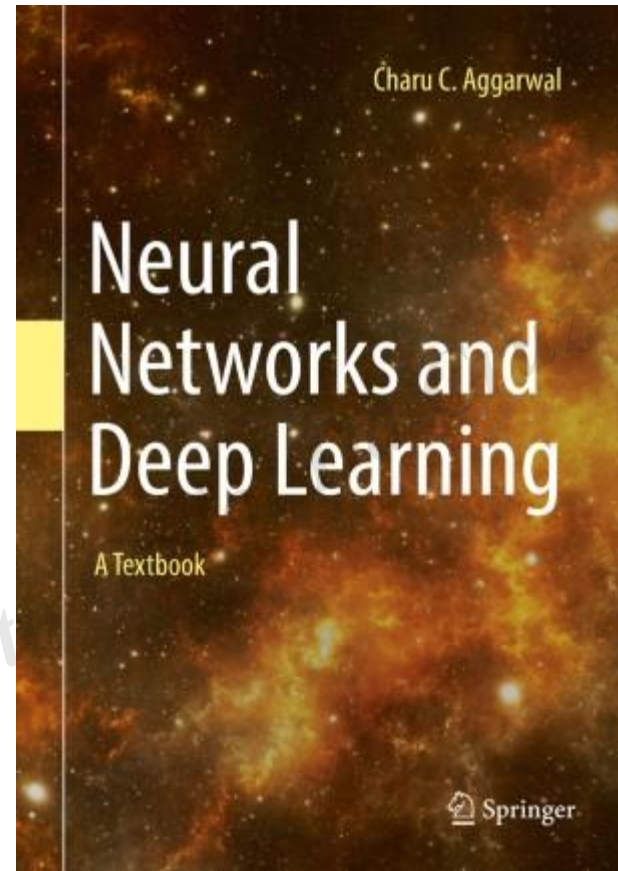
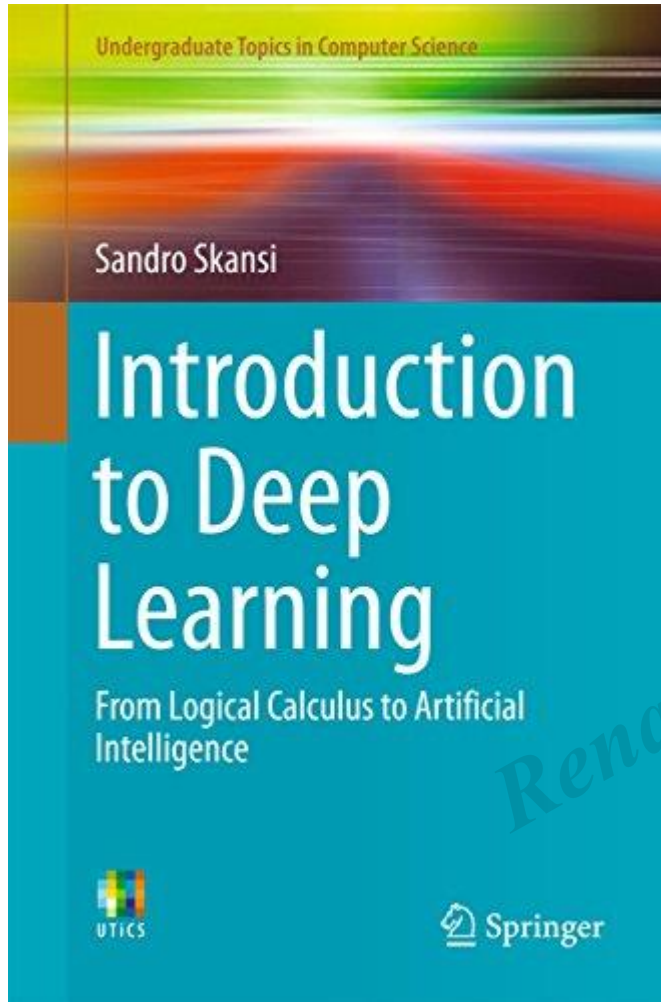


**MBA
USP
ESALQ**

DEEP LEARNING

Prof. Dr. Jeronimo Marcondes

Introdução



Introdução

- Alguns problemas importantes:

1. Dados de texto
2. Séries Temporais
3. Assistir um filme

RNN

- Exemplo de uma série temporal.



Introdução

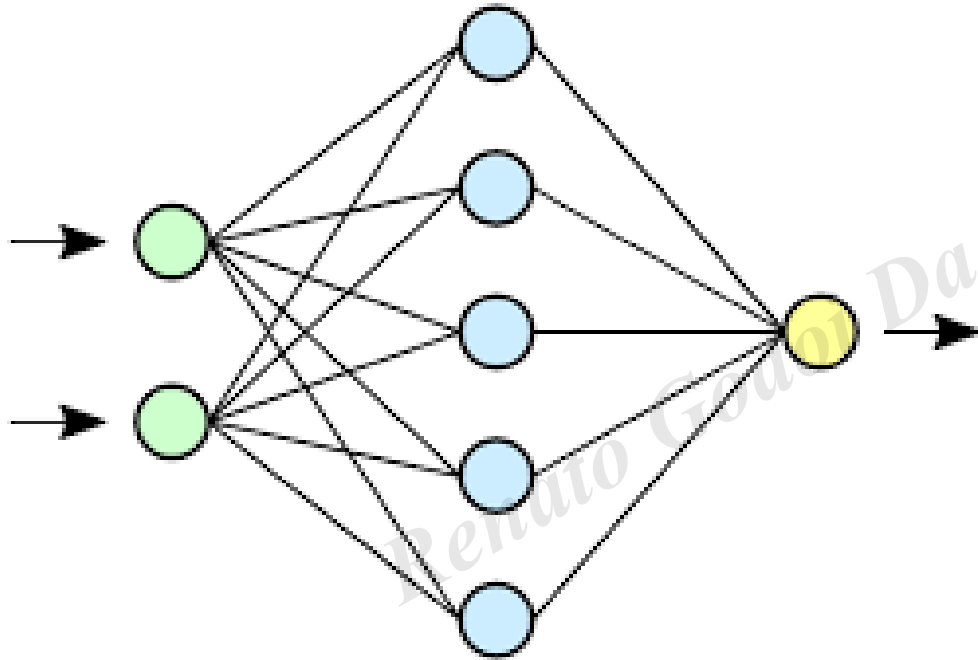
- Qual é a característica desses dados? Sequência.
- FNN pode ser usada para reproduzir esse processo, mas não é a melhor escolha.

“O gato perseguiu o rato”

“O rato perseguiu o gato”

RNN

- O que ocorreria se usássemos uma FNN?



RNN

('\$', 'all')

('\$ all', 'I')

('\$ all I', 'want')

('\$ all I want', 'for')

('\$ all I want for', 'Christmas')

('\$ all I want for Christmas', 'is')

('\$ all I want for Christmas is', 'you')

('\$ all I want for Christmas is you', '&').

Fonte: Introduction to Deep Learning from Logical Calculus to Artificial Intelligence

RNN

- RNN constrói distribuição de probabilidade

- Exemplo

‘My name is Cassidy’

‘My name is Myron’

‘My name is Marcus’

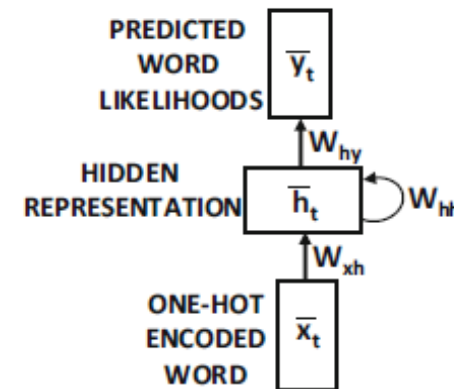
‘My name is Marcus’

‘My name is Marcus’.

Fonte: Introduction to Deep Learning from Logical Calculus to Artificial Intelligence

RNN

- Objetivo: rede neural com “memória”.
- Recorrente: realiza a mesma tarefa para todos os elementos e sua saída depende dos cálculos anteriores.

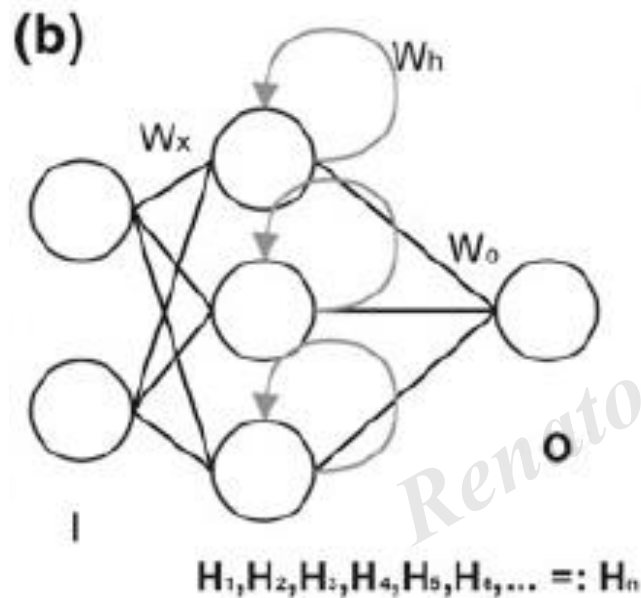


(a) RNN

Fonte: Neural Networks and Deep Learning

RNN

- Podemos visualizar a RNN de outra forma:



Fonte: Introduction to Deep Learning from Logical Calculus to Artificial Intelligence

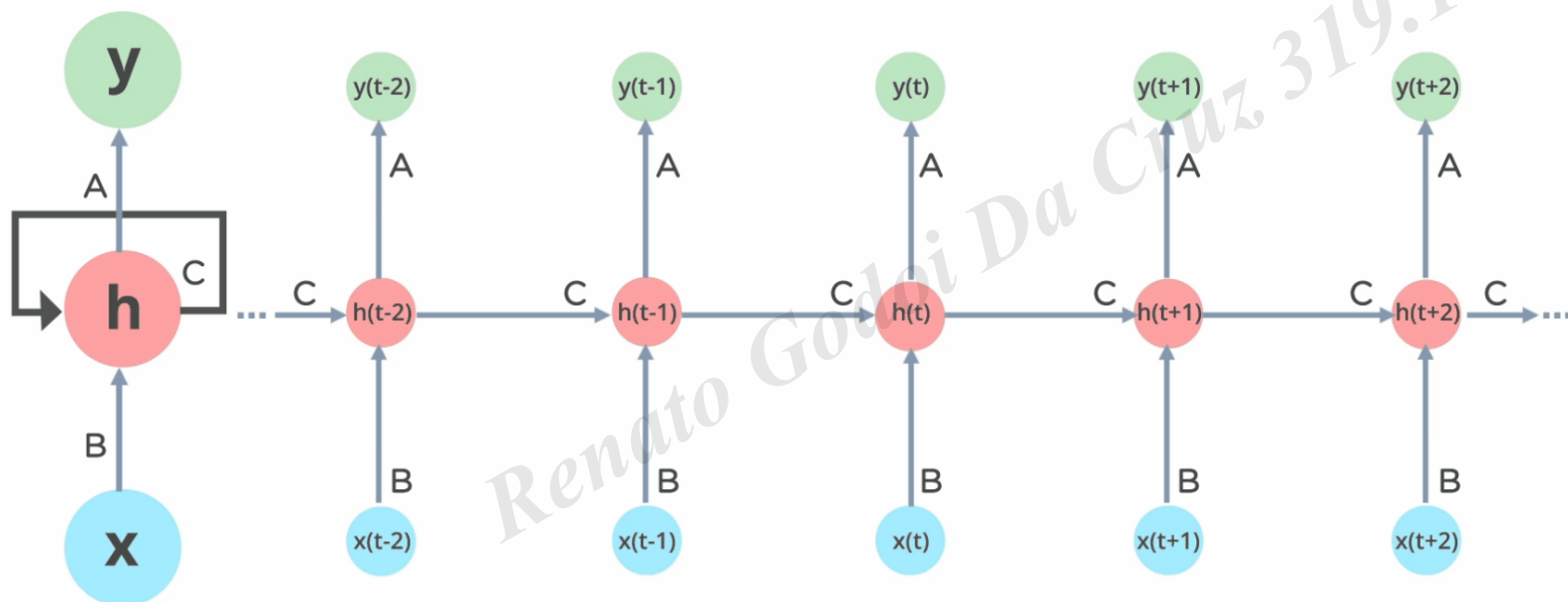
RNN

A camada de entrada 'x' recebe a entrada para a rede neural e a processa e a passa para a camada intermediária.

A camada intermediária 'h' pode consistir em várias camadas ocultas, cada uma com suas próprias funções de ativação e pesos e vieses.

Fonte: <https://www.simplilearn.com/tutorials/deep-learning-tutorial/rnn>

RNN

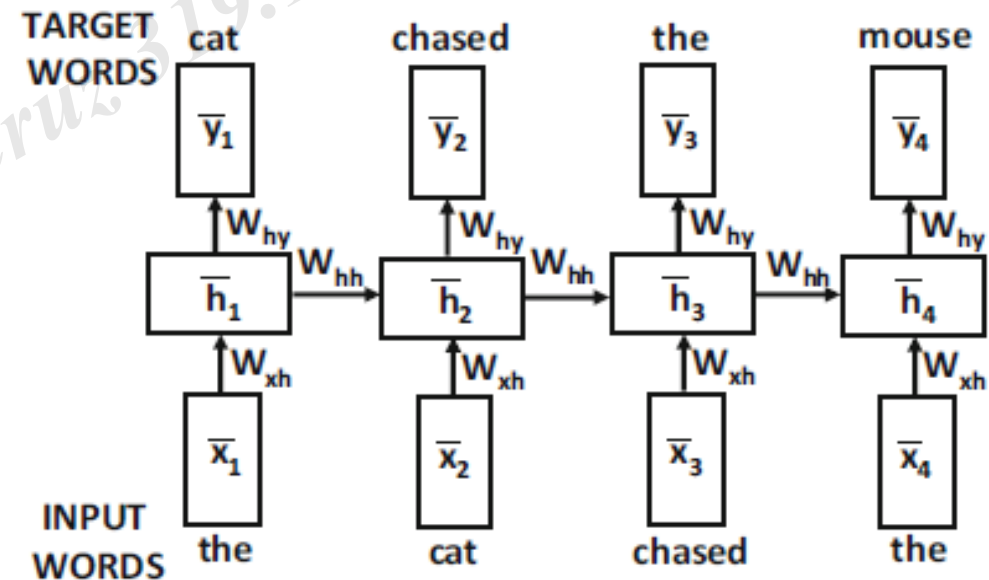


Fonte: <https://www.simplilearn.com/tutorials/deep-learning-tutorial/rnn>

RNN

- Podemos visualizar a RNN de outra forma:
- The cat chased the mouse

Fonte: Neural Networks and Deep Learning



(b) Time-layered representation of (a)

RNN

- Isso significa que temos uma rede que “guarda” todo o passado
- Explicar crescimento econômico com base em nível de confiança.
- Simple Recurrent Neural Network

$$h(t) = f_h(\mathbf{w}_h^\top h(t-1) + \mathbf{w}_x^\top x(t))$$

$$y(t) = f_o(\mathbf{w}_o^\top h(t)),$$

RNN

- Rede de Elman

$$y(t) = f(\mathbf{w}_o^\top h(t)) = \quad (7.1)$$

$$= f(\mathbf{w}_o^\top f(\mathbf{w}_h^\top h(t-1) + \mathbf{w}_x^\top x(t))) = \quad (7.2)$$

$$= f(\mathbf{w}_o^\top f(\mathbf{w}_h^\top f(\mathbf{w}_h^\top h(t-2) + \mathbf{w}_x^\top x(t-1)) + \mathbf{w}_x^\top x(t))) = \quad (7.3)$$

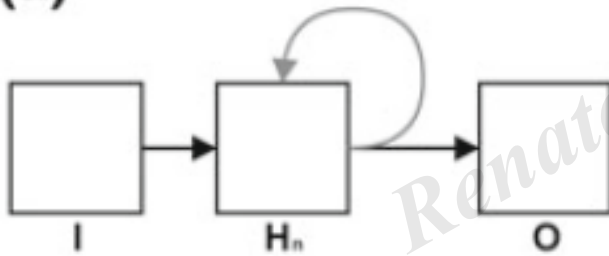
$$= f(\mathbf{w}_o^\top f(\mathbf{w}_h^\top f(\mathbf{w}_h^\top f(\mathbf{w}_h^\top h(t-3) + \mathbf{w}_x^\top x(t-2)) + \mathbf{w}_x^\top x(t-1)) + \mathbf{w}_x^\top x(t))). \quad (7.4)$$

- Tenho resultado que os valores da camada escondida estarão multiplicados por valores de pesos elevados a potencias maiores quanto mais antiga for essa informação

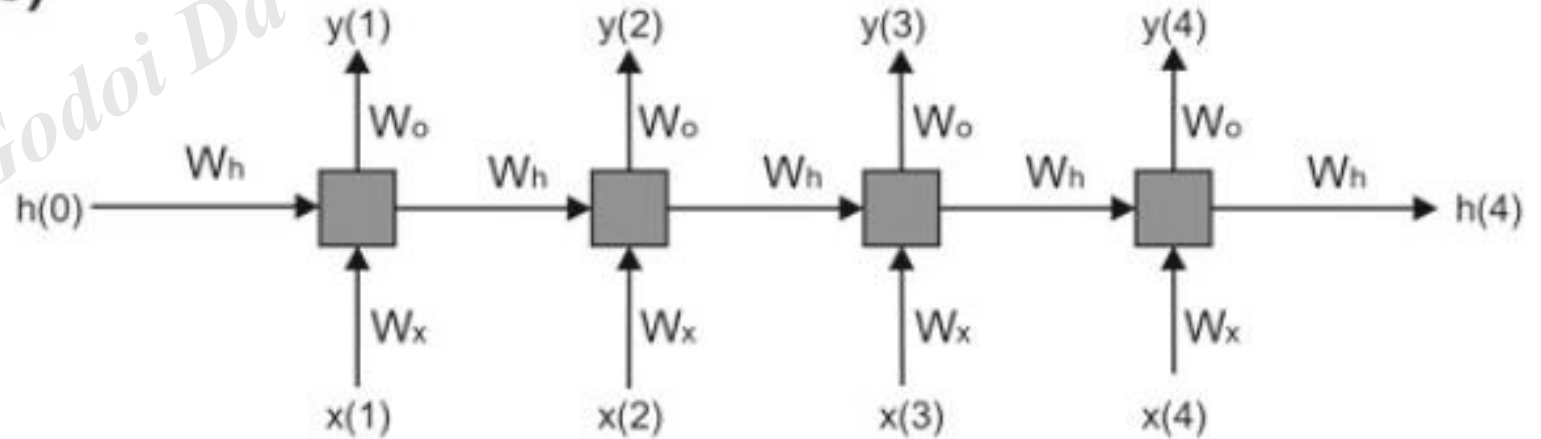
Backpropagation

- Tudo se afeta ao mesmo tempo!
- Backpropagation Through Time

(a)



(c)



Backpropagation

- Como é feito o cálculo do gradiente?
- O quanto varia o erro para uma dada variação do peso?
- Surge um problema: passado multiplicado por pesos elevados a potencias maiores quanto mais antigo for.
- Intuição: quanto mais longe no passado, mais difícil ver a influência, pois muita coisa aconteceu.

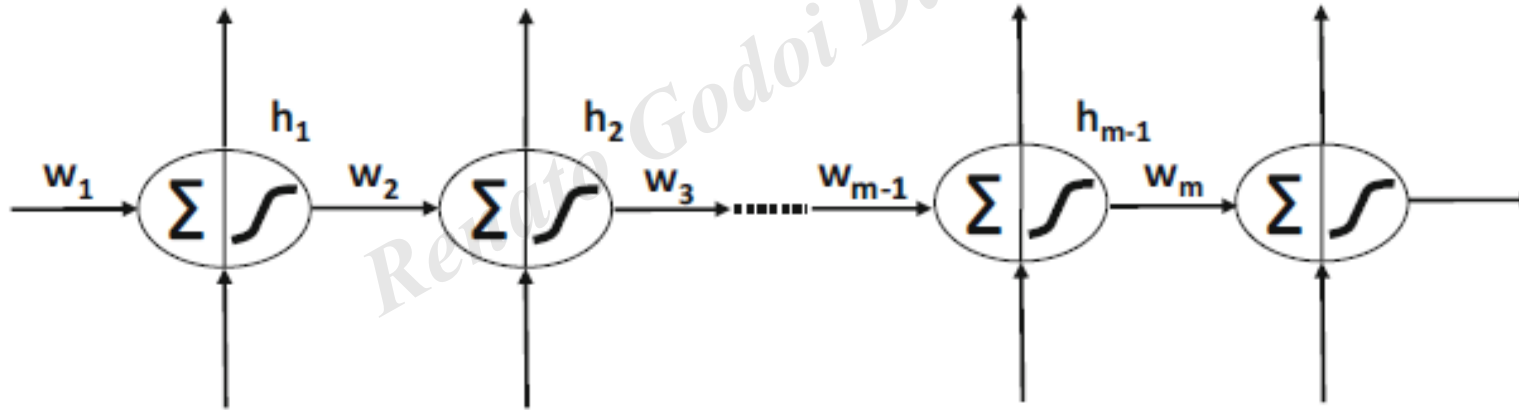
Dissipação do Gradiente

- Intuição
- Multiplicação de números menores do que 1
- Multiplicação de números maiores do que 1

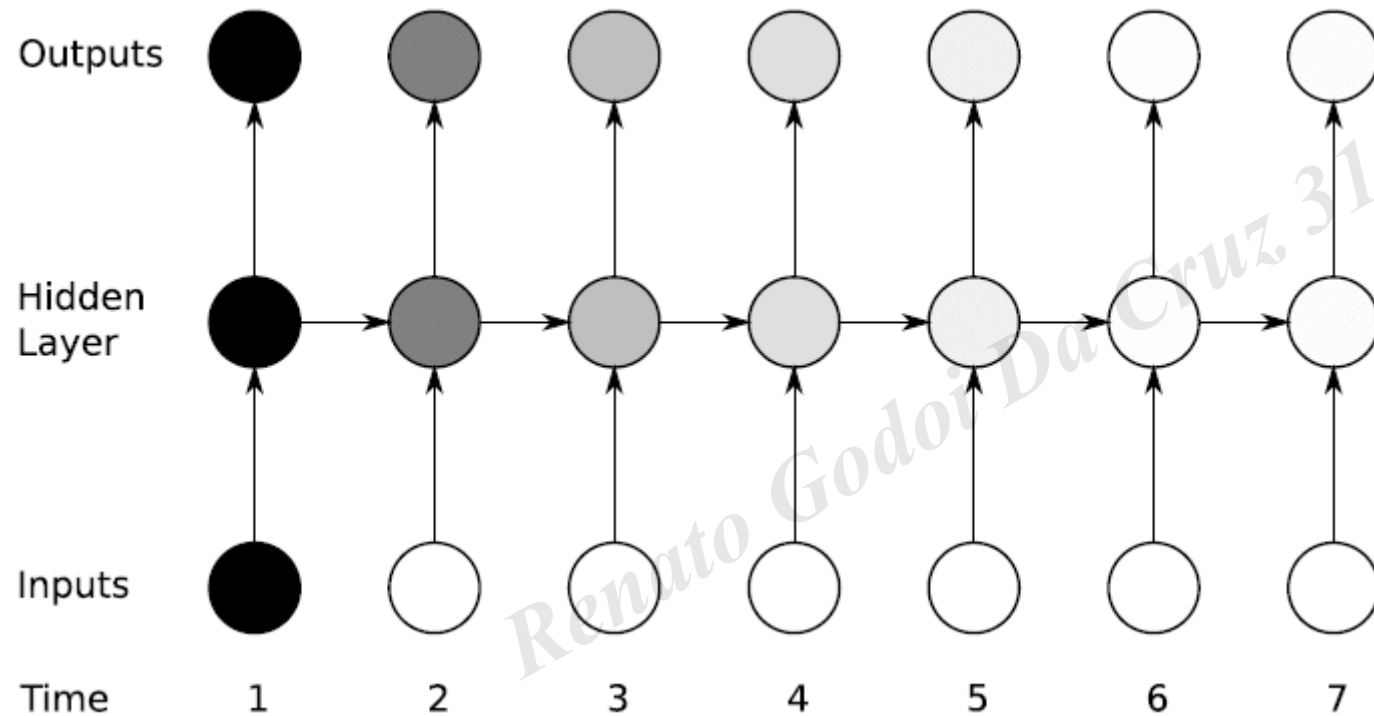
Renato Góes Da Cruz 219.150.228-94

Problemas com Gradiente

- “Sumiço” (Dissipação) e Explosão do Gradiente
- $w > 1$ ou $w < 1$



Dissipação do Gradiente



Problemas com Gradiente

- Ocorre em qualquer rede – mais comum em RNN
- RNN é mais profunda?
- Como resolver?

Backpropagation Truncada

- O processo de backpropagation truncada consiste em parar a avaliação de mudanças de pesos até certo ponto. A atualização não levará em conta todo o passado, mas só até certo limite.
- Custo computacional
- Solução Arbitrária

Resolver Vanishing Gradient

- Inicialização de matriz de peso
- Função de ativação ReLU:

$$f(x) = \max(0, x)$$

Renato Gomes Da Cruz 319.150.228-94

Clipping do Gradiente

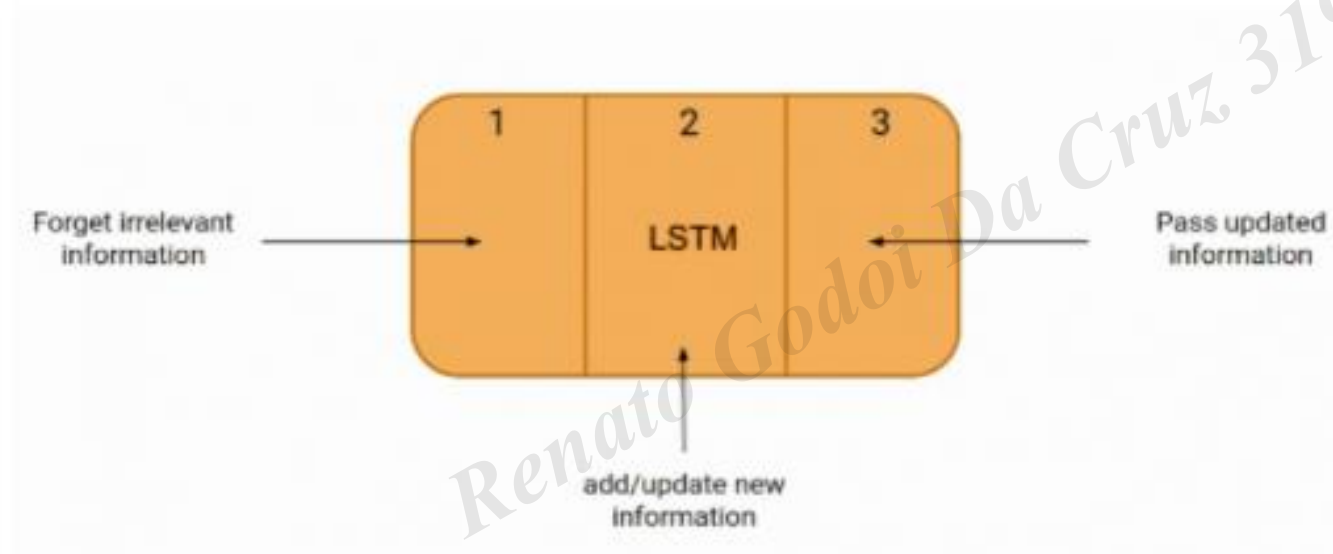
- Solução possível para Dissipação e Explosão

O clipping define um valor limite definido nos gradientes, o que significa que, mesmo se um gradiente aumentar além do valor predefinido durante o treinamento, seu valor ainda será limitado ao limite definido. Dessa forma, a direção do gradiente permanece inalterada e apenas a magnitude do gradiente é alterada. (deeplearningbook.com).

LSTM

- Long Short-Term Memory
- Mesma coisa que na RNN – mas temos o “estado da célula”
- Com base em “portões” – gates
- Será que devemos manter ou guardar uma informação?

LSTM - intuição



Fonte: <https://www.analyticsvidhya.com/>

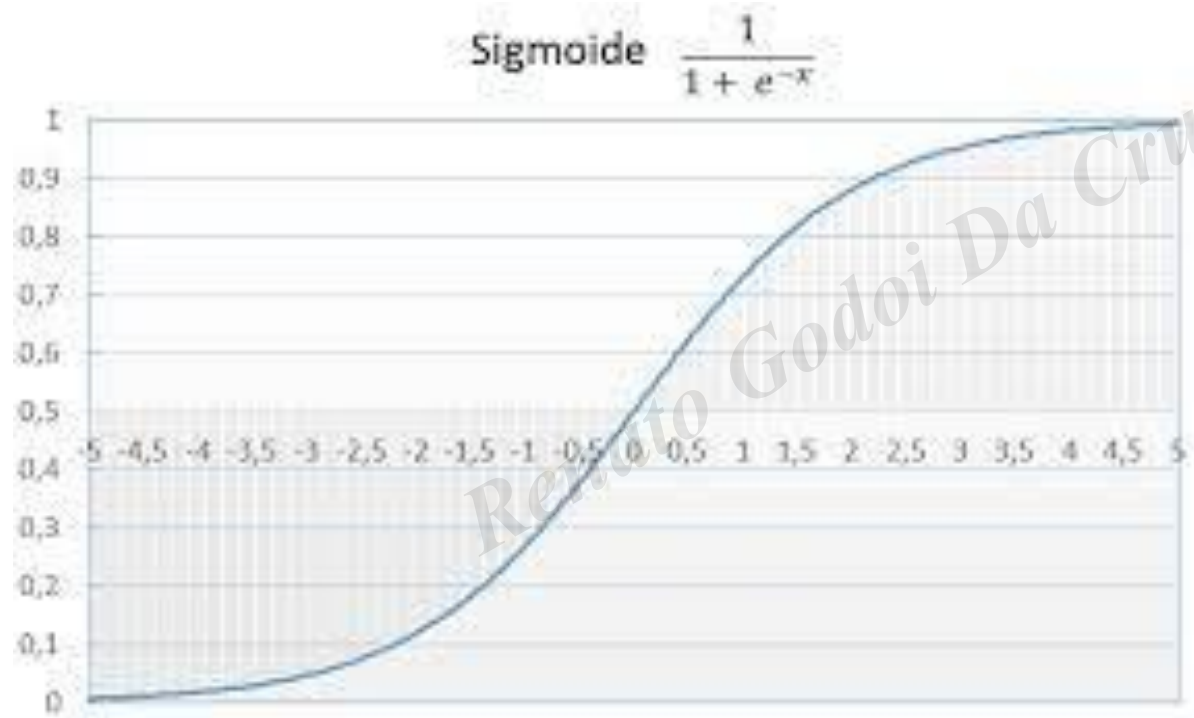
LSTM

- Algumas funções importantes
- TANH – tangente hiperbólica
- Resultado entre -1 e 1 => “negativo”, “neutro” e “positivo”

$$\tanh = \frac{\sinh(t)}{\cosh(t)}$$

LSTM

- Sigmoidoide
- Resultado entre 0 e 1 => “sim” ou “não”



Fonte: Research Gate

LSTM

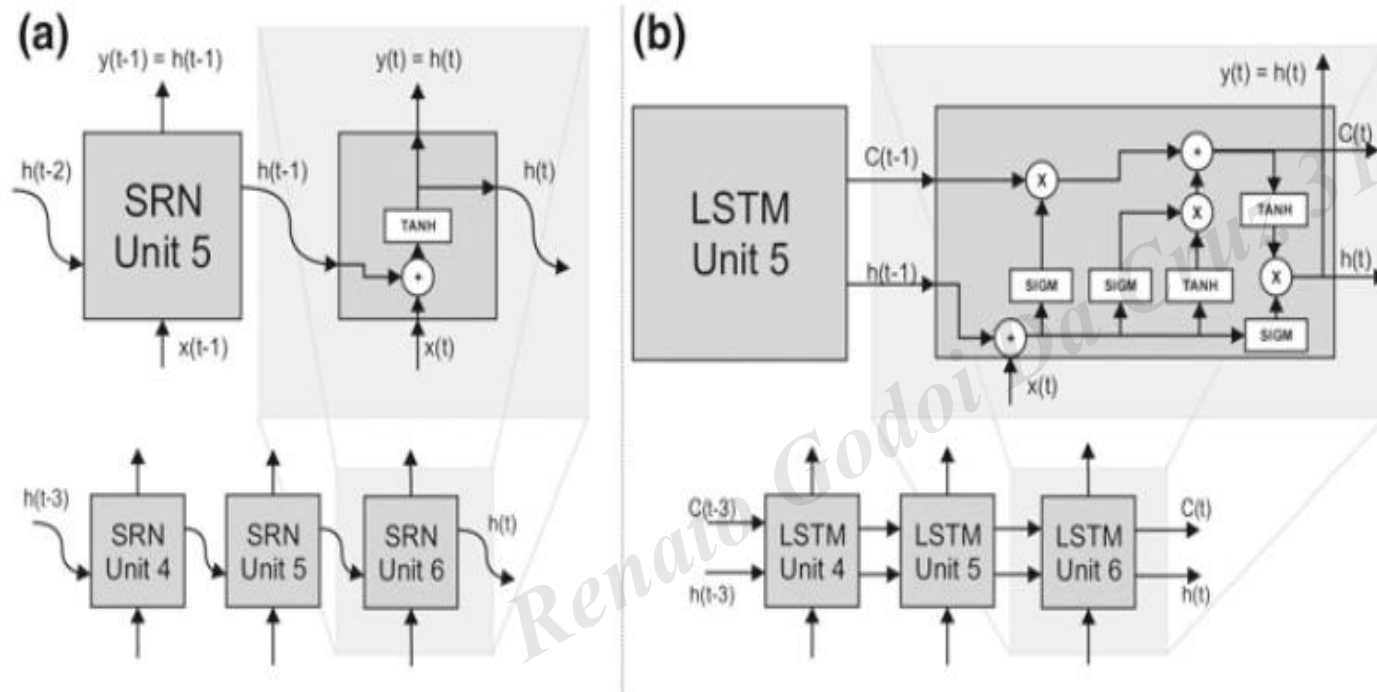
- Ligação da entrada com cada escondida – igual
- Estado da célula – memória longa

Renato Goabí Da Cruz 319.150.228-94

LSTM

- Forget Gate – Quanto lembrar?
- Input Gate – Quanto manter dos inputs? O que adicionar ao estado da célula?
- Output Gate – O que do estado da célula e do hidden state será utilizado como resultado?

LSTM



LSTM

É possível observar que dado um instante no tempo t , a célula LSTM possui como entradas o instante atual de alimentação de informação da rede, identificado como x_t , o estado oculto (hidden state) h_{t-1} e o estado da célula (cell state) c_{t-1} , ambos estados provindos da recorrência do instante de tempo passado $t-1$. As saídas da célula são o cell state c_t do instante atual, o hidden state h_t e a saída de informação y_t . Para o caso da célula pertencer a última camada da rede, o h_t é entendido como a saída final y_t , para o caso da camada ser interna a rede, o h_t servirá como o h_{t-1} para a próxima camada adiante na rede.

LSTM

Além das entradas e saídas, uma célula LSTM é composta internamente por combinações entre funções de ativação, adições e produtos. Estas operações internas da célula LSTM se denominam de gates, sendo eles, forget gate, input gate, cell gate e output gate. Além destes gates, a célula LSTM possui uma região responsável por agrupar a saída de alguns destes gates para produzir o c_t , que é uma das saídas da célula. (OLIVEIRA, E.V., 2020)

LSTM – Forget Gate



LSTM

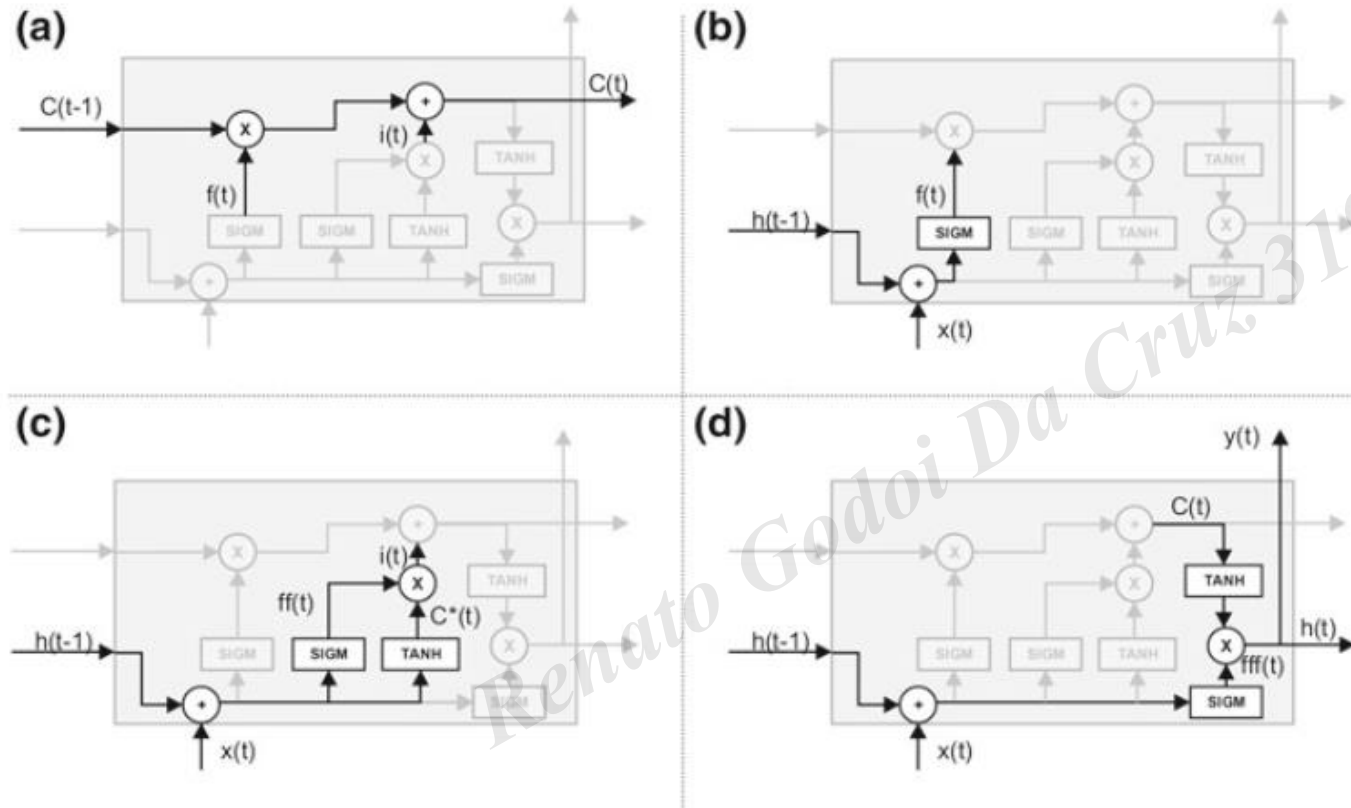
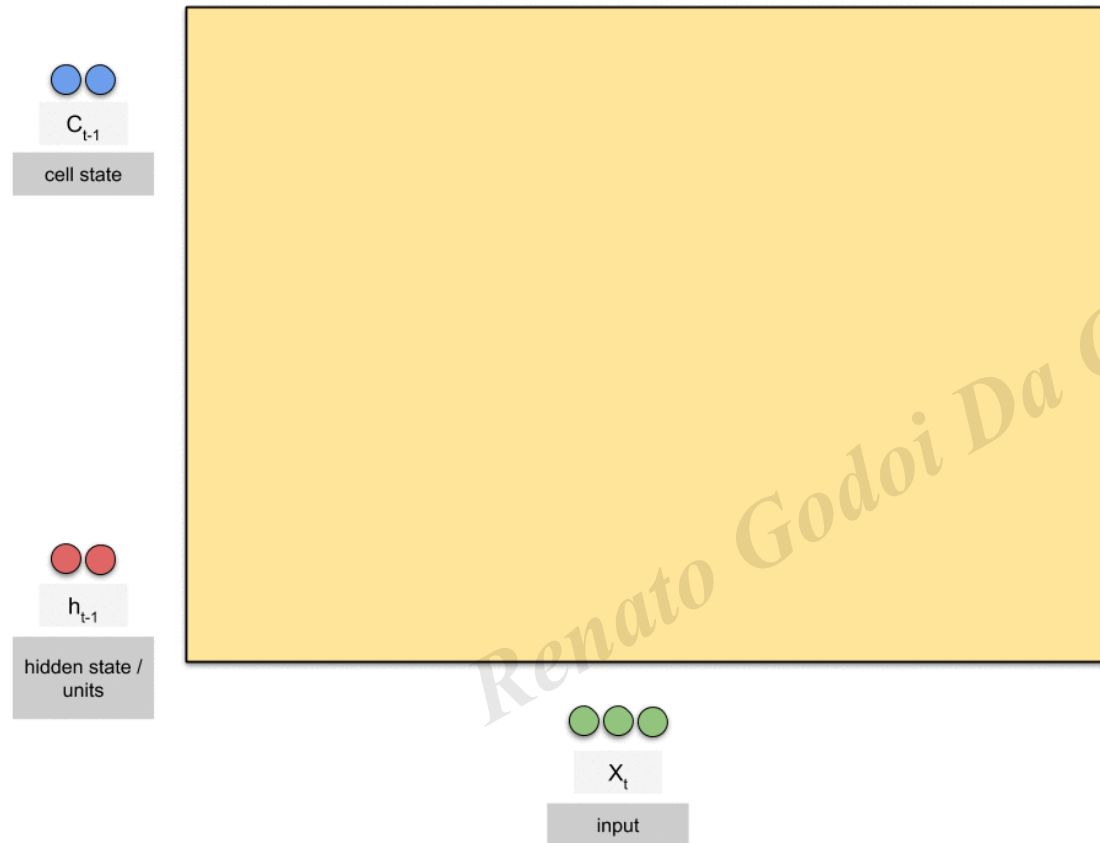


Fig. 7.4 Cell state (a), forget gate (b), input gate (c) and output gate (d)

Animação LSTM



GRU – Gate Recurrent Unit

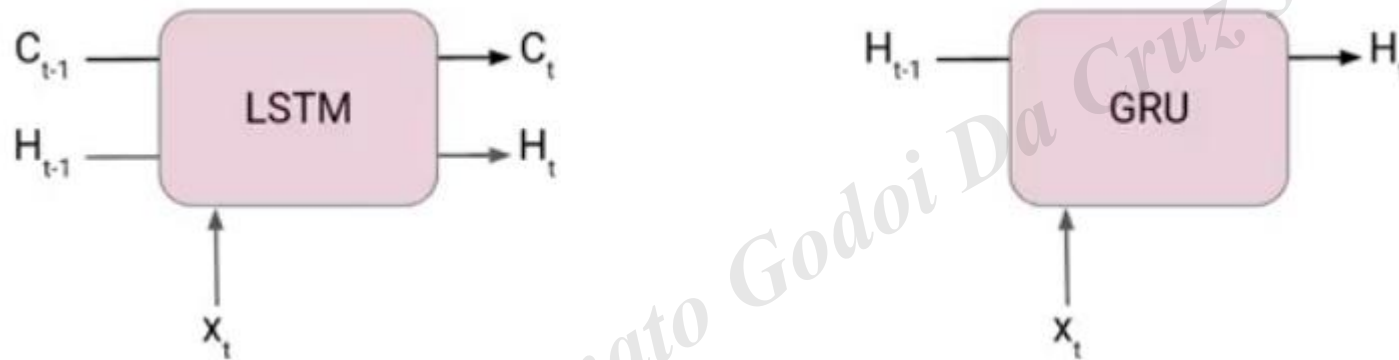
- Resolve problema da dissipação do gradiente
- Baseada em portões: reinicialização e atualização
- Apenas 1 estado oculto

GRU

- Guarda dependências longas
- Reset Gate: Quanto de informação anterior iremos ignorar
- Update Gate: Quanto de informação anterior iremos manter

Renato Góes Da Cruz 319.150.228-94

GRU



Fonte: <https://www.analyticsvidhya.com/>

Extra: Transformadores

- Maior parte de PLN era feito com RNN
- Attention is all you need
- Uso de RNN => perde informação conforme se distância do início de uma série
- Contexto é essencial em PLN

Extra: Transformadores

- Codificador – decodificador
- Codificador – processa informações sobre a entrada e relacionamentos entre si
- Decodificador - faz o oposto, pegando todas as codificações e as processando, usando suas informações contextuais incorporadas para gerar uma sequência de saída.

Extra: Transformadores

- A lógica: o que é mais rápido para achar uma solução, ler um livro inteiro ou procurar no índice?
- Vetor de contexto – guarda posição dentro da sequência
- Solução: passar todos estados ocultos



<https://www.linkedin.com/in/jeronymo-marcondes-585a26186>

Renato Cadoi Da Cruz 319.150.228-94