Part1

1. 根据上述前向传播和结构图，推导RNN中的U，V，W的更新过程（W，U的更新仅需要给出第三时刻，即t=3时的更新）

因为损失函数具有累加性，故：



处于第t时刻的V的梯度为：



梯度：



更新：



处于第3时刻的W的梯度为：



处于第2时刻的W的梯度为：



处于第1时刻的W的梯度为：



梯度：



更新：



处于第3时刻的U的梯度为：



处于第2时刻的U的梯度为：



处于第1时刻的U的梯度为：



梯度：



更新：



1. 我们常用的激活函数有sigmoid(),tanh(),ReLU()三种函数，请证明三种函数应用在RNN中时，会出现那种梯度问题，并给出简要证明。

从上述图形中我们可以观察得到，在某个时刻的对W和U的偏导数，可以追溯到这个时刻之前所有时刻的信息。并且上述的损失函数是累加，也就是损失函数是前面所有时刻损失的求和。上述损失函数的偏导数的通式是：





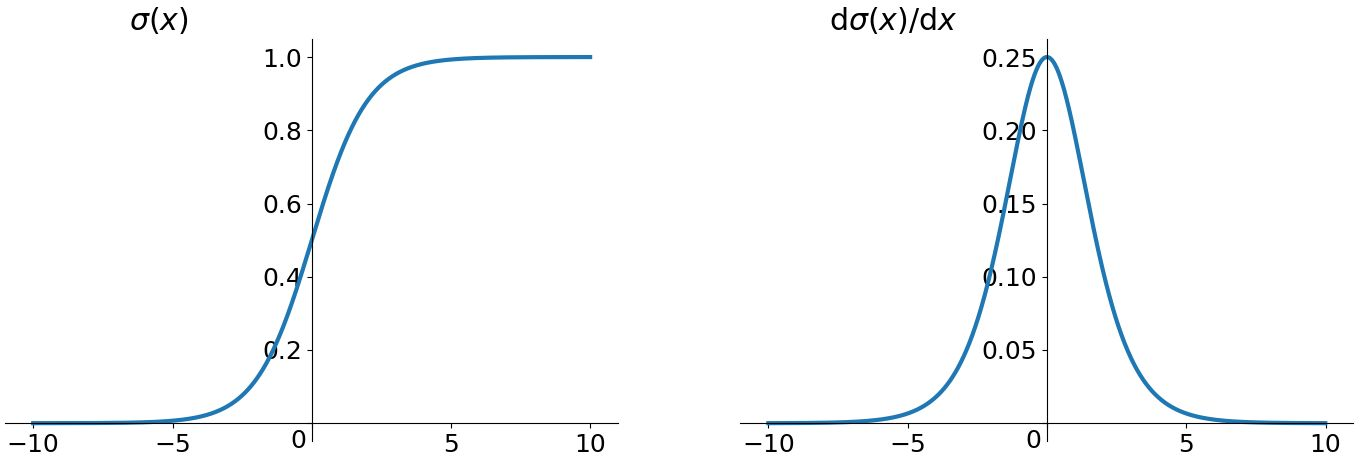
由于激活函数都是嵌套在里面的，我们将激活函数放进去，拿出中间累乘的部分：





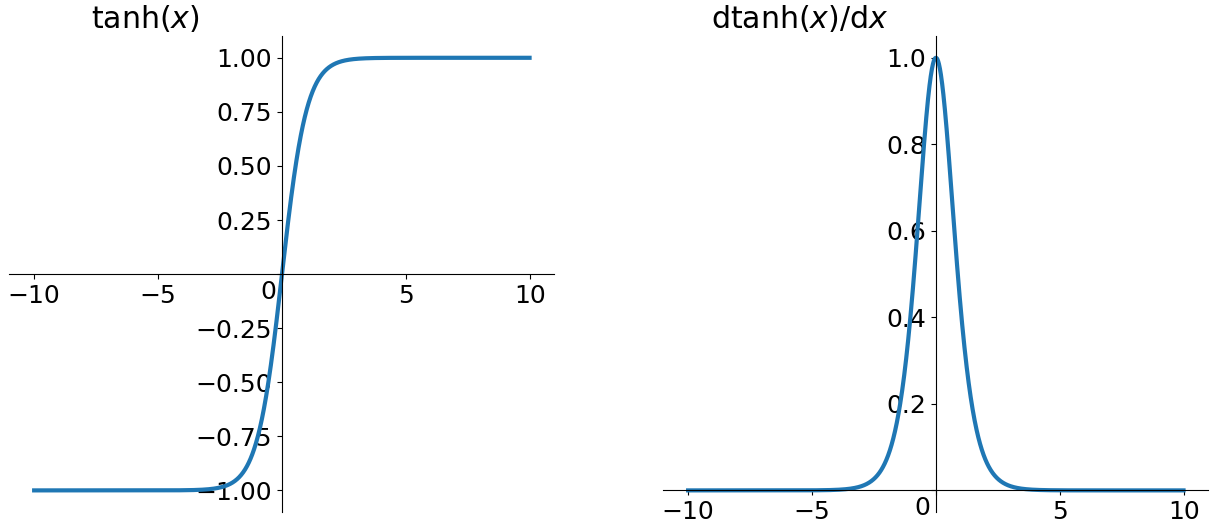


Sigmoid函数的函数图与导数图。



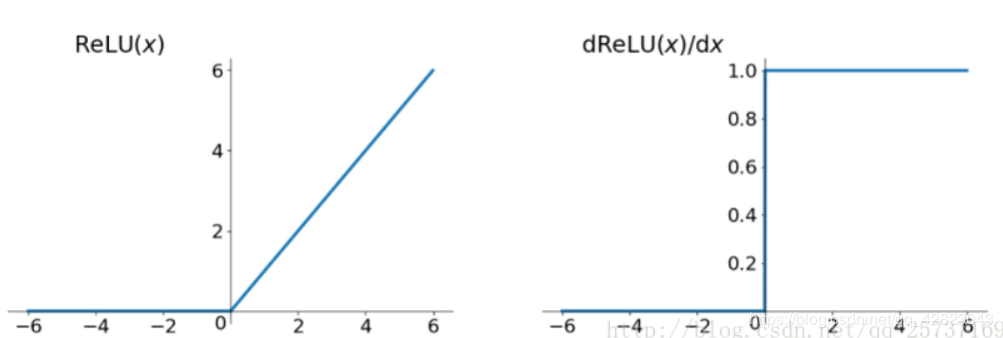
从图中可以看出，sigmoid函数的导数范围在(0,0.25]，导数最大值小于1。那么随着时间序列的不断深入，小数的累乘会使得梯度越来越小，这就会出现梯度消失现象。

Tanh函数的函数图与导数图。



从图中可以看出，tanh函数的导数范围在(0,1]，导数最大值为1，但是也不是时刻为1，那么随着时间序列的不断深入，小数的累乘会使得梯度越来越小，这就会出现梯度消失现象。

ReLU函数的函数图与导数图。



从图中可以看出，ReLU函数的左侧导数为0，右侧导数恒为1，各时刻损失函数梯度的累加会出现梯度爆炸现象。但是因为左侧恒为0，又导致神经元消失。