



循环神经网络

作者: Calvin

QQ: 179209347

Mail: 179209347@qq.com

介绍

笔记简介:

- 面向对象: 深度学习初学者
- 依赖课程: **线性代数, 统计概率**, 优化理论, 图论, 离散数学, 微积分, 信息论

知乎专栏:

<https://zhuanlan.zhihu.com/p/693738275>

Github & Gitee 地址:

https://github.com/mymagicpower/AIAS/tree/main/deep_learning

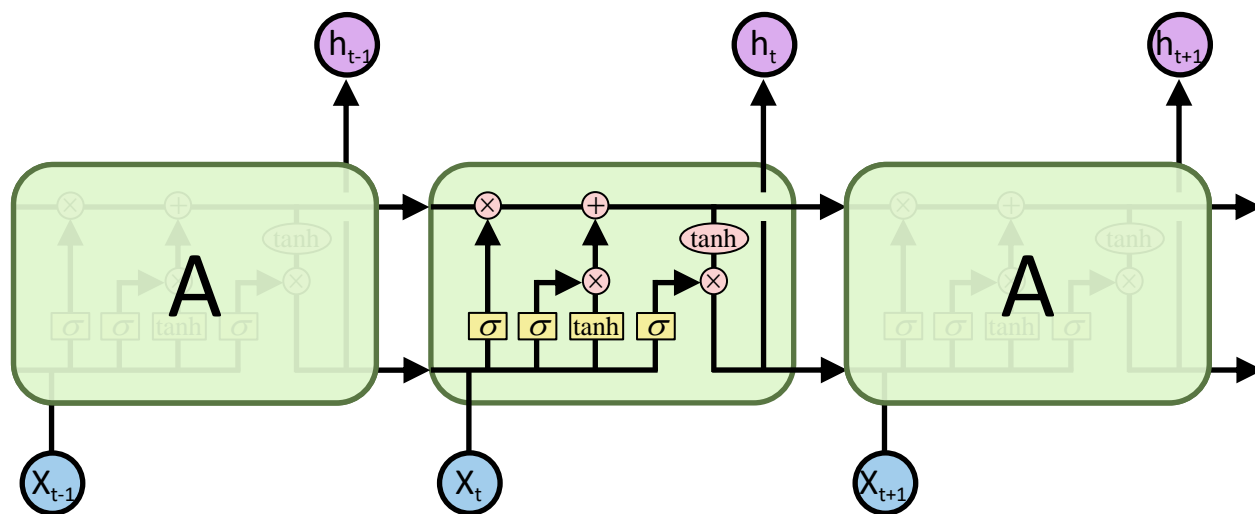
https://gitee.com/mymagicpower/AIAS/tree/main/deep_learning

* 版权声明:

- 仅限用于个人学习
- 禁止用于任何商业用途

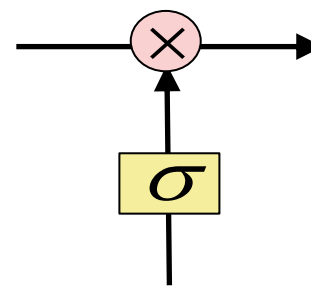
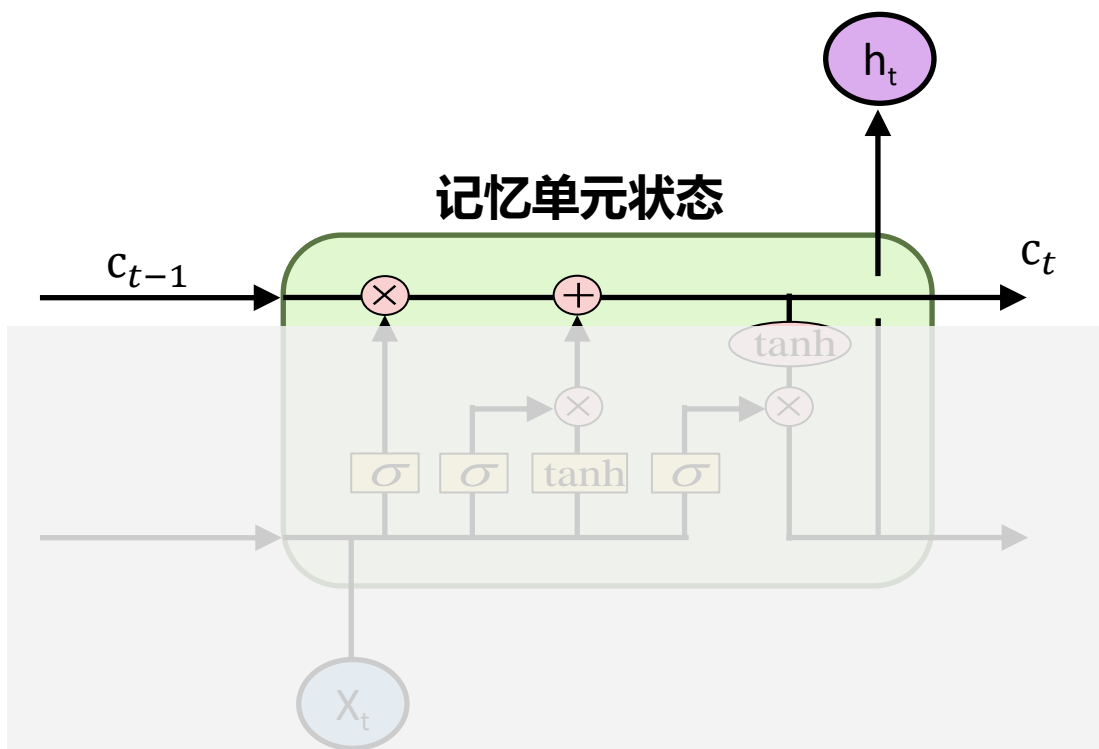
长短期记忆神经网络 (Long Short-Term Memory, LSTM)

长短期记忆神经网络 (Long Short-Term Memory, LSTM) 是一种常用于处理序列数据的深度学习模型，特别适用于需要长期记忆和捕捉时间依赖关系的任务。相比于传统的循环神经网络 (RNN)，LSTM在解决梯度消失和梯度爆炸等问题上表现更好。



长短期记忆神经网络的原理

长短期记忆神经网络通过门结构来有效地捕捉和记忆长期依赖关系。LSTM中的门结构包括三种门：遗忘门（Forget Gate）、输入门（Input Gate）和输出门（Output Gate），以及一个单元状态（Cell State）用于存储信息。

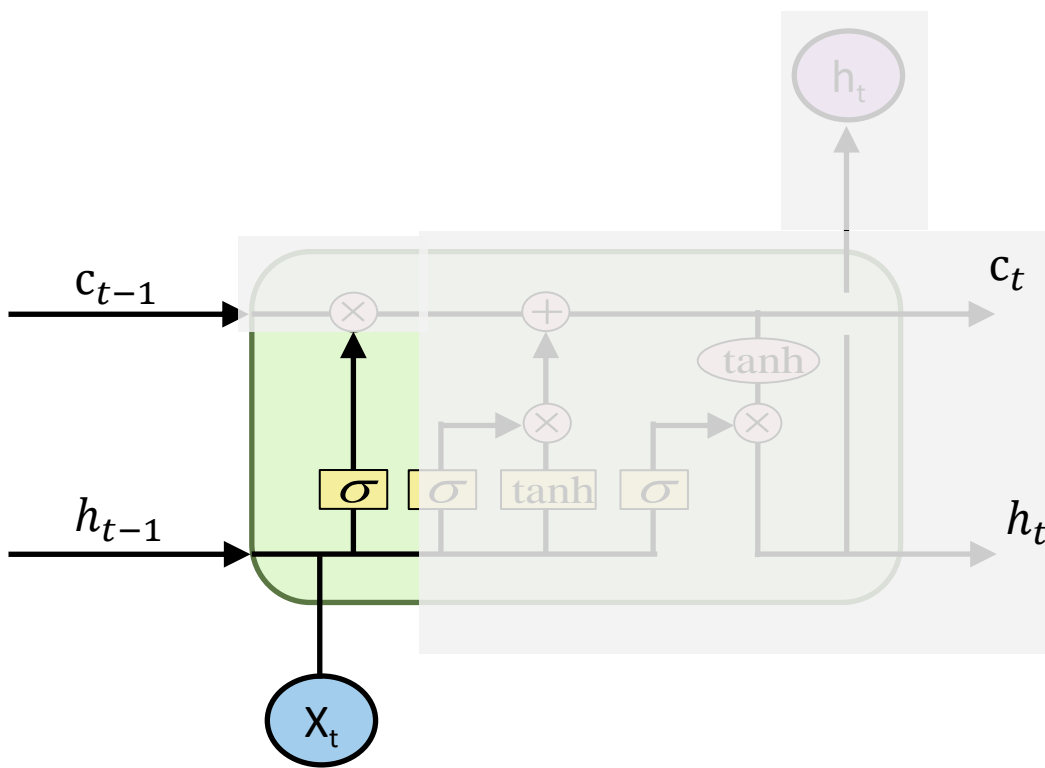


“门”结构

门可以选择性地让信息通过。它们由 S 形神经网络层和逐点乘法运算组成。

长短期记忆神经网络的原理 - 遗忘门

遗忘门 (Forget Gate)：遗忘门决定了在当前时间步应该忘记多少过去的信息。它通过一个sigmoid激活函数来输出一个0到1之间的值，表示每个单元状态中的信息应该被保留 (1) 还是被遗忘 (0)。



遗忘门计算步骤：

- ① 输入：数据为 h_{t-1} 和 x_t
- ② Sigmoid 激活函数计算
- ③ 输出： f_t 决定遗忘哪些数据，0表示“完全遗忘”，1表示“完全保留”。

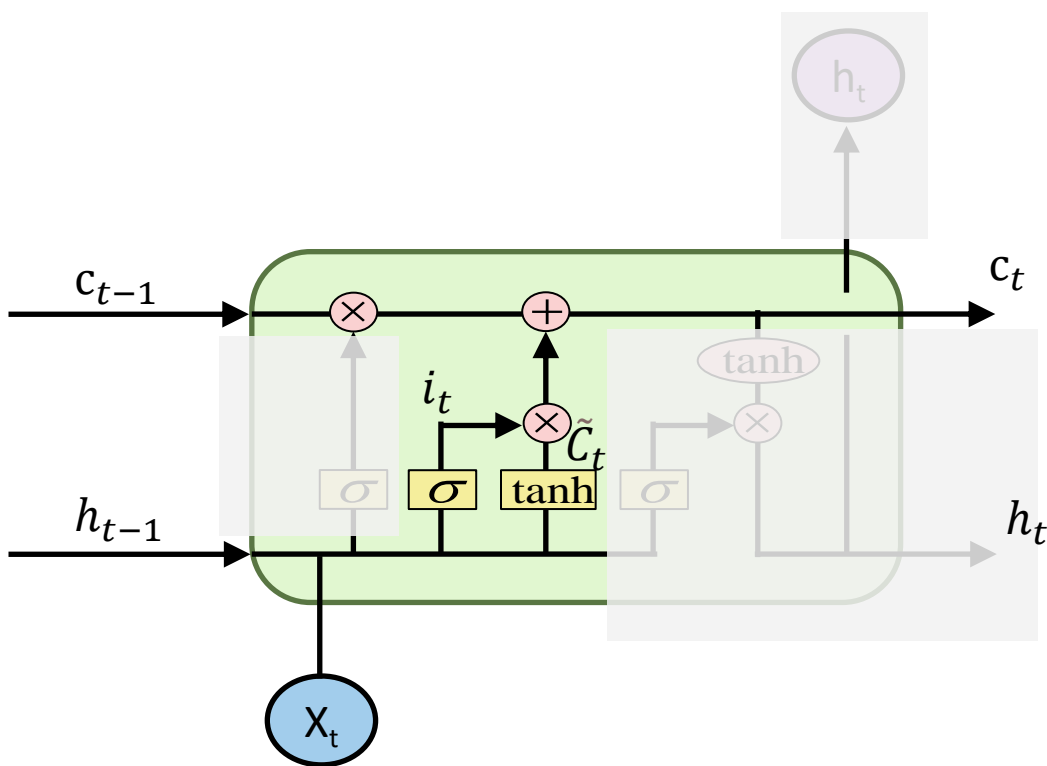
计算公式如下：

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

- $\sigma(\)$ 为 Sigmoid 函数
- “ \cdot ” 为点乘运算，
- $[h_{t-1}, x_t]$ 为输入数据
- W_f 为遗忘门的权重
- b_f 为遗忘门的偏置。

长短期记忆神经网络的原理 - 输入门

输入门 (Input Gate)：输入门决定了在当前时间步应该更新多少新的信息到单元状态中。它包括两部分：一个sigmoid激活函数用于确定更新哪些信息，以及一个tanh激活函数用于创建一个新的候选值向量，这个向量将会被加入到单元状态中。



输入门计算步骤:

- ① 输入数据为 h_{t-1} 和 x_t
- ② Sigmoid 激活函数计算，利用Sigmoid激活函数输出数值在0 ~ 1的特性，决定更新哪些数据
- ③ Tanh层创建新的候选值向量 \tilde{C}_t

计算公式如下:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

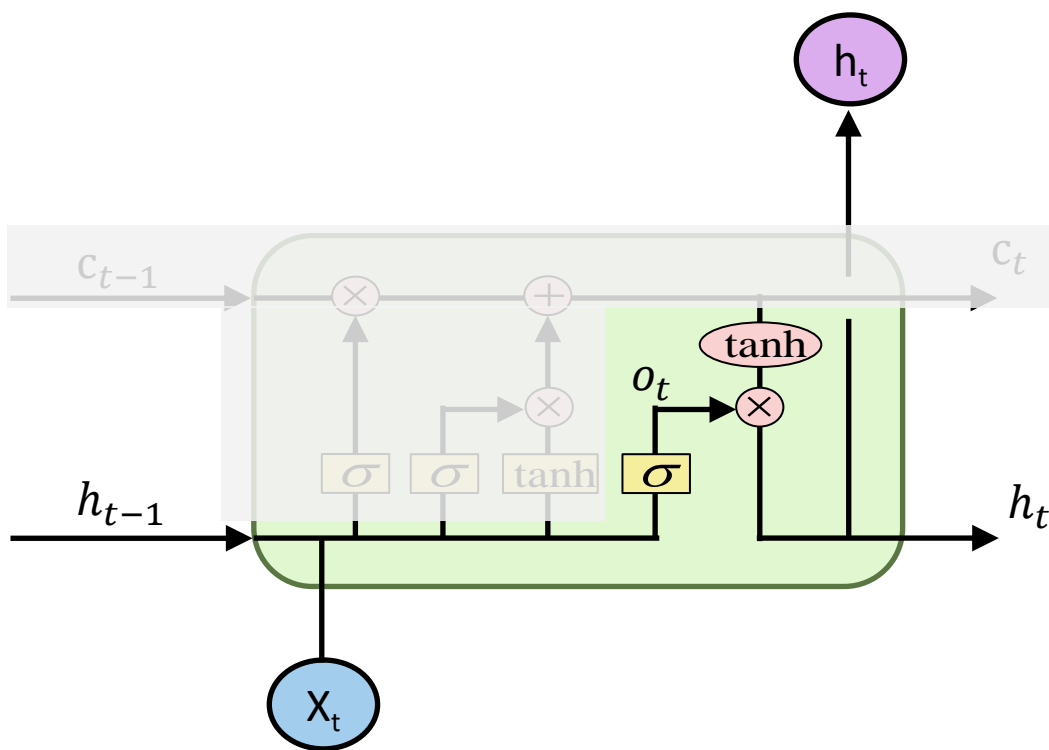
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

- $\sigma(\)$ 为Sigmoid函数
- “ \cdot ” 为点乘运算,
- $[h_{t-1}, x_t]$ 为输入数据
- W_i 为输入门的权重
- b_i 为输入门的偏置。

长短期记忆神经网络的原理 - 输出门

输出门 (Output Gate)：输出门决定了在当前时间步应该输出多少信息到下一个时间步。它通过一个sigmoid激活函数来确定单元状态的哪部分将被输出，然后通过一个tanh激活函数将细胞状态进行缩放，最终输出到下一个时间步。



输出门计算步骤：

- ① 输入：数据为 h_{t-1} 和 x_t
- ② Sigmoid 激活函数计算 o_t ，利用Sigmoid激活函数输出数值在0~1的特性，决定更新哪些数据
- ③ Tanh层再和 S 形网络层的输出值相乘

计算公式如下：

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

- $\sigma(\)$ 为Sigmoid函数
- “.” 为点乘运算，
- $[h_{t-1}, x_t]$ 为输入数据
- W_o 为输出门的权重
- b_o 为输出门的偏置。



Thank

You