



Formação Inteligência Artificial

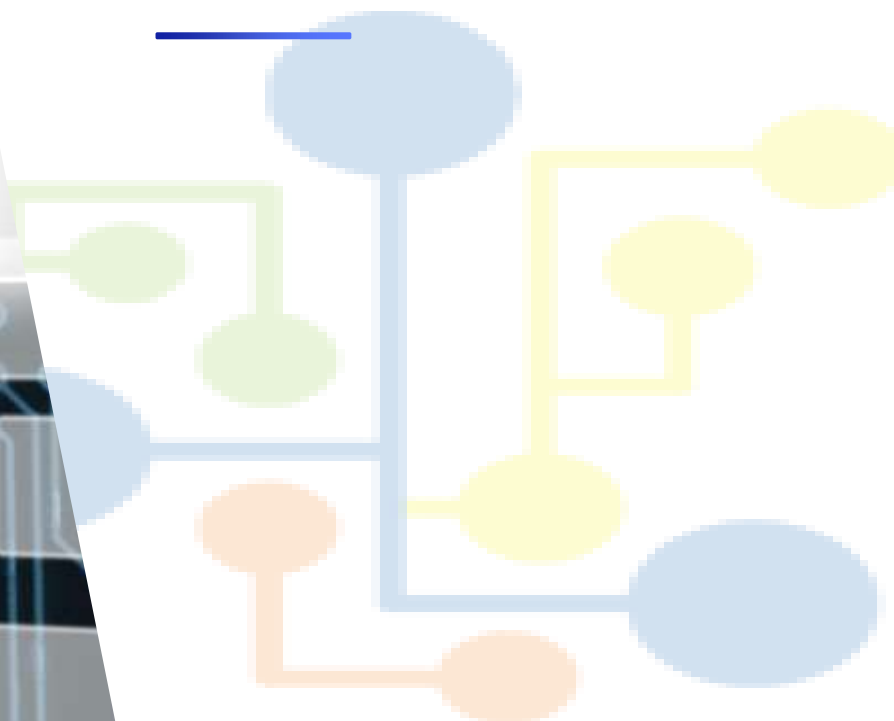


Processamento de Linguagem Natural



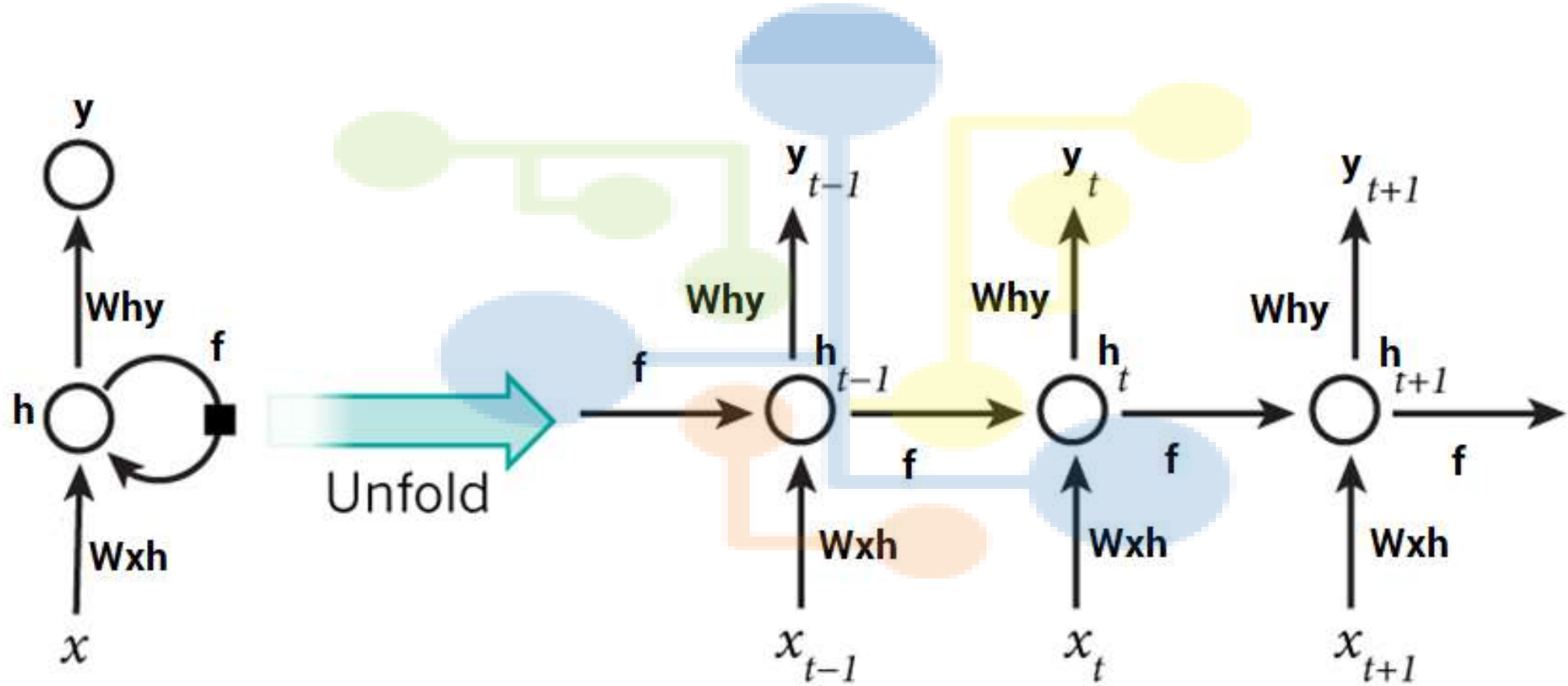


Deep Learning - Redes Neurais Recorrentes





Deep Learning - Redes Neurais Recorrentes





Deep Learning - Redes Neurais Recorrentes

A decorative background diagram consisting of a central vertical blue line. From this line, several horizontal and vertical lines branch out, connecting to various colored circles (nodes) in light blue, green, yellow, and orange. Two large dark blue rounded rectangles are positioned on either side of the central structure, containing the text 'Redes Neurais Recorrentes' and 'Long-Short Term Memory' respectively.

Redes Neurais
Recorrentes

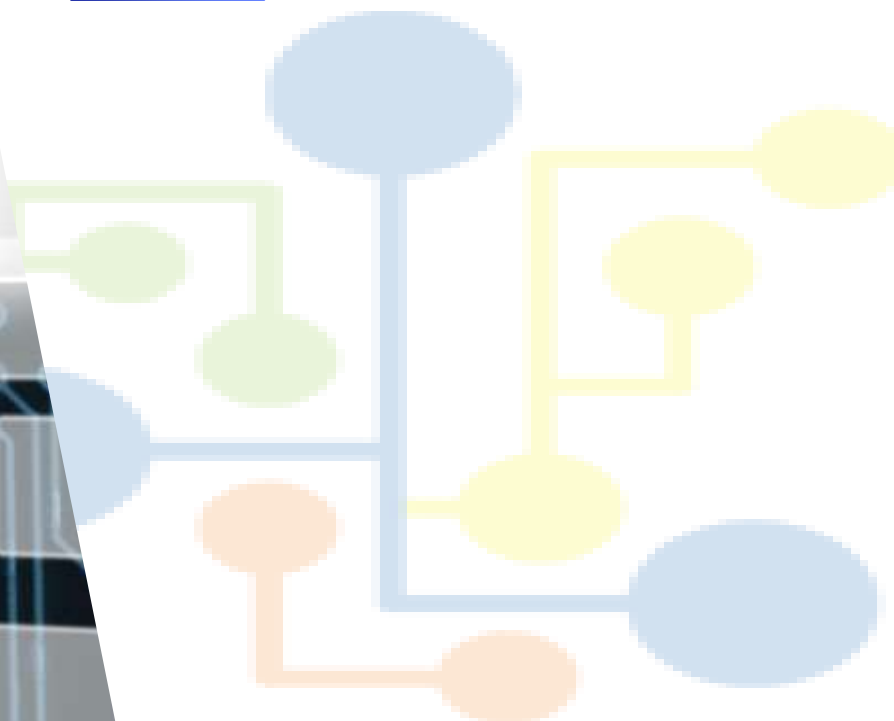
Long-Short Term
Memory



Data Science
Academy

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

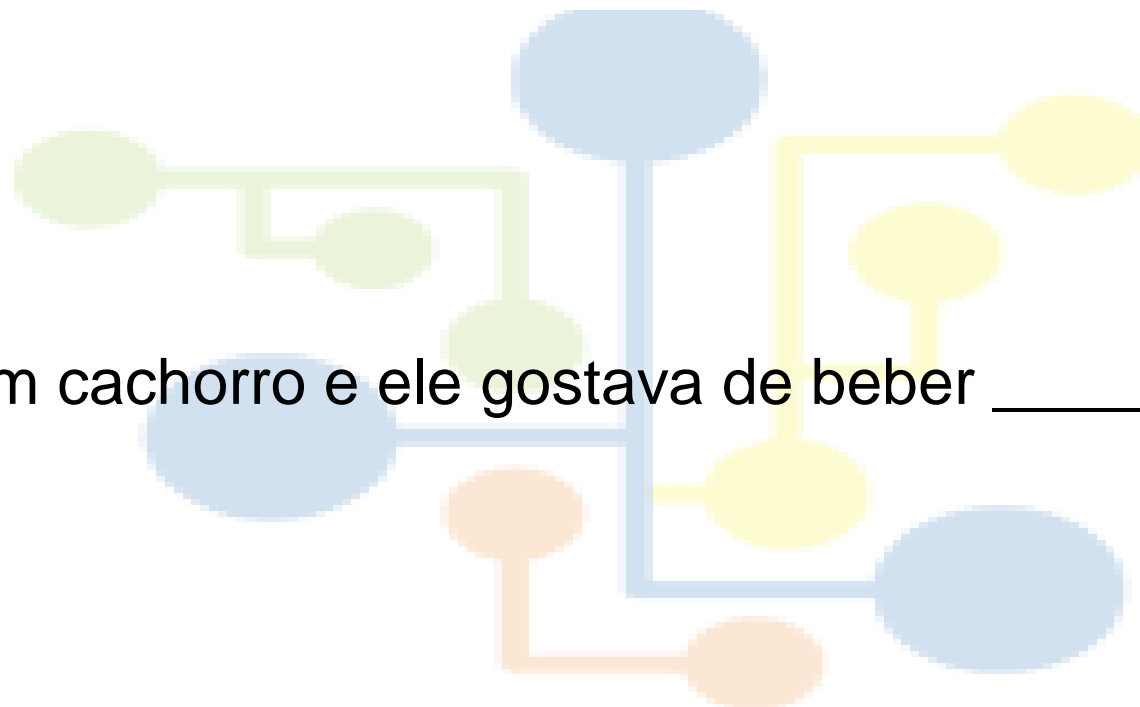
O Problema das Redes Feed-Forward Para Dados Sequenciais





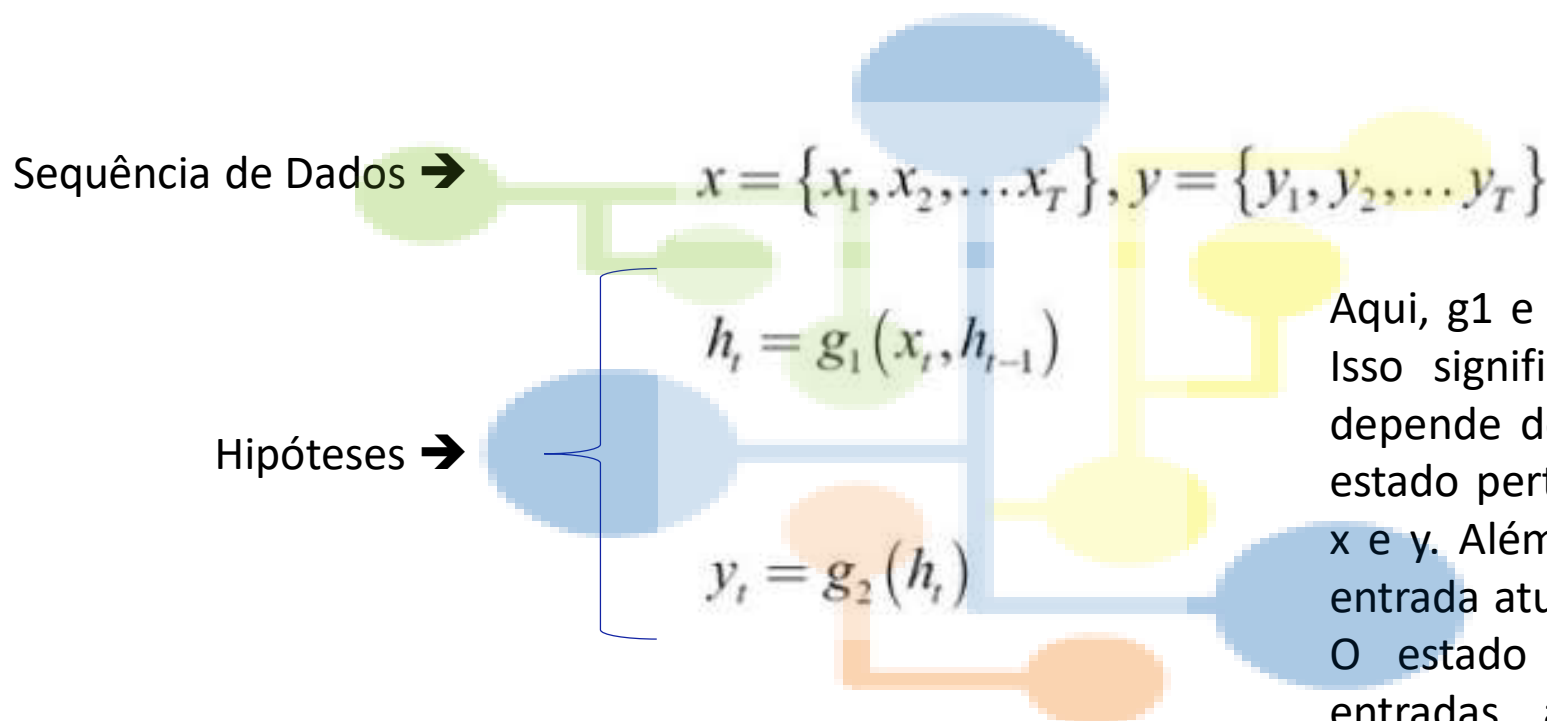
O Problema das Redes Feed-Forward Para Dados Sequenciais

Matias tinha um cachorro e ele gostava de beber _____.





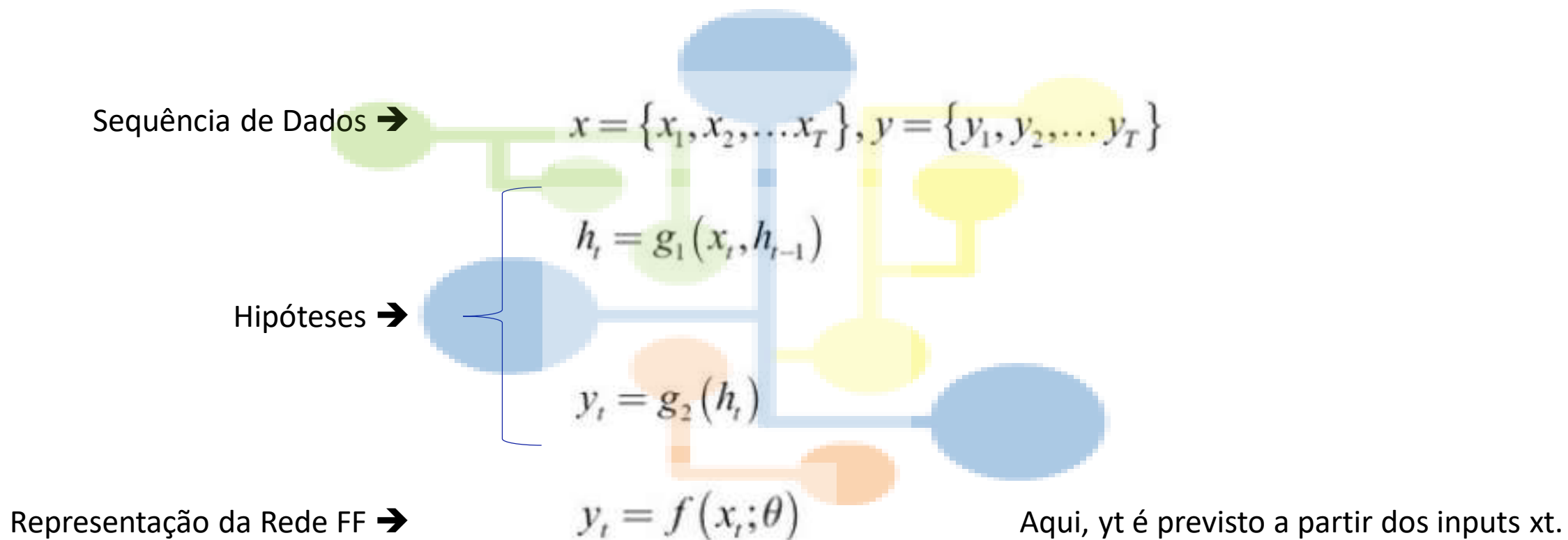
O Problema das Redes Feed-Forward Para Dados Sequenciais



Aqui, g_1 e g_2 são funções do algoritmo. Isso significa que a saída corrente y_t depende do estado atual h_t para algum estado pertencente ao modelo que gera x e y . Além disso, h_t é calculado com a entrada atual x_t e o estado anterior h_{t-1} . O estado codifica informações sobre entradas anteriores observadas pelo modelo.



O Problema das Redes Feed-Forward Para Dados Sequenciais

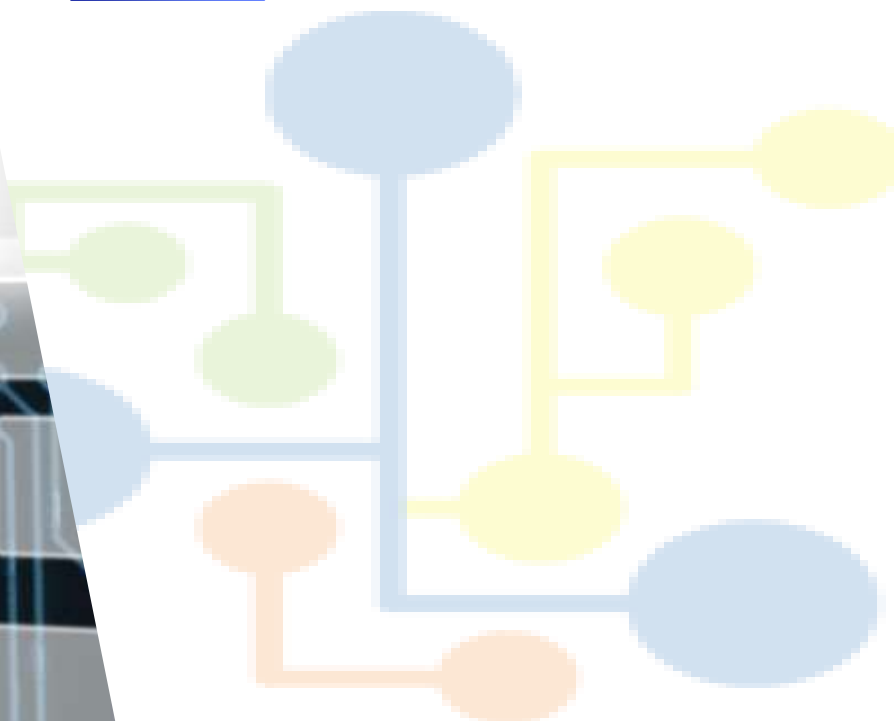




Data Science
Academy

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

Modelagem de Sequências de Dados com Redes Neurais Recorrentes





Modelagem de Sequências de Dados com Redes Neurais Recorrentes

Matias tinha um cachorro e ele gostava de beber _____.



Modelagem de Sequências de Dados com Redes Neurais Recorrentes

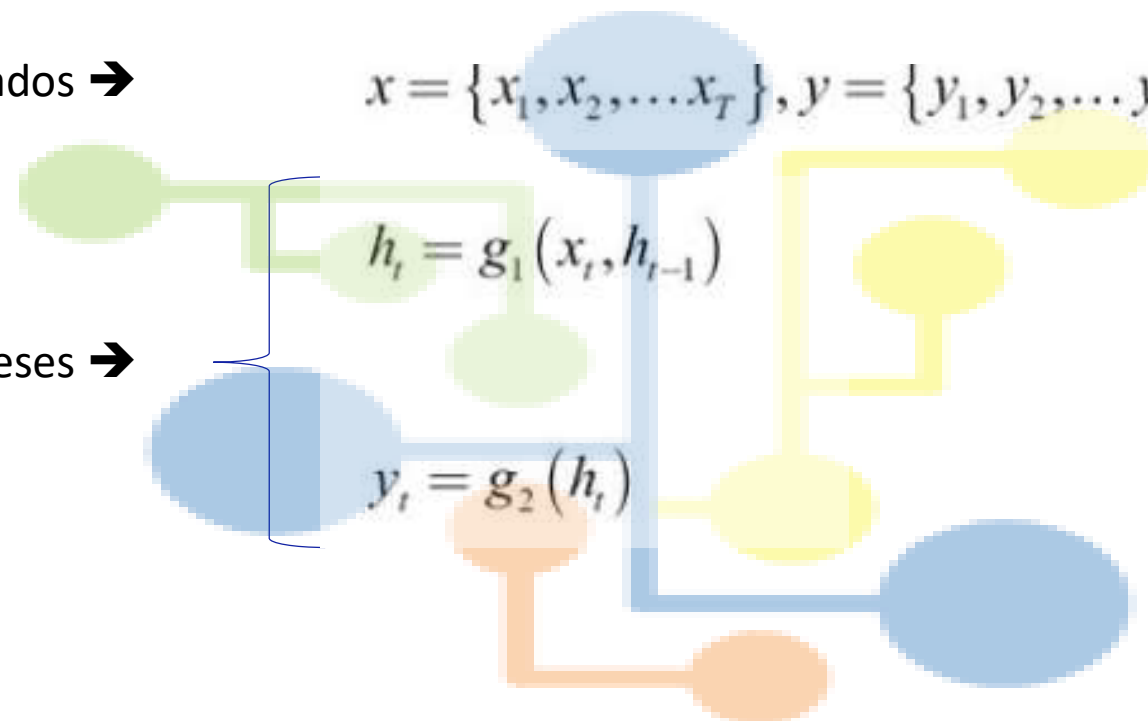
Sequência de Dados →

$$x = \{x_1, x_2, \dots, x_T\}, y = \{y_1, y_2, \dots, y_T\}$$

Hipóteses →

$$h_t = g_1(x_t, h_{t-1})$$

$$y_t = g_2(h_t)$$





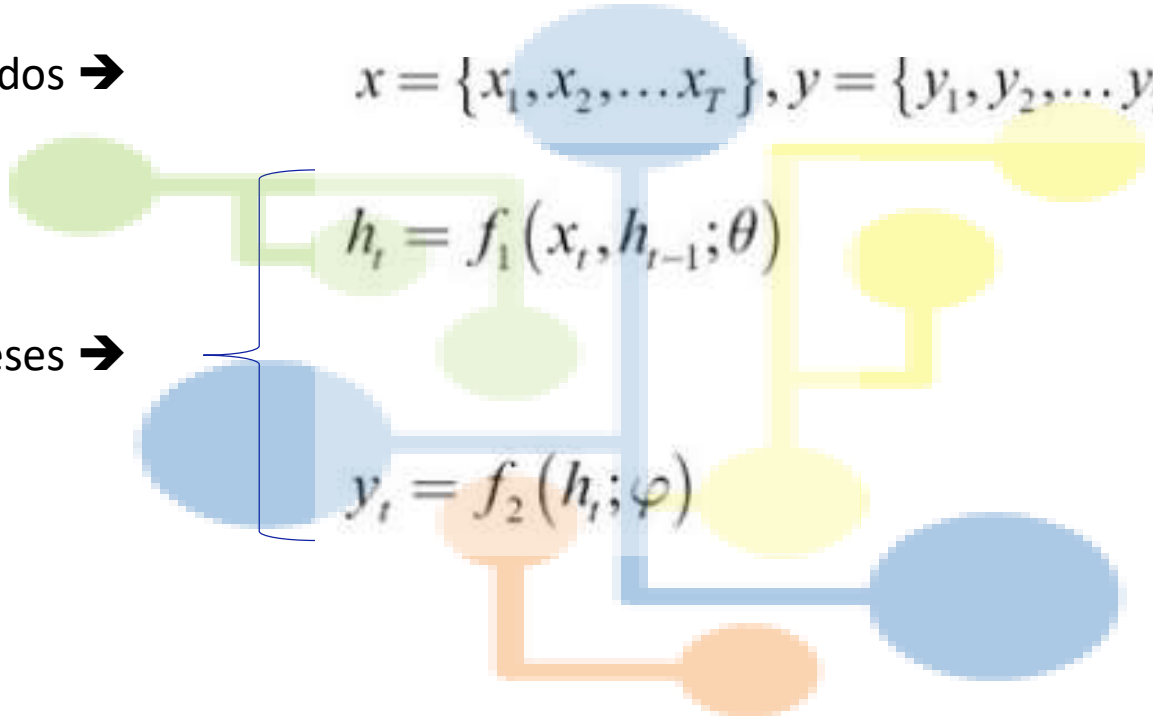
Modelagem de Sequências de Dados com Redes Neurais Recorrentes

Sequência de Dados →

$$x = \{x_1, x_2, \dots, x_T\}, y = \{y_1, y_2, \dots, y_T\}$$

Hipóteses →

$$h_t = f_1(x_t, h_{t-1}; \theta)$$

$$y_t = f_2(h_t; \varphi)$$




Modelagem de Sequências de Dados com Redes Neurais Recorrentes

Sequência de Dados →

$$x = \{x_1, x_2, \dots, x_T\}, y = \{y_1, y_2, \dots, y_T\}$$

Hipóteses →

$$h_t = f_1(x_t, h_{t-1}; \theta)$$

$$y_t = f_2(h_t; \varphi)$$

Representação da RNN →

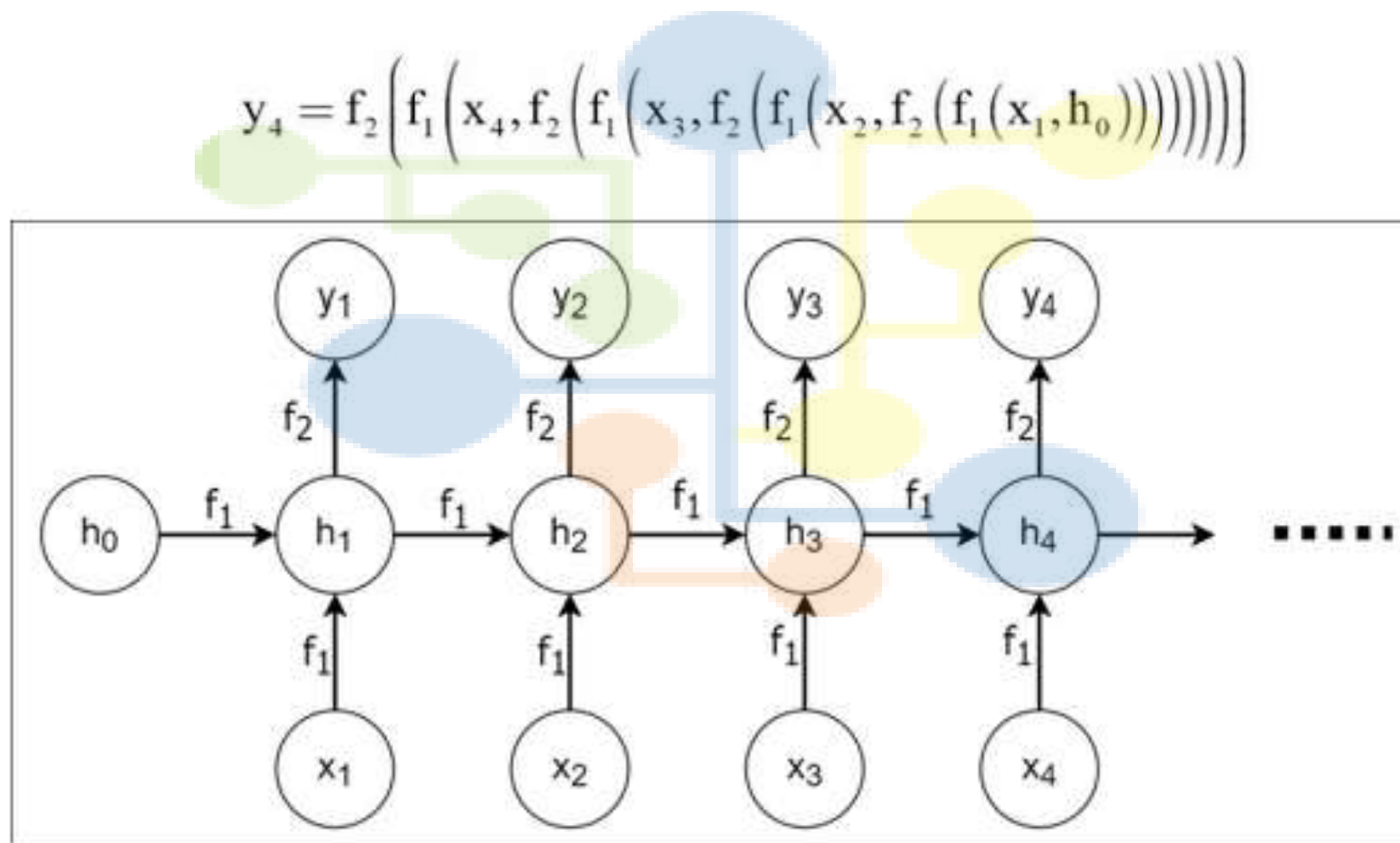
$$y_t = f_2(f_1(x_t, h_{t-1}; \theta); \varphi)$$

Representação da RNN com y_4
(por exemplo) →

$$y_4 = f_2 \left(f_1 \left(x_4, f_2 \left(f_1 \left(x_3, f_2 \left(f_1 \left(x_2, f_2 \left(f_1 \left(x_1, h_0 \right) \right) \right) \right) \right) \right) \right) \right)$$

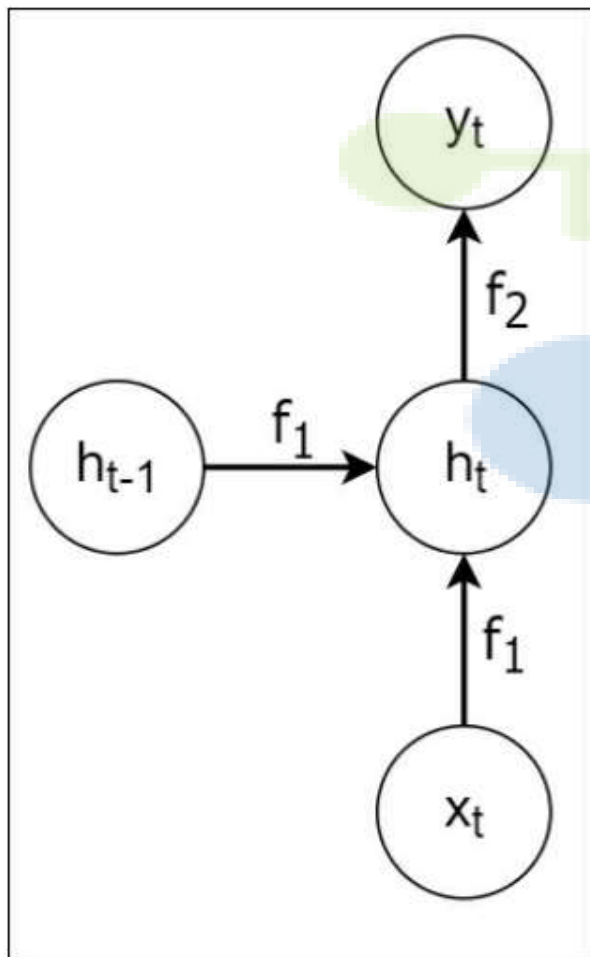


Modelagem de Sequências de Dados com Redes Neurais Recorrentes





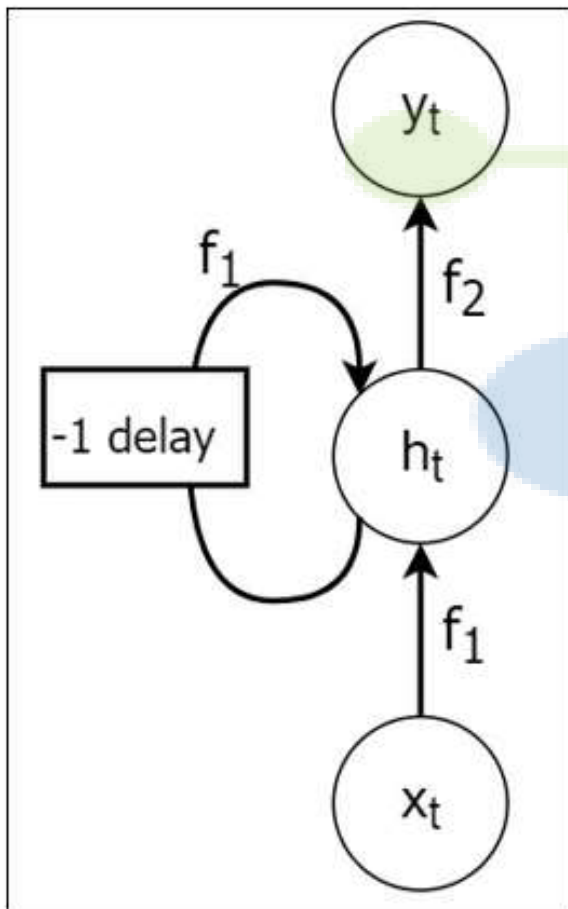
Modelagem de Sequências de Dados com Redes Neurais Recorrentes



Com isso, podemos prever uma palavra para um dado passo no tempo t .



Modelagem de Sequências de Dados com Redes Neurais Recorrentes



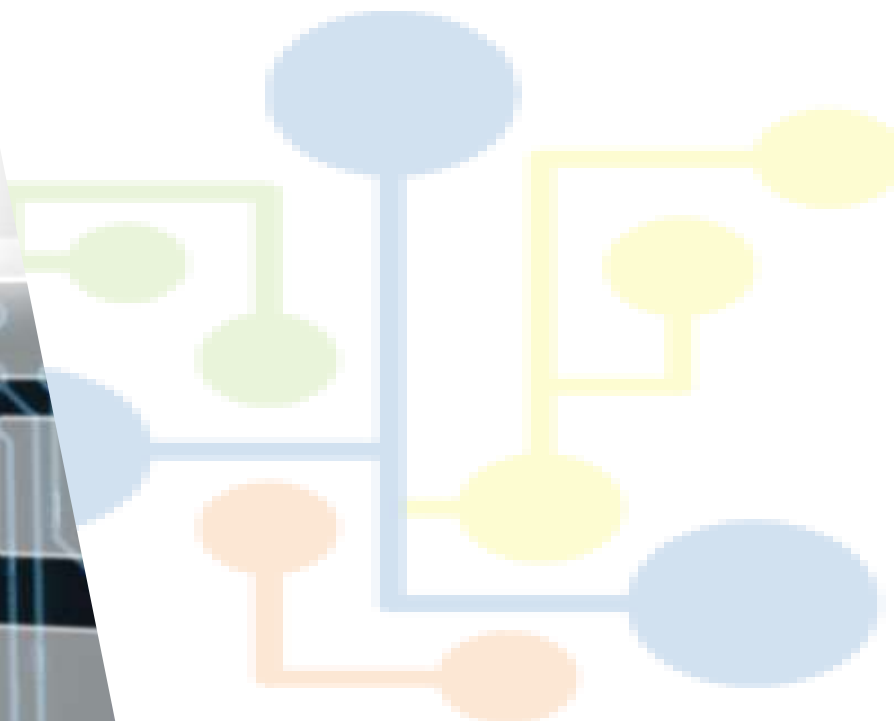
No entanto, deve ser entendido que h_{t-1} , de fato, é o que h_t era antes de receber x_t . Em outras palavras, h_{t-1} é h_t antes de um passo de tempo. Portanto, podemos representar o cálculo de h_t com uma repetição.



Data Science
Academy

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

BPTT - Backpropagation Through Time





Data Science
Academy

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

BPTT - Backpropagation Through Time

A faint, stylized diagram of a recurrent neural network (RNN) is visible in the background. It consists of several colored nodes (blue, green, yellow, orange) connected by lines, representing the flow of information and gradients over time in a recurrent architecture.

www.deeplearningbook.com.br



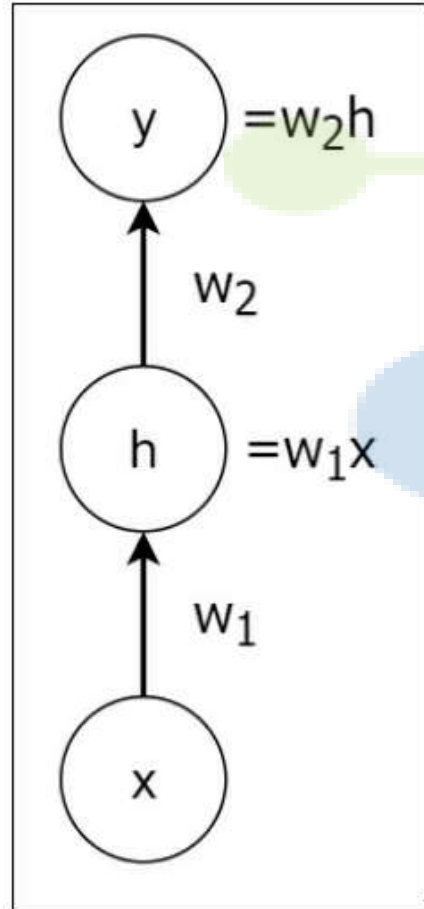
BPTT - Backpropagation Through Time

Como funciona o Backpropagation:

- 1- Calculamos a previsão para um dado input.
- 2- Calculamos o Erro E da previsão, comparando a previsão com o label original do input (usamos mean squared error ou cross-entropy).
- 3- Atualizamos os pesos em uma rede feed-forward para minimizar a perda calculada no passo anterior e fazemos isso com pequenos passos na direção oposta do gradiente, para todos os pesos em cada camada.



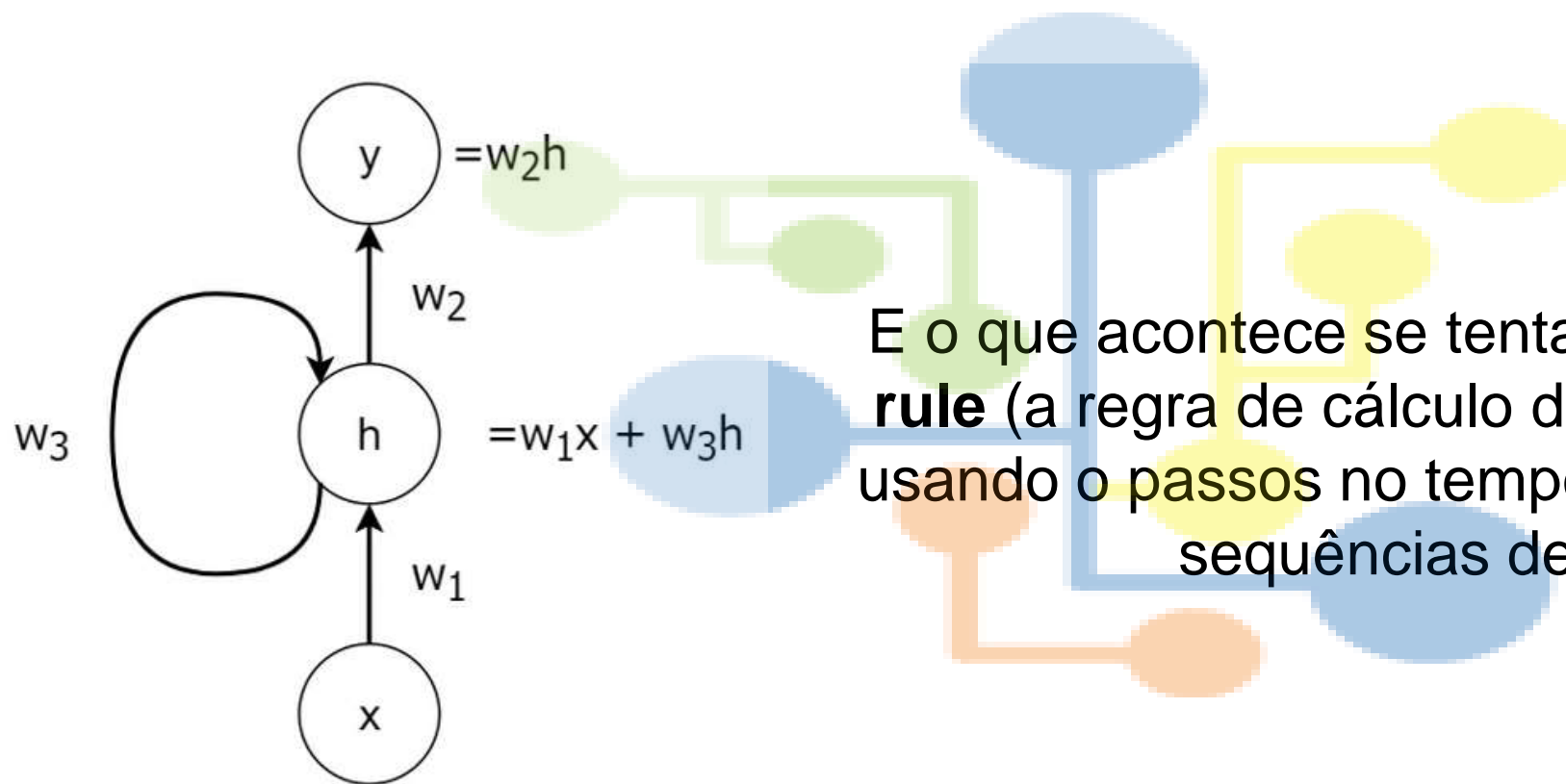
BPTT - Backpropagation Through Time



Repetimos o processo durante todas as passadas de treinamento.



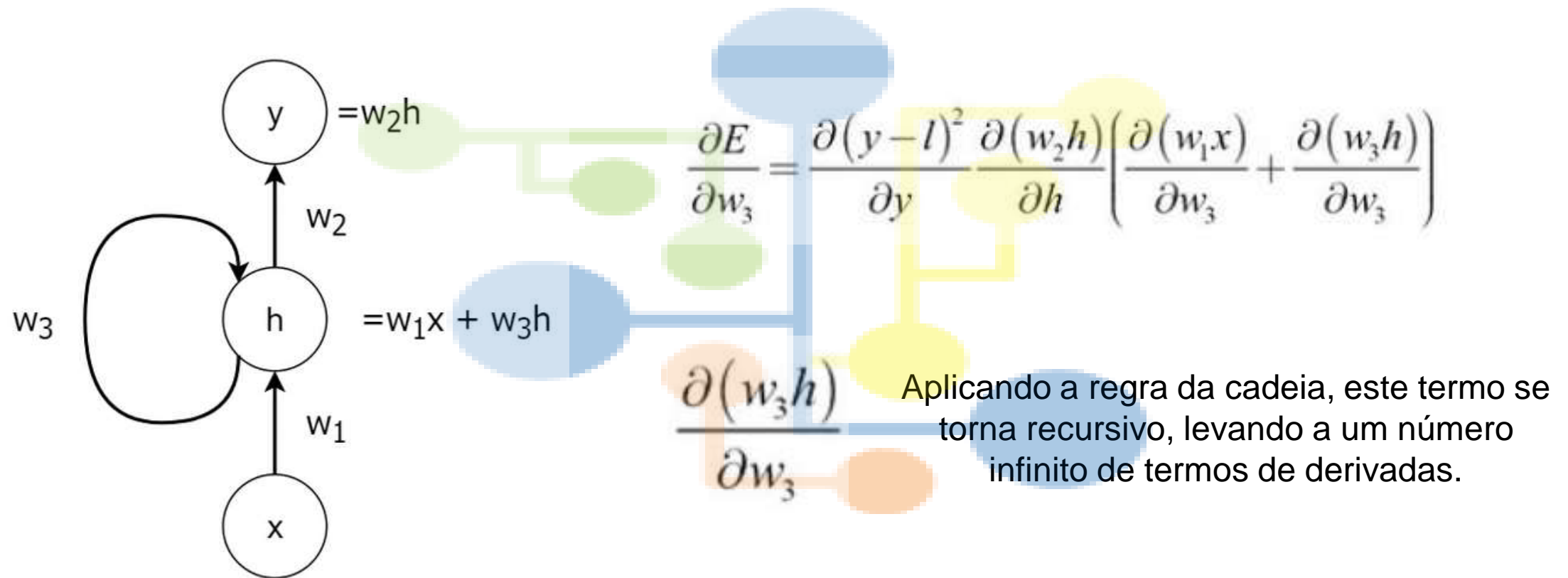
BPTT - Backpropagation Through Time



E o que acontece se tentamos incluir a **chain rule** (a regra de cálculo do Backpropagation) usando o passos no tempo que caracteriza as sequências de dados?

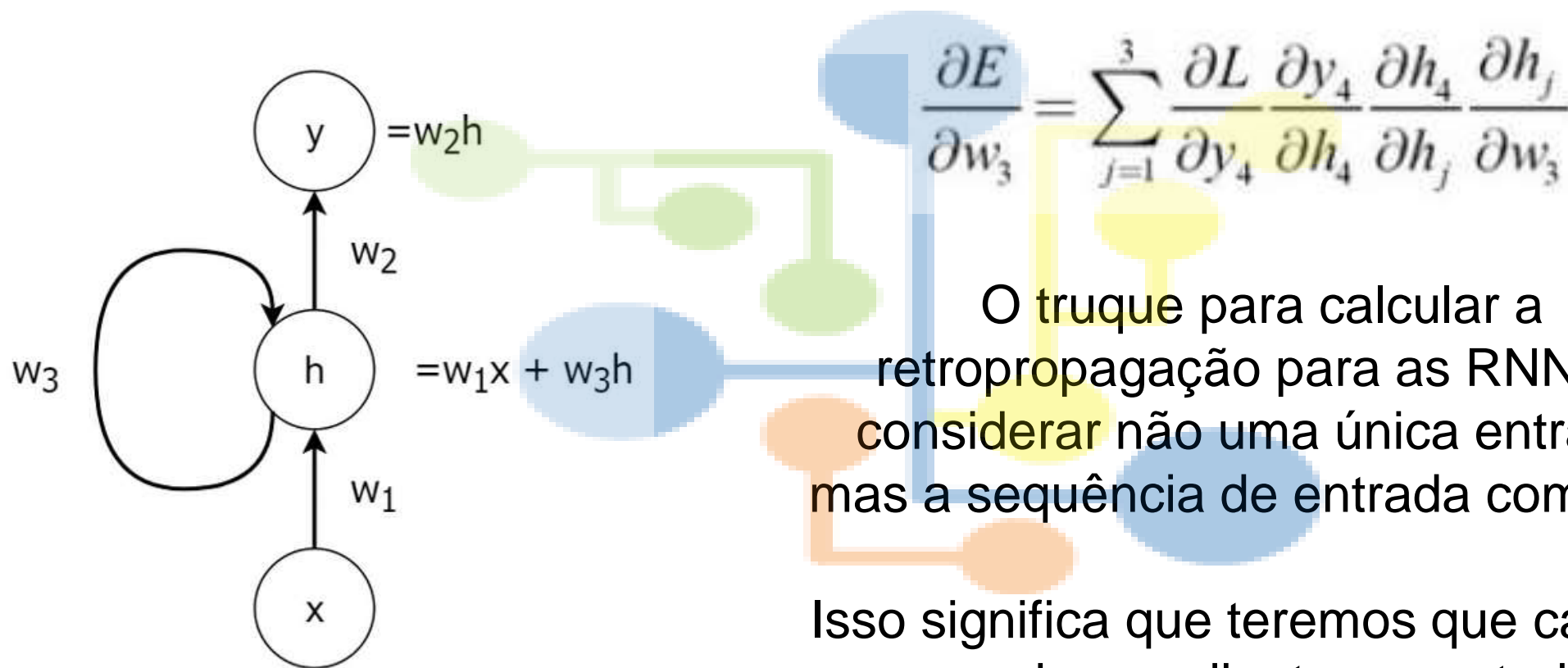


BPTT - Backpropagation Through Time





BPTT - Backpropagation Through Time



O truque para calcular a retropropagação para as RNNs é considerar não uma única entrada, mas a sequência de entrada completa.

Isso significa que teremos que calcular a soma dos gradientes para todos os passos de tempo.



TBPTT – Truncated Backpropagation Through Time

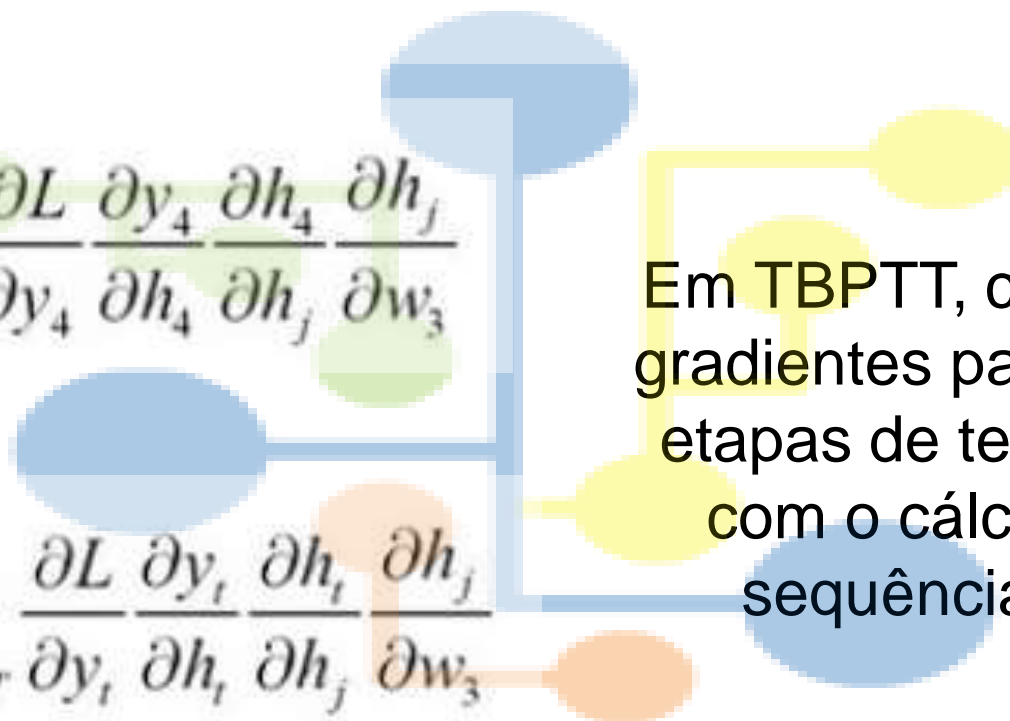
BPTT →

$$\frac{\partial E}{\partial w_3} = \sum_{j=1}^3 \frac{\partial L}{\partial y_4} \frac{\partial y_4}{\partial h_4} \frac{\partial h_4}{\partial h_j} \frac{\partial h_j}{\partial w_3}$$

TBPTT →

$$\frac{\partial E}{\partial w_3} = \sum_{j=t-T}^{t-1} \frac{\partial L}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_j} \frac{\partial h_j}{\partial w_3}$$

Em TBPTT, calculamos apenas os gradientes para um número fixo de etapas de tempo T (em contraste com o cálculo para o início da sequência como no BPTT).





Data Science
Academy

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

Obrigado