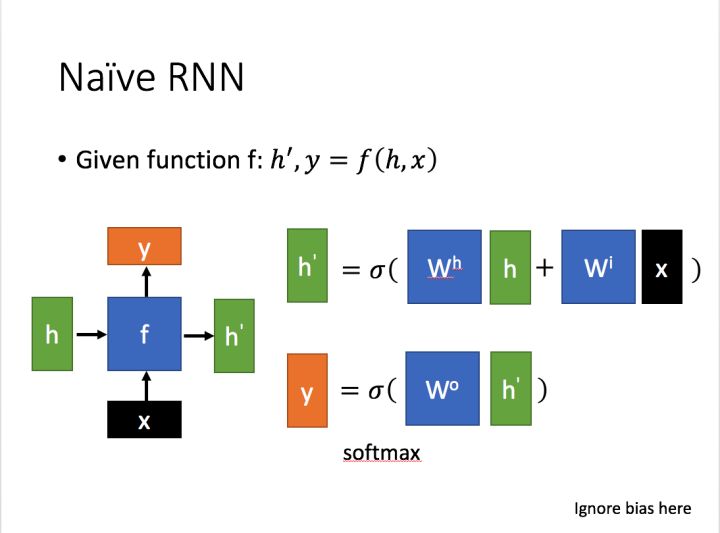
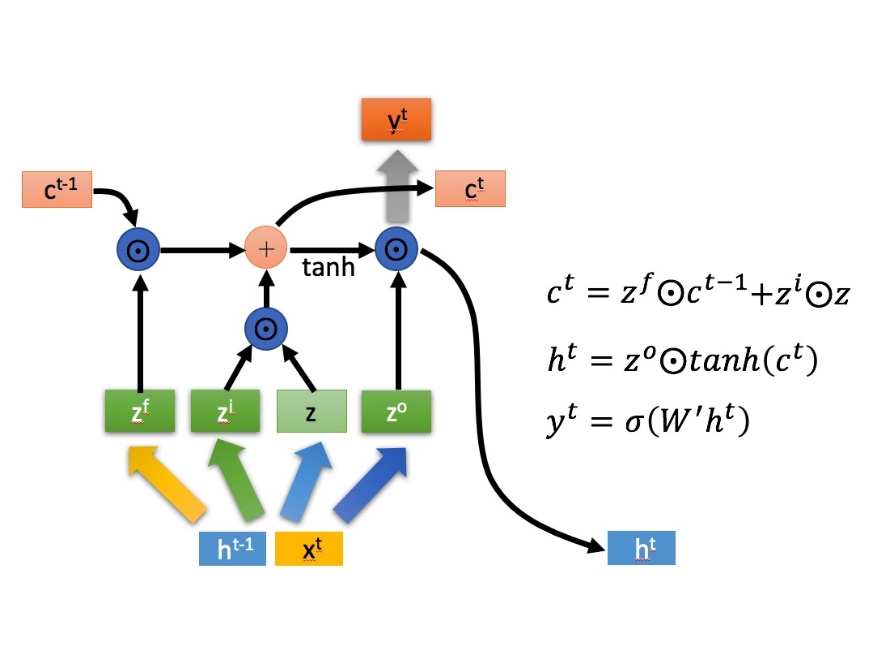
# 常见网络结构

## RNN



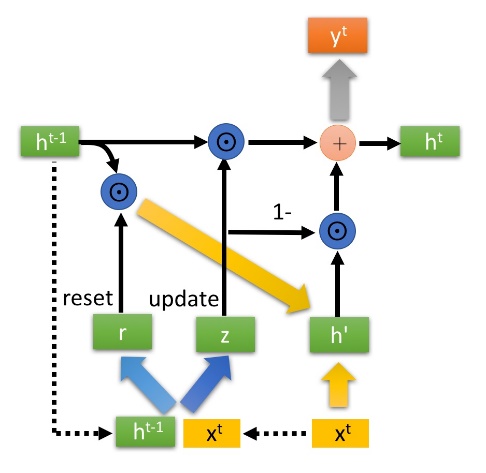
1. 是激活函数

## LSTM



1. 先根据上一个状态和输入计算各种门和新输入
   1. 遗忘门：
   2. 输入门：
   3. 输出门：
   4. 新输入：
2. 更新
3. 更新
4. 更新

## GRU



1. 类似LSTM，先算门
   1. 重置门：
   2. 更新门：
2. 计算
3. 更新

## LSTM和GRU的区别

1. LSTM拥有两个数据通道(流)，一个是不断累积的信息C，一个是每个循环的输出H，H根据当前信息的累积情况C产生
2. LSTM使用遗忘门和输入门控制信息的累积，使用输出门产生当前的输出
3. GRU只有一条通道，参数更少

# 梯度消失/爆炸问题

## RNN梯度推导

1. 假设输入为，标签为
   1. 标签很多情况下只有一个怎么办？只算最终循环的梯度吗？？
2. 则各参数梯度如下，最终的梯度是每个时间维度(循环)上梯度的累积
3. 以为例进行推导，因为
   1. 因为，所以
4. 一般小于1，所以受影响，很容易梯度爆炸()或者梯度消失()
5. 对的求导则没有这个问题

## LSTM梯度推导

此时

而

此时连乘部分从参数变成了激活函数的取值，在(0~1)之间，实际参数更新时，可以通过调控参数()使得激活函数输出的值接近1，这样连乘就不会梯度消失或者爆炸了。