# 基础知识点

## 梯度下降

https://www.jianshu.com/p/c7e642877b0e

## 反向传播

https://www.jianshu.com/p/964345dddb70

## 梯度不稳定性和梯度饱和

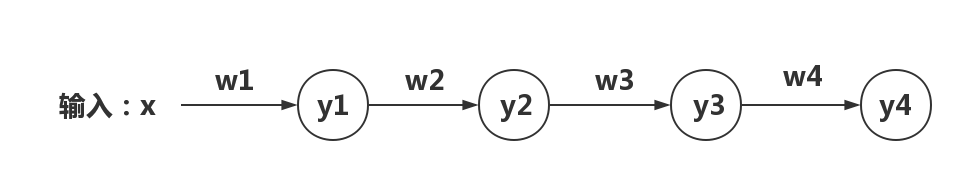
<https://blog.csdn.net/qq_17130909/article/details/80582226>

<https://blog.csdn.net/u014595019/article/details/52562159>

# 线性反向传播

## 反向传播梯度计算

如图，输入x经过4个神经元的计算结果分别是，也是最终的计算结果。



最终损失, , ,,

所以的梯度：

的梯度：

…

的梯度：

## 非线性反向传播

反向传播中的神经元：

输入：

输出：

参数：b, W

激活函数：f(x)

误差： , g的初值为，输出和标签之间的损失函数值

也就是最终的损失J关于的梯度

激活函数梯度 =

偏置梯度：

权重梯度：

前一层的误差：

求得梯度之后就可以更新参数了，比如参数W

## 线性反向传播

在计算梯度时忽略激活函数的梯度

偏置梯度：

权重梯度：

前一层的误差：

## 优点

1. 计算量低
2. 避免了激活函数的梯度饱和导致的神经元饱和，无法学习的情况