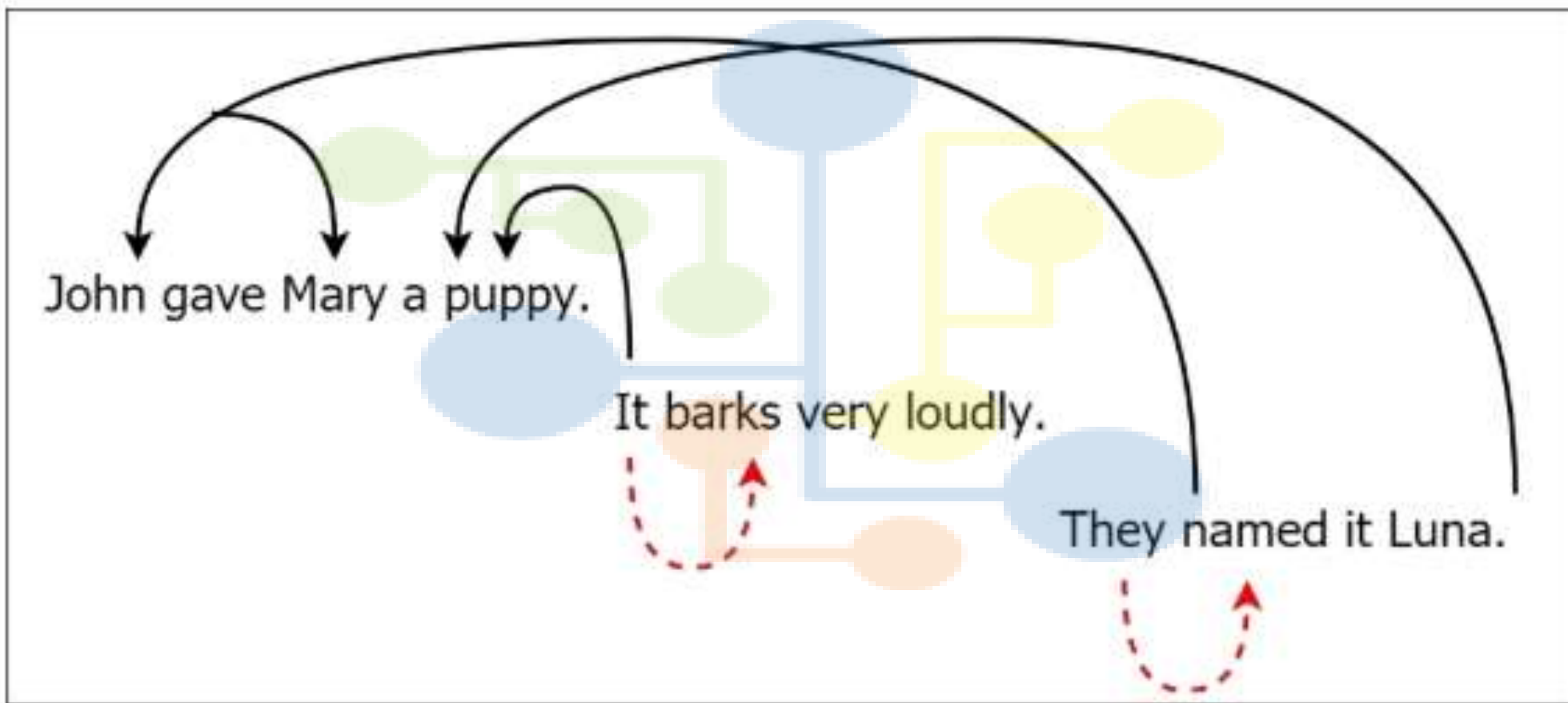
John gave Mary a puppy. *It barks very loudly. They named it Luna.*

John gave **Mary** a puppy. *It barks very loudly. **They** named it Luna.*

John gave Mary a **puppy**. ***It barks** very loudly. They named **it** **Luna**.*



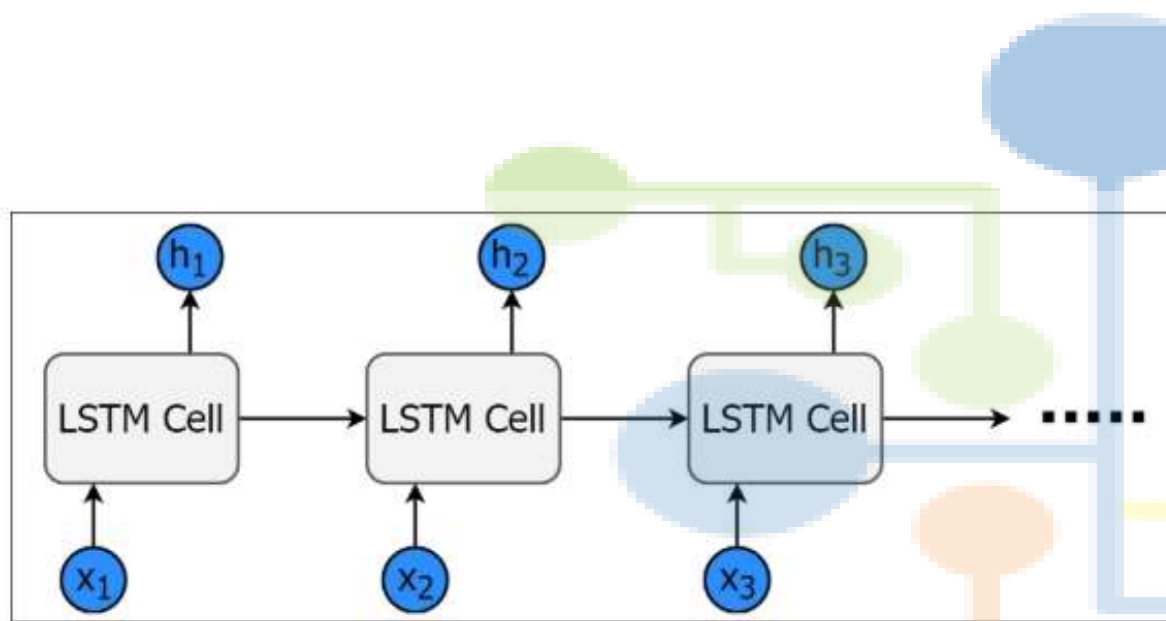
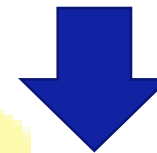
Compreendendo as LSTMs





Compreendendo as LSTMs

Isso aqui é o que acontece dentro da LSTM cell:



$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i)$$

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f)$$

$$\tilde{c}_t = \tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c)$$

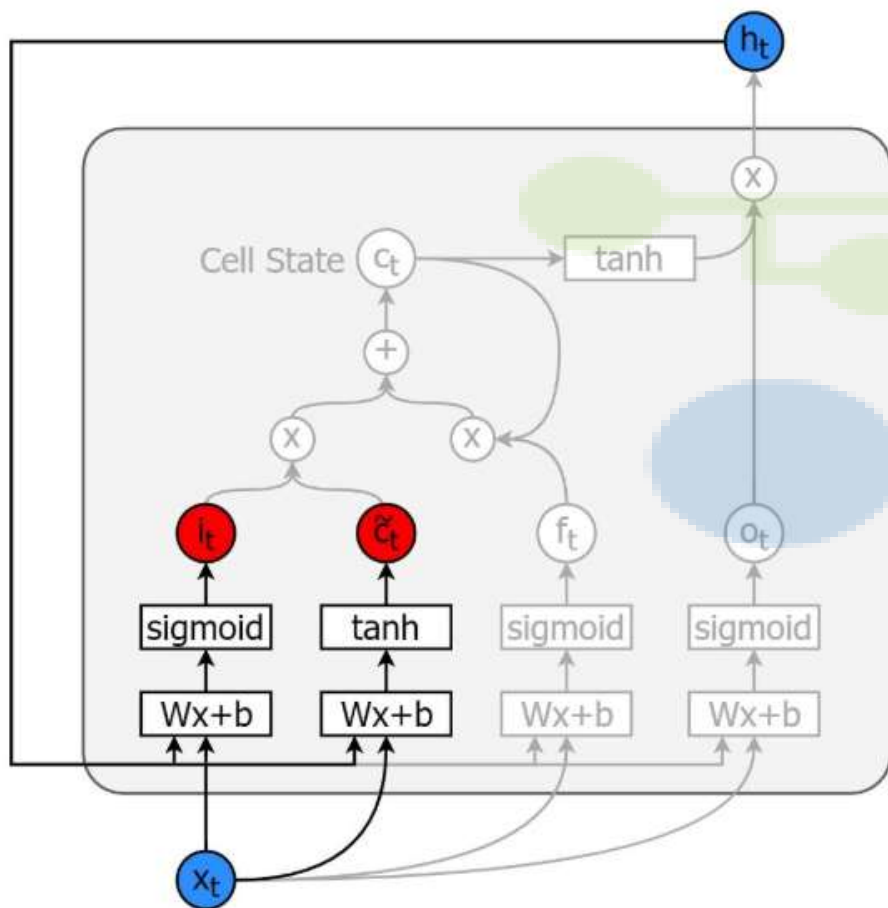
$$c_t = f_t c_{t-1} + i_t \tilde{c}_t$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o)$$

$$h_t = o_t \tanh(c_t)$$



Compreendendo as LSTMs



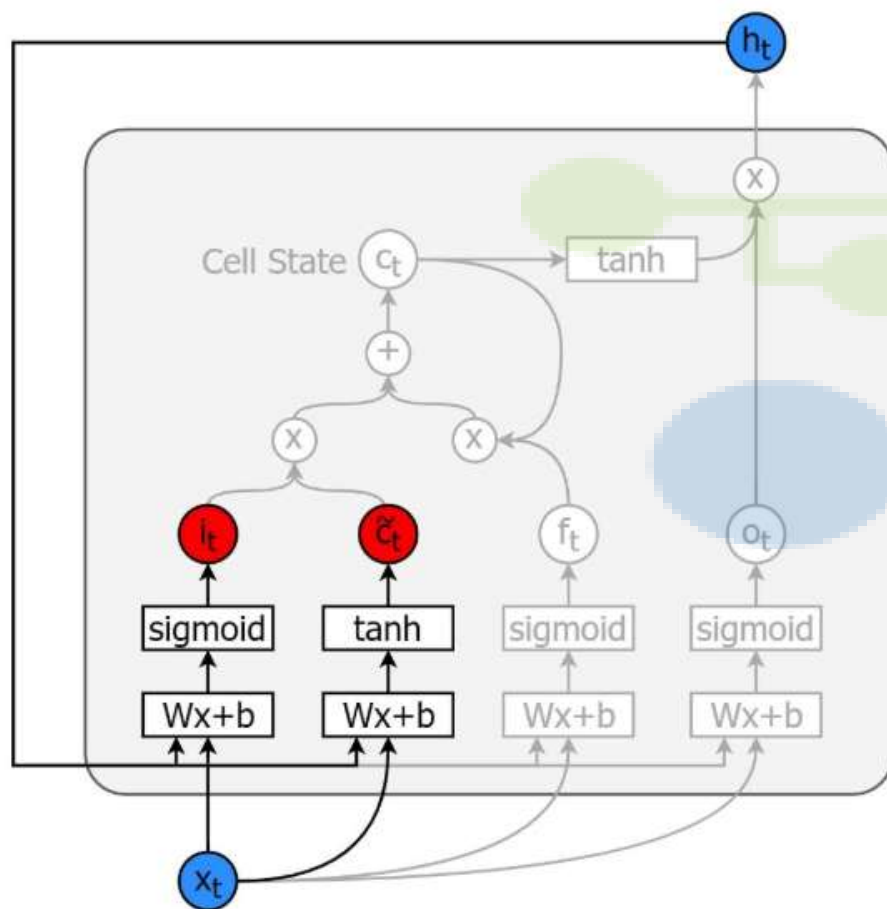
Primeiro o Input Gate (i) recebe o current input (x) e o hidden state anterior (h_{t-1}) e calcula da seguinte forma:

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i)$$

Input Gate – gate com valores entre 0 (o input corrente não é gravado na cell state) e 1 (o input corrente é totalmente gravado na cell state). Ativação Sigmóide é usada para garantir que os valores fiquem entre 0 e 1.



Compreendendo as LSTMs



Em seguida, é calculado um valor chamado **candidate value**, que é adicionado para calcular o valor final da cell state.

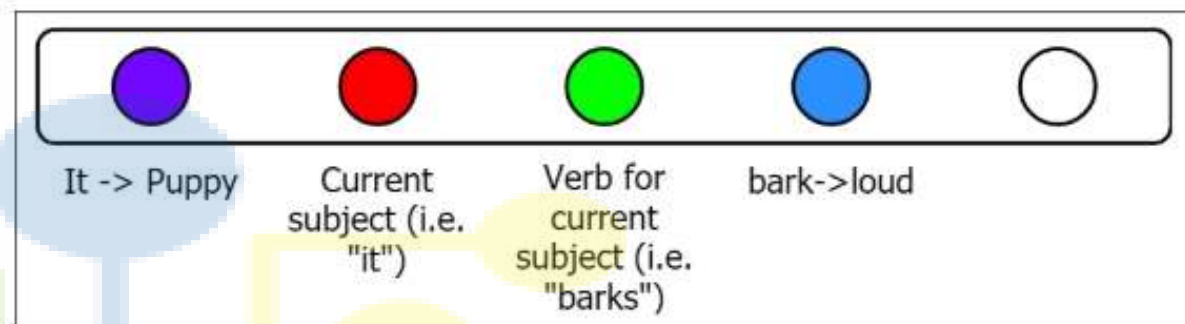
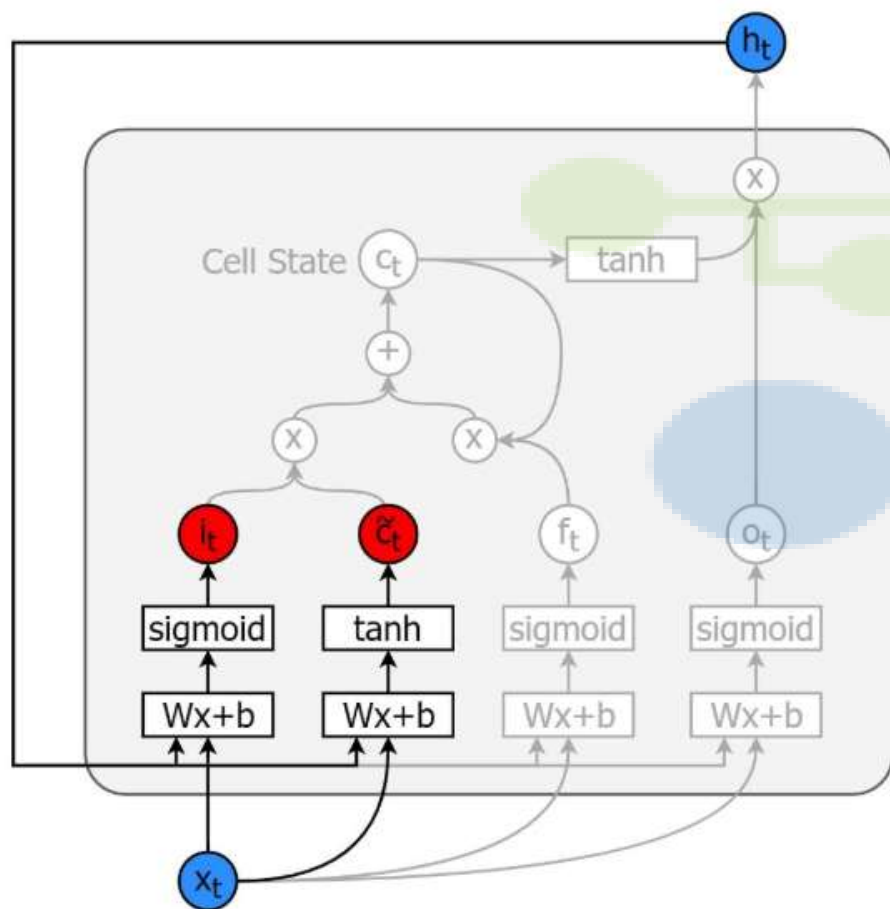
$$\tilde{c}_t = \tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c)$$

Compreendendo as LSTMs



Data Science
Academy

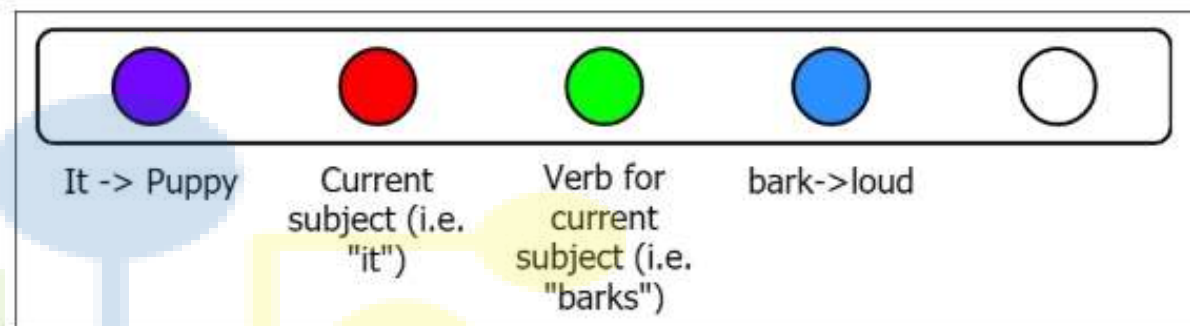
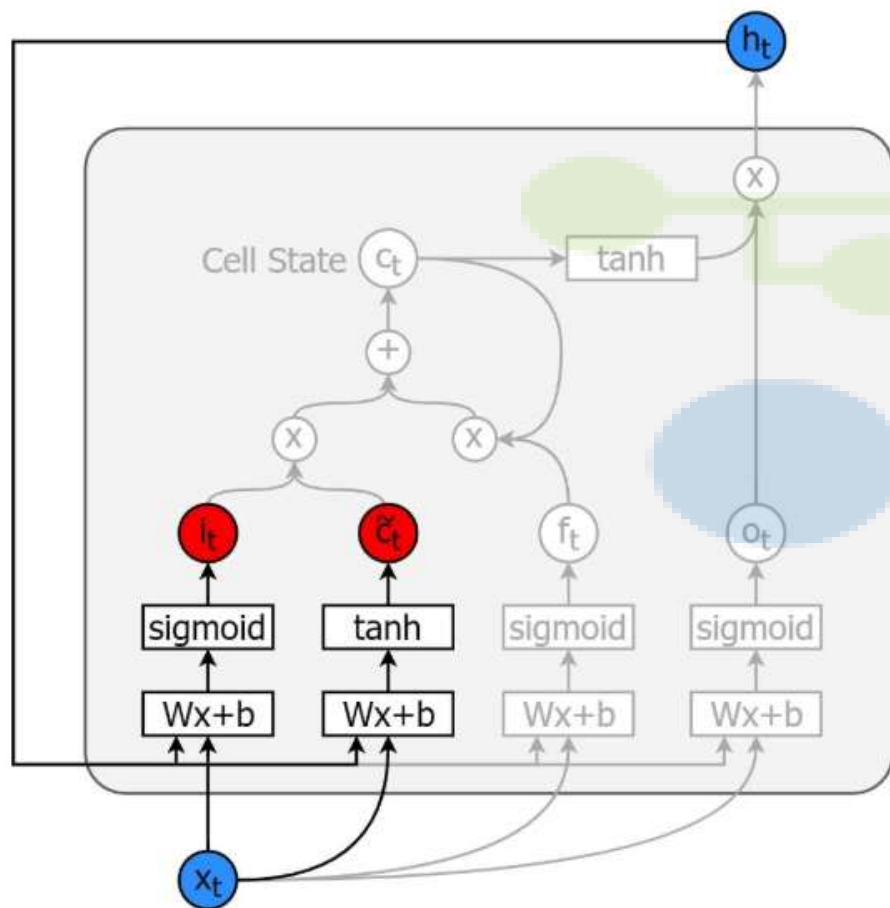
Data Science Academy felipe.oliveiras2000@gmail.com 5f8a0b3ee32fc37d576ba60d



Vamos supor que nossa LSTM tem 5 neurônios para armazenar o estado. A primeira palavra que será a saída do nosso modelo LSTM, será a palavra it (considerando nosso exemplo), logo nosso modelo precisa aprender que it é o pronome usado para representar a palavra puppy. Além disso, nosso modelo precisa guardar na memória que it se refere a puppy. Nosso modelo deve aprender ainda que o s deve ser usado ao final do verbo na terceira pessoa do singular em inglês. Calculamos então a saída da Forget Gate.



Compreendendo as LSTMs

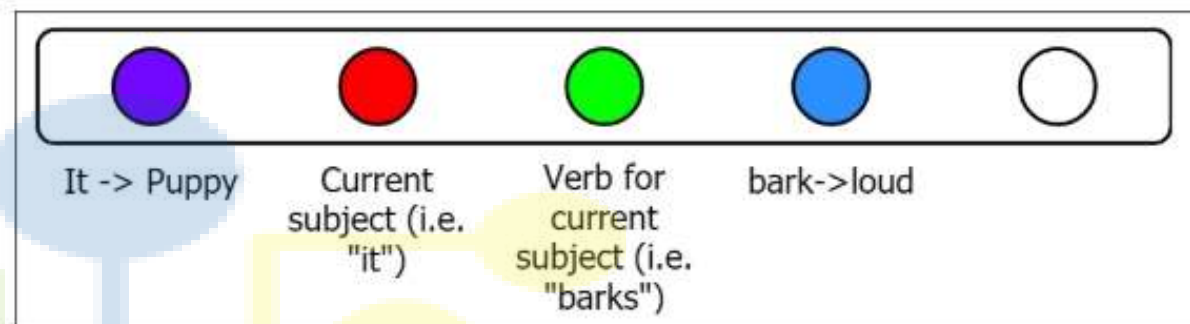
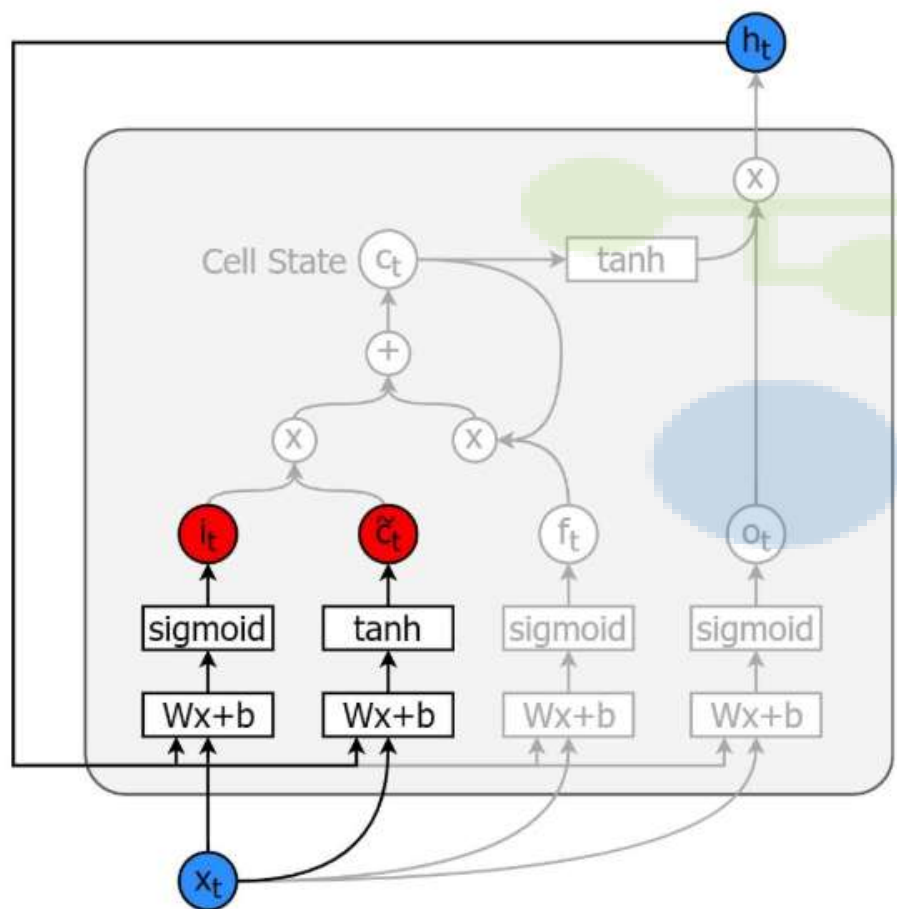


$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f)$$

Forget Gate – gate sigmoidal com valores entre 0 (a cell state anterior é totalmente esquecida para o cálculo da cell state corrente) e 1 (a cell state anterior é totalmente usada para calcular a cell state corrente).



Compreendendo as LSTMs

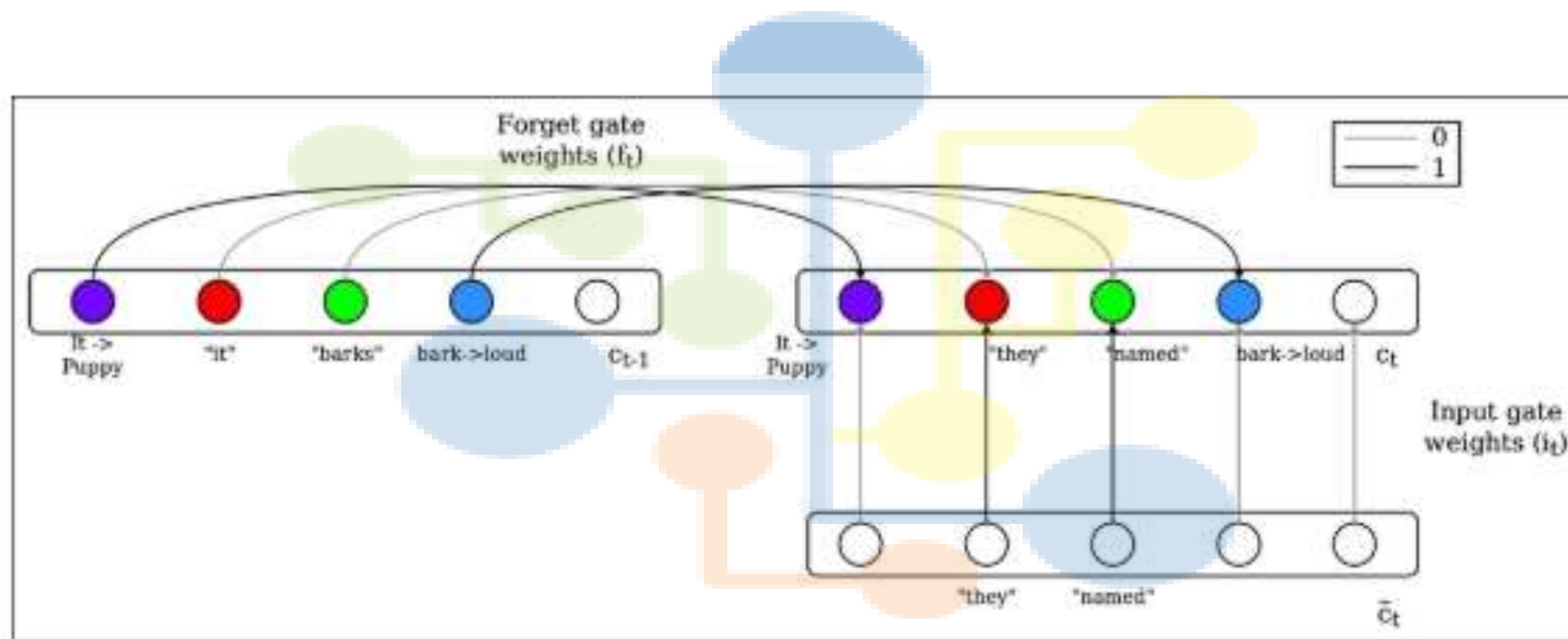


$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f)$$

Um valor igual a 0 indica que nenhuma informação de $ct-1$ será passada para calcular ct e um valor igual a 1 significa que toda informação de $ct-1$ será propagada para calcular ct .

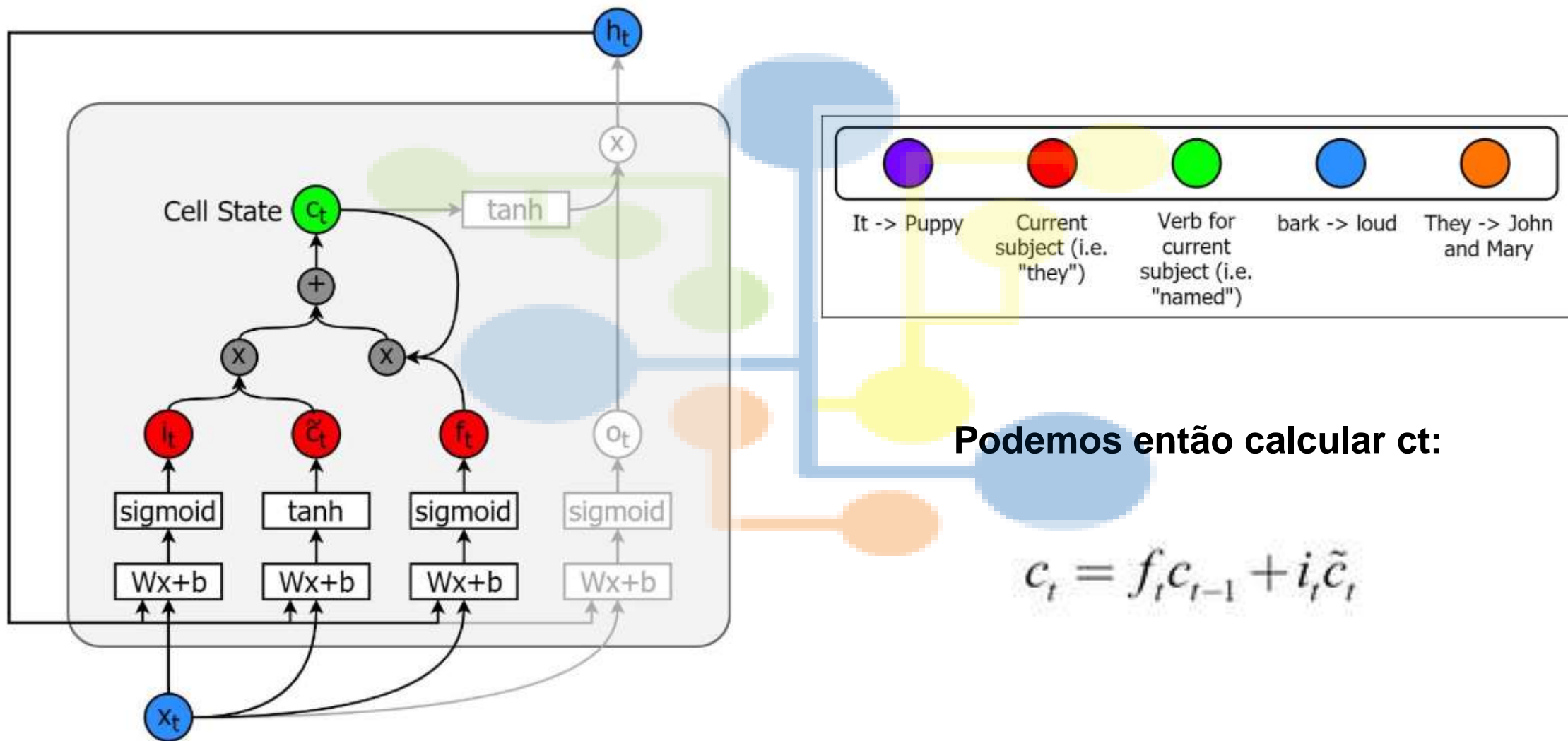


Compreendendo as LSTMs



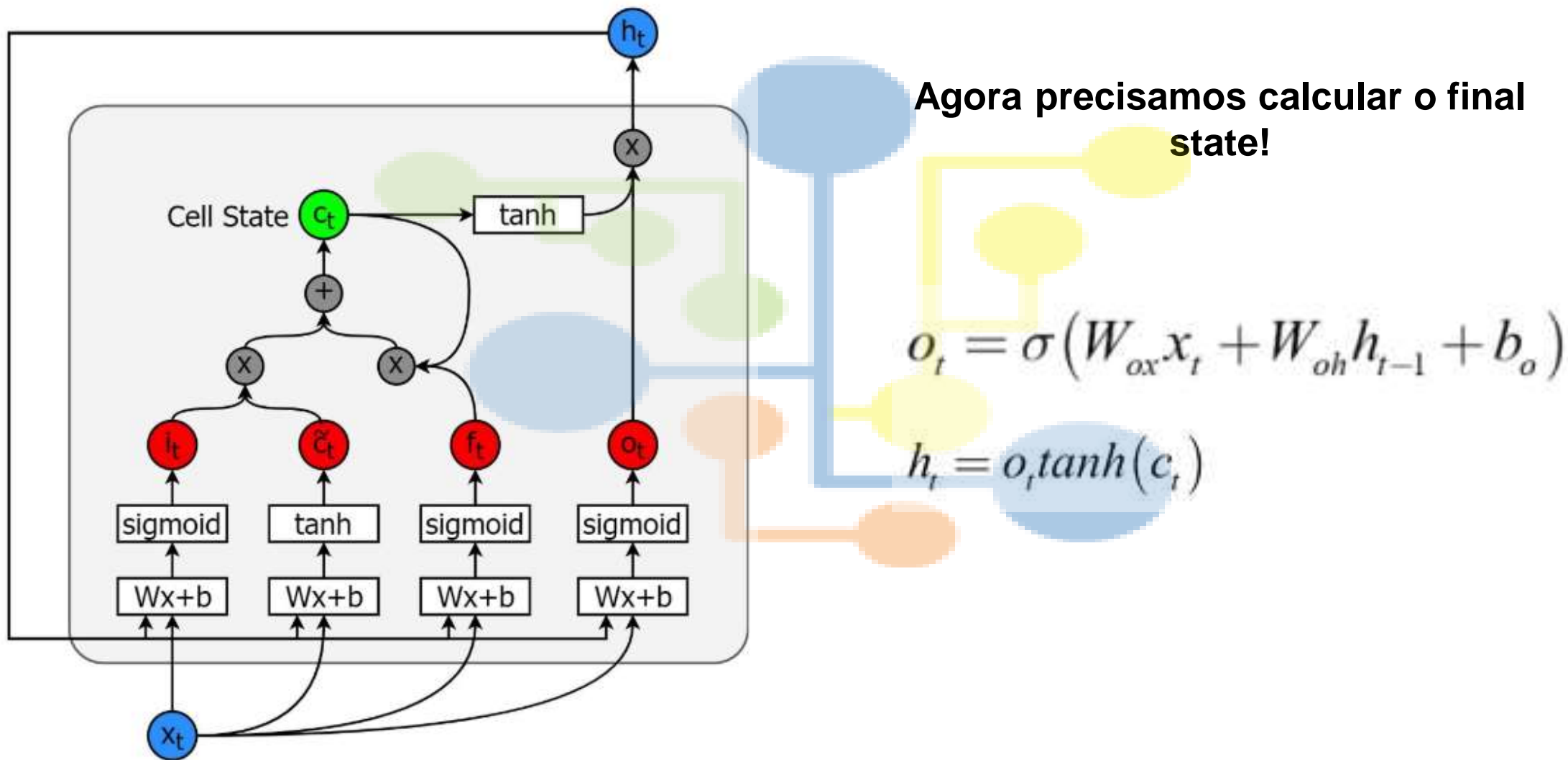


Compreendendo as LSTMs





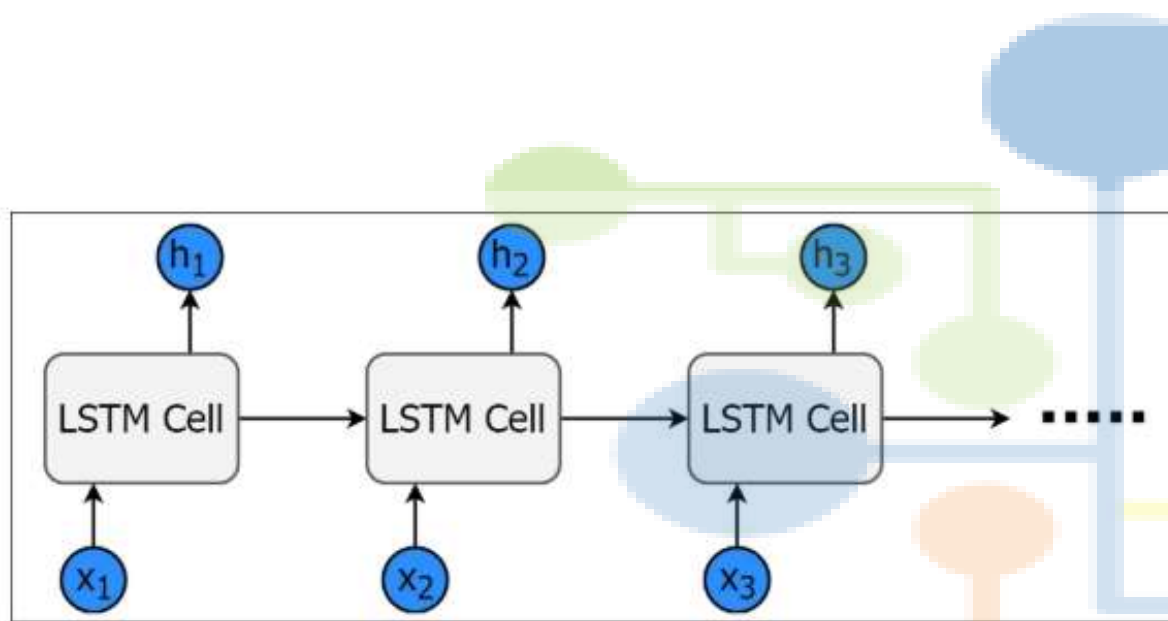
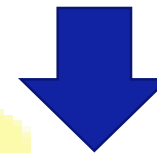
Compreendendo as LSTMs





Compreendendo as LSTMs

Isso aqui é o que acontece dentro da LSTM cell:



$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i)$$

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f)$$

$$\tilde{c}_t = \tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c)$$

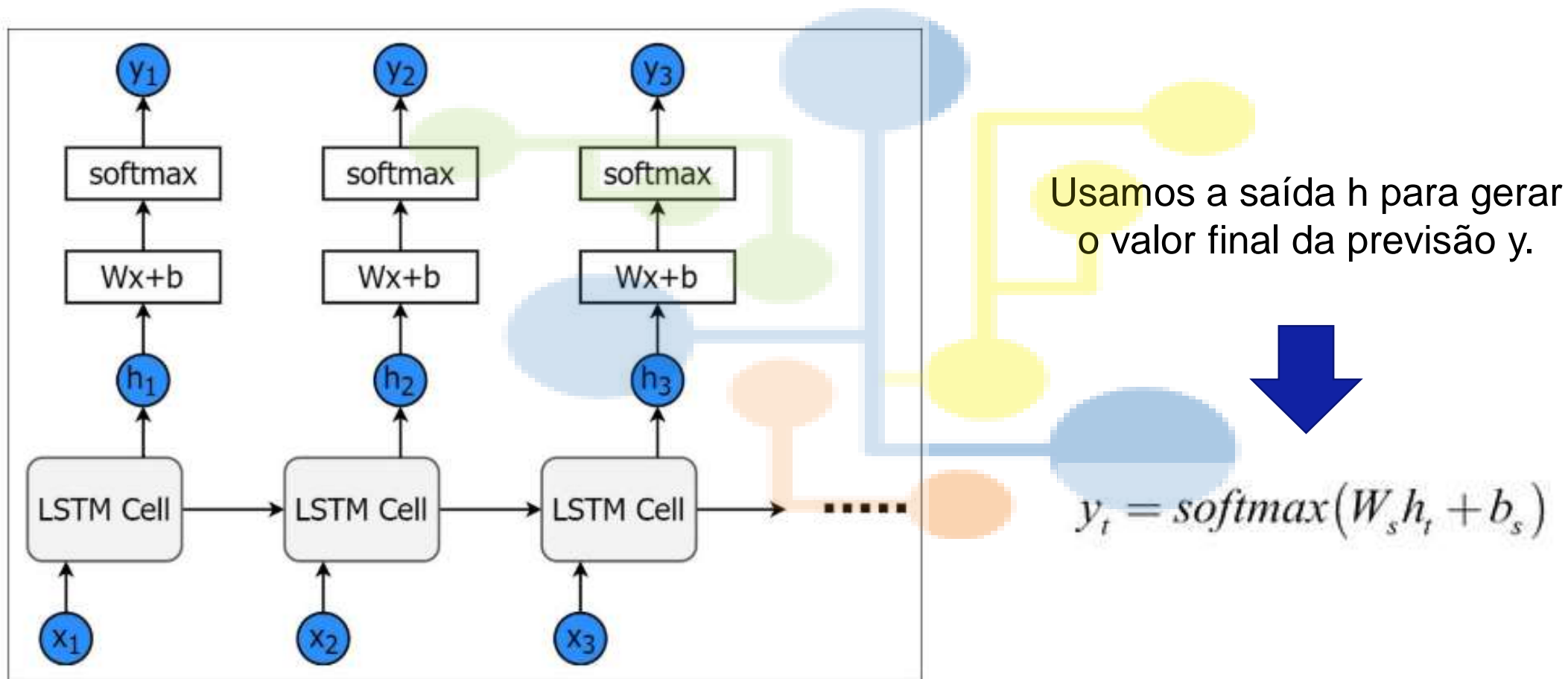
$$c_t = f_t c_{t-1} + i_t \tilde{c}_t$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o)$$

$$h_t = o_t \tanh(c_t)$$

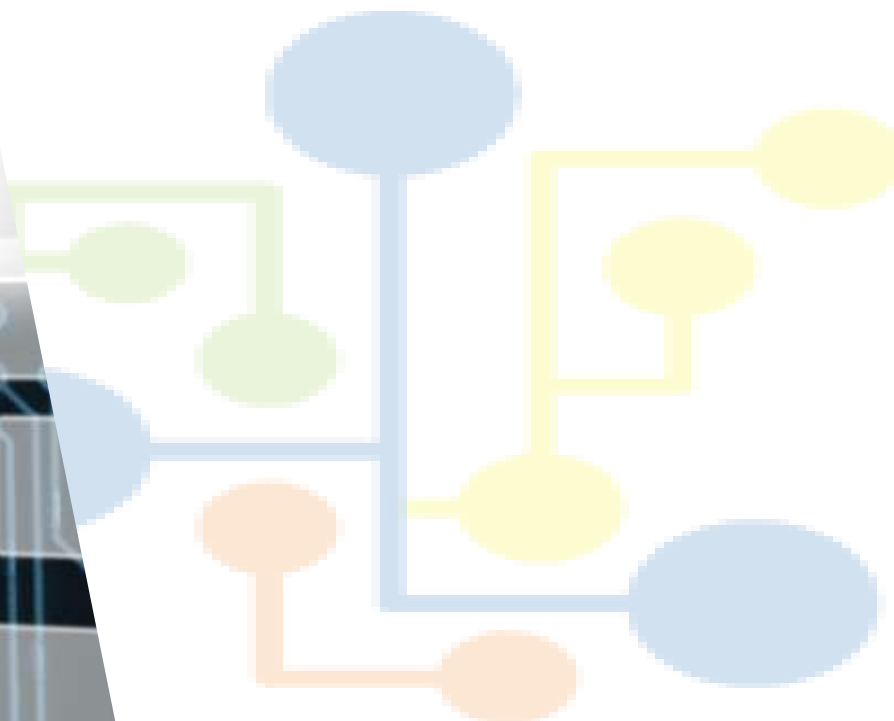


Compreendendo as LSTMs





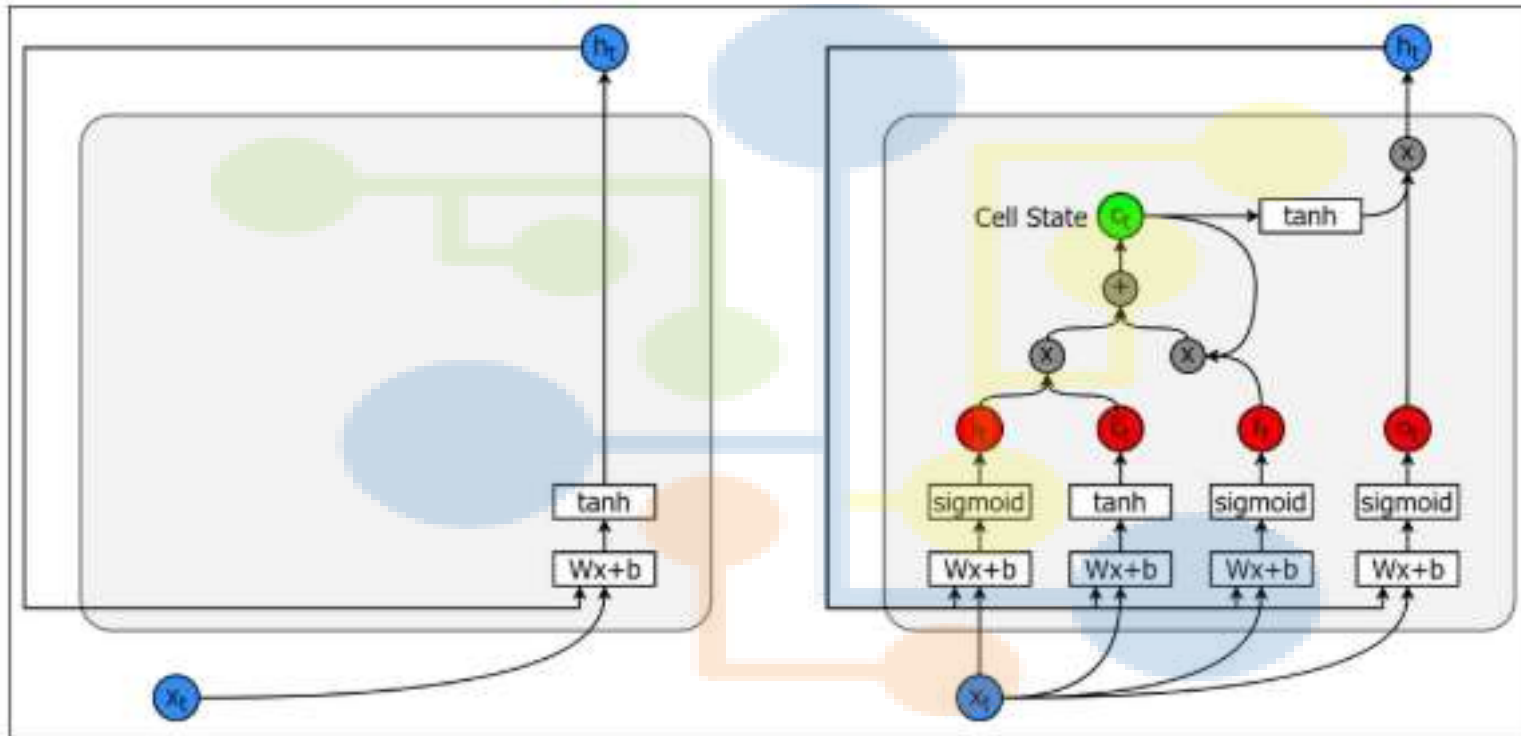
Como as LSTMs Diferem das RNNs





Como as LSTMs Diferem das RNNs

A RNN possui apenas um hidden state h_t .



RNN

LSTM

A LSTM possui 2 estados diferentes, cell state (c_t) e final hidden state (h_t).



Extensões de Melhoria de Performance Para as LSTMs





Extensões de Melhoria de Performance Para as LSTMs

Extensões que melhoram a performance de uma LSTM e tornam as saídas mais realísticas:

- Greedy Sampling
- Beam Search
- Usar Word Vectors ao invés de representações one-hot encoding

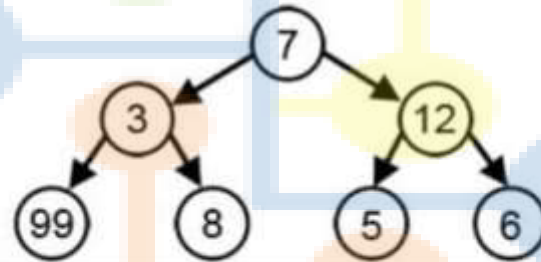
Greedy Sampling



Data Science
Academy

Data Science Academy felipe.oliveiras2000@gmail.com 5f8a0b3ee32fc37d576ba60d

Greedy algorithm



Greedy Sampling



Data Science
Academy

Data Science Academy felipe.oliveiras2000@gmail.com 5f8a0b3ee32fc37d576ba60d

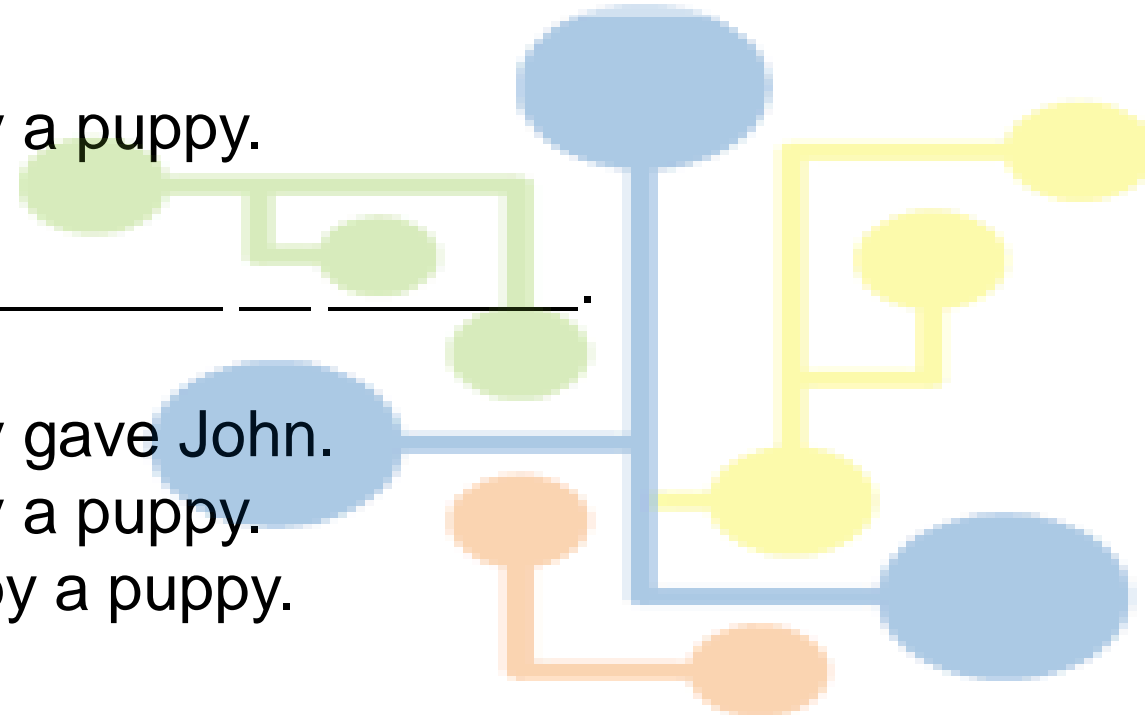
John gave Mary a puppy.

John _____.

John gave Mary gave John.

John gave Mary a puppy.

John gave puppy a puppy.



Greedy Sampling



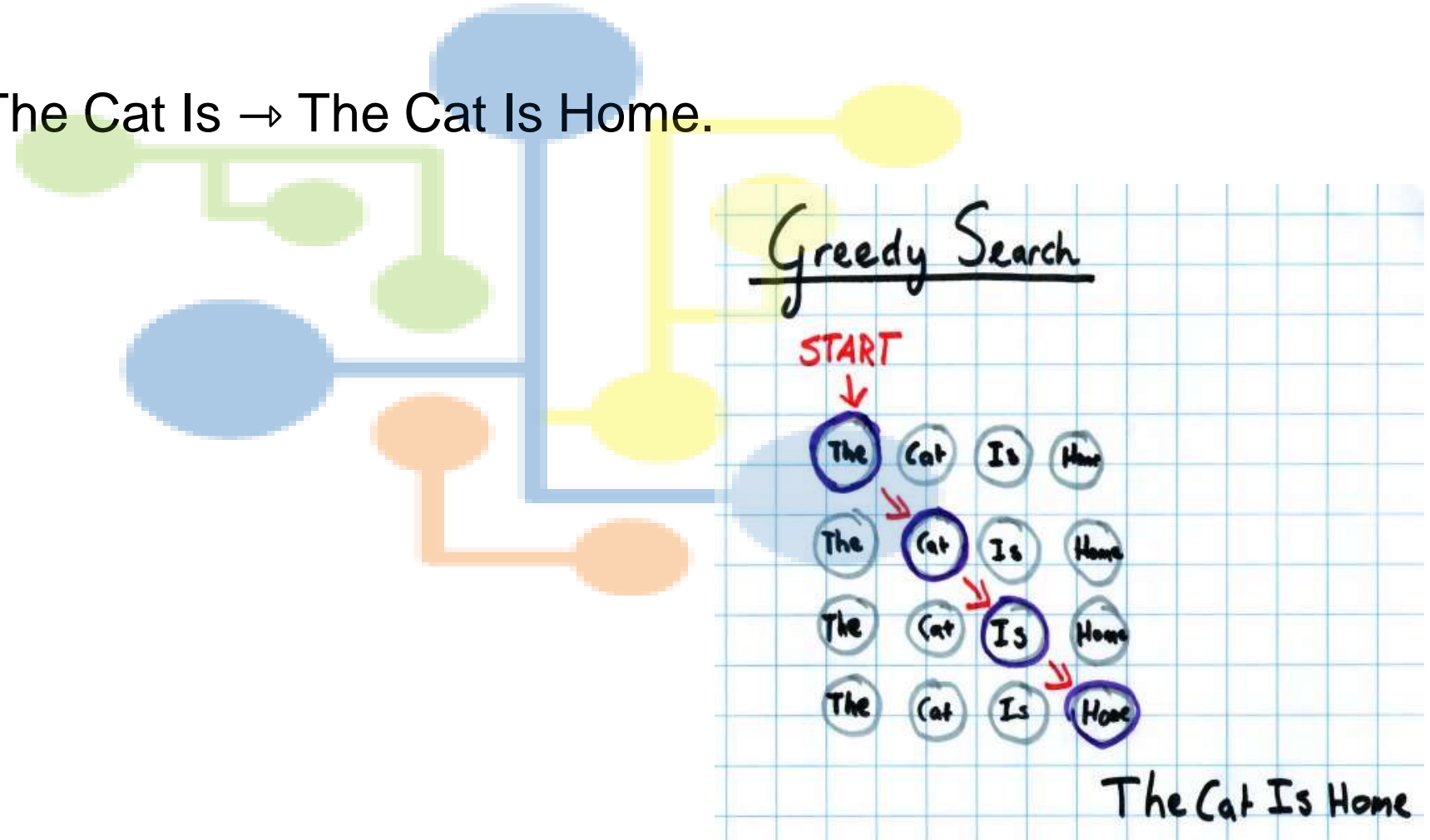
Data Science
Academy

Data Science Academy felipe.oliveiras2000@gmail.com 5f8a0b3ee32fc37d576ba60d

The cat is home.

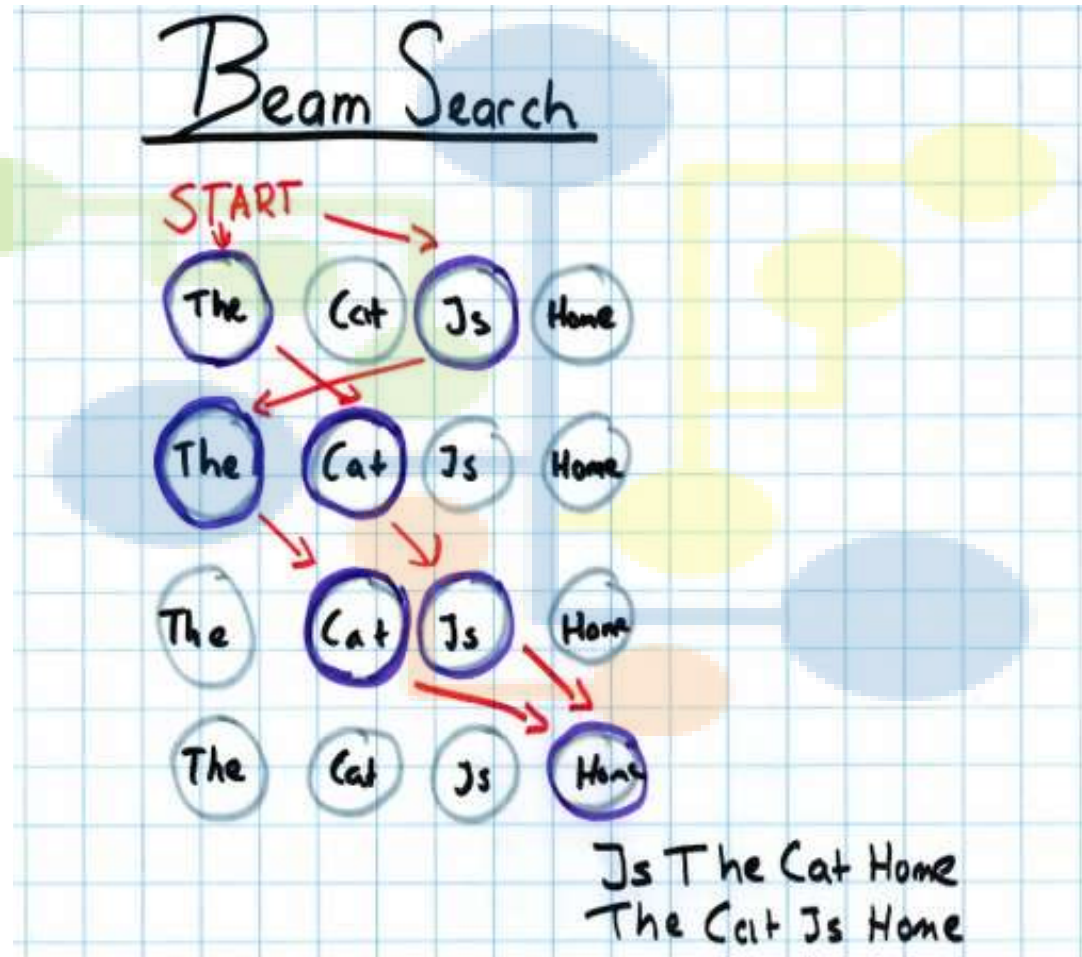
The → The Cat → The Cat Is → The Cat Is Home.

The cat The cat.



Beam Search

Data Science
Academy



O Greedy Sampling faz uma amostragem entre as palavras com maiores probabilidades de serem a próxima palavra em uma frase e seleciona uma de forma aleatória.

O objetivo do Beam Search é aumentar a qualidade das previsões geradas pelos modelos LSTM. As melhores previsões são encontradas resolvendo um problema de busca.

A principal diferença entre os 2 métodos, é que o Beam Search utiliza *joint probability (probabilidade conjunta)*, enquanto o Greedy Sampling utiliza apenas a maior probabilidade.

Beam Search



Data Science
Academy

Data Science Academy felipe.oliveiras2000@gmail.com 5f8a0b3ee32fc37d576ba60d

$$y_t, y_{t+1}, \dots, y_{t+b}$$

$$p(y_t, y_{t+1}, \dots, y_{t+b} | x_t)$$

$$p(y_t | x_t)$$

Beam Search gera **b outputs** ao invés de um único output y_t . Aqui, b outputs é chamado de Beam.

Com Beam Search nós obtemos o Beam de mais alta probabilidade conjunta.

Com Greedy Sampling obtemos apenas as palavras de probabilidade mais alta e dessas palavras pegamos uma amostra para ser a próxima palavra no texto.



Data Science
Academy

Data Science Academy felipe.oliveiras2000@gmail.com 5f8a0b3ee32fc37d576ba60d

Obrigado