

# 第五周

## Day4-Day5

注意：回顾题一般为课本上的题目，需要用自己的话进行总结打卡。补充题则为课本上没有另行补充的题目，会提供一些参考答案。

### 1. 回顾：为什么RNN会出现梯度消失或梯度爆炸，分别的处理方式？

参照10.2节总结

### 2. 回顾：LSTM中有哪几个门，分别使用什么激活函数？

参照10.4节总结

### 3. 回顾：LSTM相比于传统的RNN解决了什么问题？

参照10.4节总结

### 4. 补充：LSTM为什么可以解决梯度消失的问题

LSTM 中梯度的传播有很多条路径，在一些路径上梯度流与RNN类似，照样会发生连乘产生梯度弥散，而有一些路径上梯度流非常稳定，比如：

$$c_{t-1} \rightarrow c_t = f_t \odot c_{t-1} + i_t \odot \hat{c}_t$$

这条路径上只有逐元素相乘和相加的操作。由于总的远距离梯度 = 各条路径的远距离梯度之和，即便其他远距离路径梯度消失了，只要保证有一条远距离路径梯度不消失，总的远距离梯度就不会消失（正常梯度 + 消失梯度 = 正常梯度）。因此 LSTM 通过改善一条路径上的梯度问题拯救了总体的远距离梯度。

### 5. 回顾：Seq2Seq模型引入注意力机制是为了解决什么问题

参照10.6节总结