# 基础知识点

## 梯度下降

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

公式：



求导的目的是为了获取参数更新方向，学习率是为了调整每次更新的步长。

## 反向传播

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

## 梯度不稳定性和梯度饱和

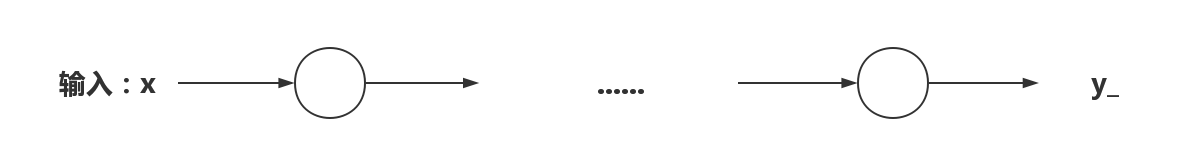
<https://blog.csdn.net/qq_17130909/article/details/80582226>

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

# 反向传播的计算方式

## 神经元参数说明

如图是一个具有个神经元的简单神经网络



记

对于第个神经元，

输入：

输出：

参数：

激活函数：

## 梯度计算

记,表示最终损失对于的偏导，其中

第个神经元梯度计算：

第个神经元梯度计算：

…

## 参数更新

对于第k个神经元：

# 线性反向传播

## 非线性反向传播

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