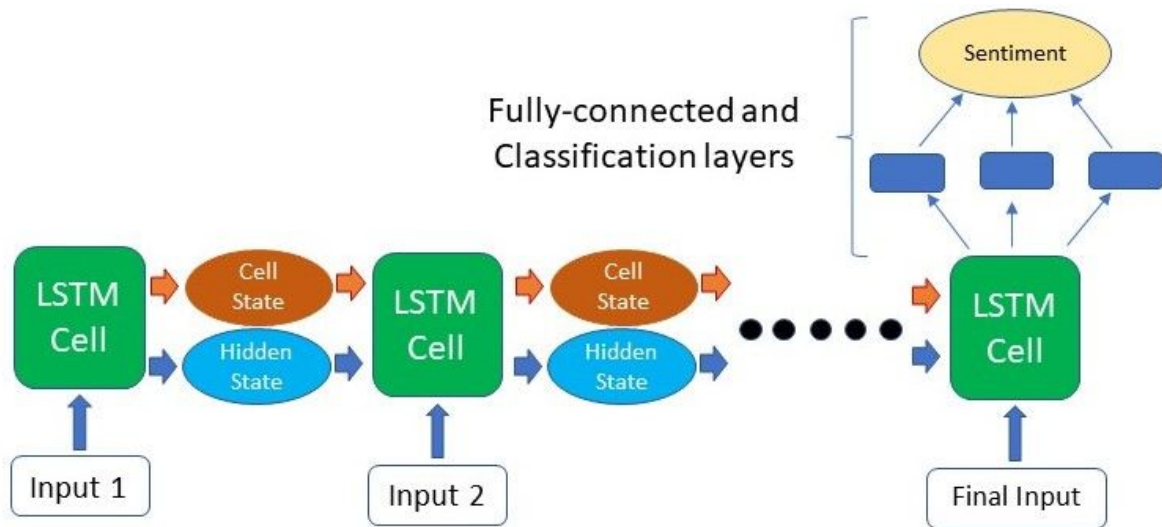


LSTM

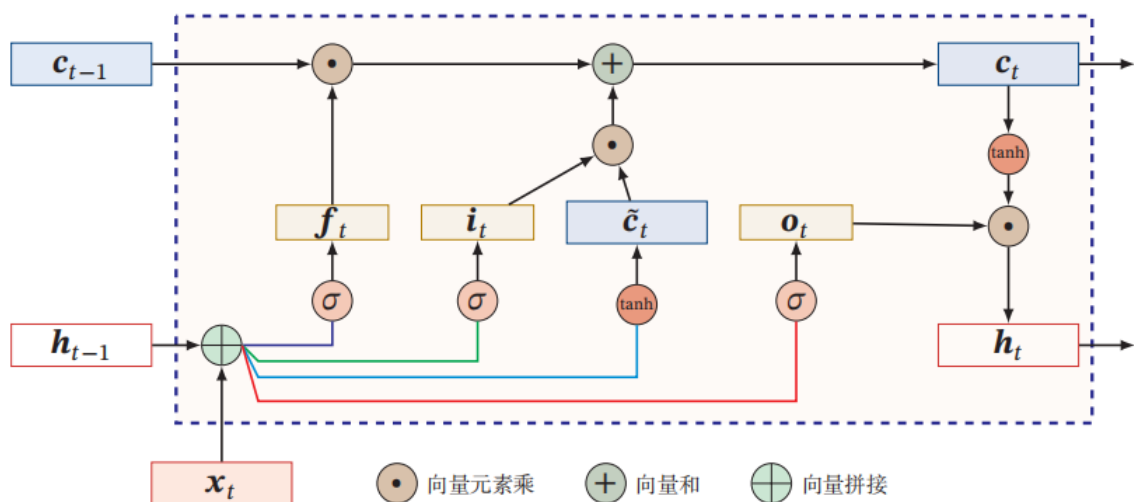
2021-12-26

1. LSTM简介
2. LSTM单元结构
3. Pytorch LSTM
 3. 1. 模型参数
 3. 2. 输入
 3. 3. 输出
4. 参考文档

1. LSTM简介



2. LSTM单元结构



LSTM 网络的循环单元结构

1. 输入

x_t : 输入数据

h_{t-1} : hidden state

c_{t-1} : cell state

2. 候选状态

\tilde{c} : 从输入数据和hidden state中提取的信息

$$\tilde{c}_t = \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c)$$

3. 门

遗忘门: 对 c_{t-1} , 遗忘其多少信息

输入门: 对 \tilde{c} , 保存其多少信息

输出门: 对 c_t , 输出其多少信息到 h_t

三个门的计算方式为:

$$\begin{aligned} \mathbf{i}_t &= \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i) \\ \mathbf{f}_t &= \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f) \\ \mathbf{o}_t &= \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o) \end{aligned} \quad (3)$$

其中, $\sigma(\cdot)$ 为 Logistic函数, \mathbf{x}_t 为当前输入, \mathbf{h}_{t-1} 为上一时刻的hidden state

4. 输出

$$\begin{aligned} \mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \\ \mathbf{h}_t &= \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \end{aligned} \quad (4)$$

3. Pytorch LSTM

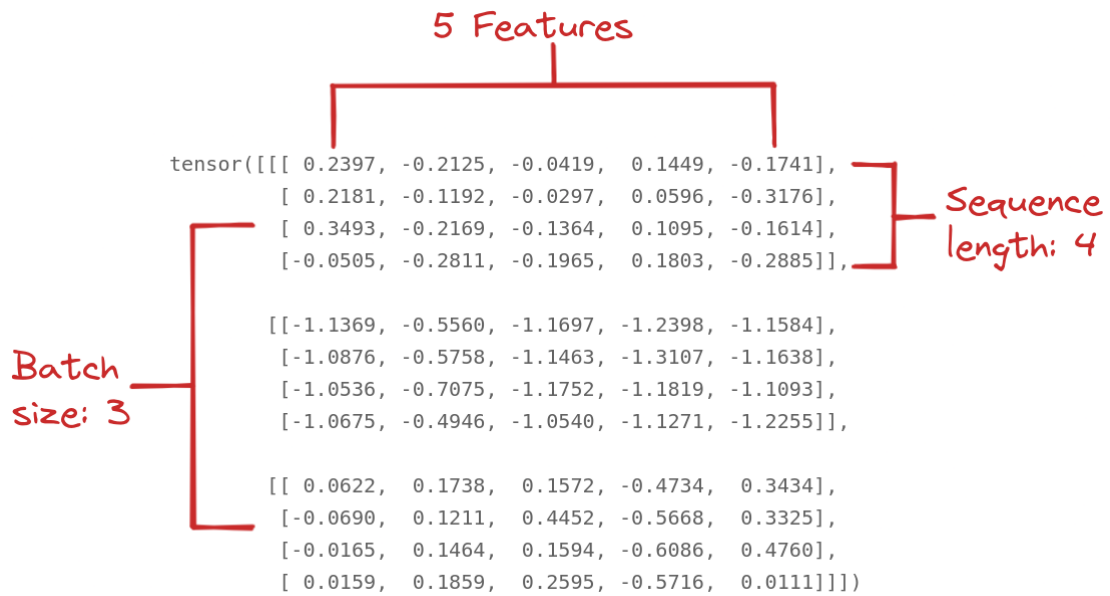
3.1. 模型参数

1. input_size: 输入数据的特征个数
2. hidden_size: hidden state的个数
3. num_layers: RNN网络的层数, 以下默认为1。
4. batch_first: 如果为true, 则输入数据的shape为(batch_size, seq_len, feature_num)。以下默认为true。

3.2. 输入

1. input

$\text{batch_size} * \text{seq_len} * \text{feature_num}$



2. h_0

$\text{num_layers} * \text{batch_size} * \text{hidden_size}$

3. c_0

$\text{num_layers} * \text{batch_size} * \text{hidden_size}$

3.3. 输出

1. output

$\text{batch_size} * \text{seq_len} * \text{hidden_size}$

2. h_n : 每个batch最后一个时刻的hidden state

$\text{num_layers} * \text{batch_size} * \text{hidden_size}$

3. c_n : 每个batch最后一个时刻的cell state

$\text{num_layers} * \text{batch_size} * \text{hidden_size}$

4. 参考文档

1. [LSTM — PyTorch 1.10.1 documentation](#)
2. [Understanding LSTM Networks -- colah's blog](#)