目录

[研究 1](#_Toc22067)

[什么是拟合？ 2](#_Toc14691)

[一个完整神经元一般长什么样？ 2](#_Toc25934)

[激活函数 3](#_Toc12744)

[导数的作用 5](#_Toc10650)

[快速理解导数在神经网络上的意义 5](#_Toc148)

[简单的导数以及链式求导 7](#_Toc32197)

[链式求导的多链路概念 8](#_Toc1927)

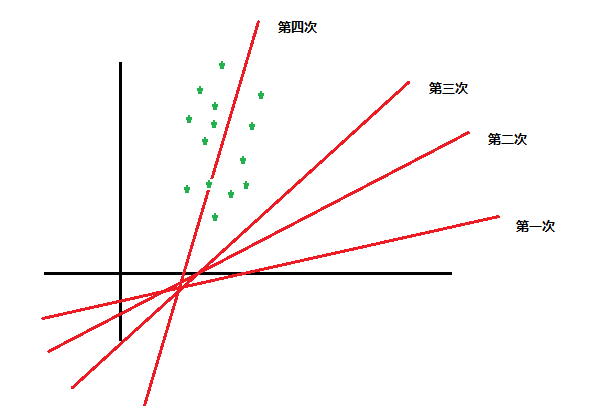
[误差函数 8](#_Toc3887)

[本地知识库实现的相关知识 10](#_Toc20121)

# **研究**

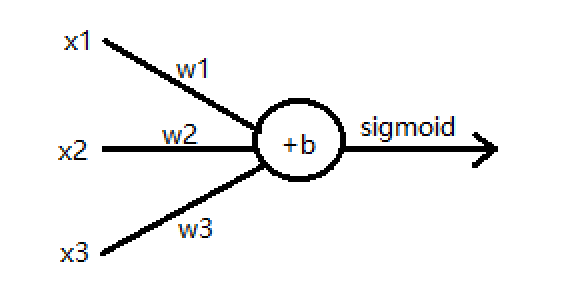
https://github.com/huggingface/transformers

# **什么是拟合？**

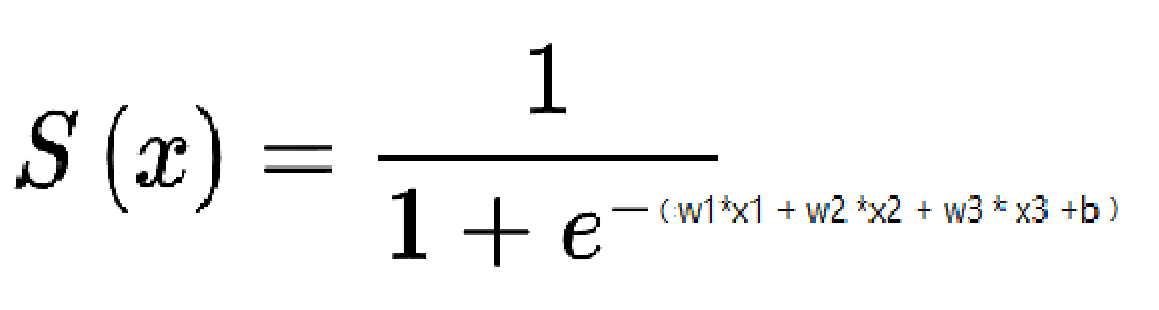


# 一个完整神经元一般长什么样？

图像表达为



数学表达为



# 激活函数

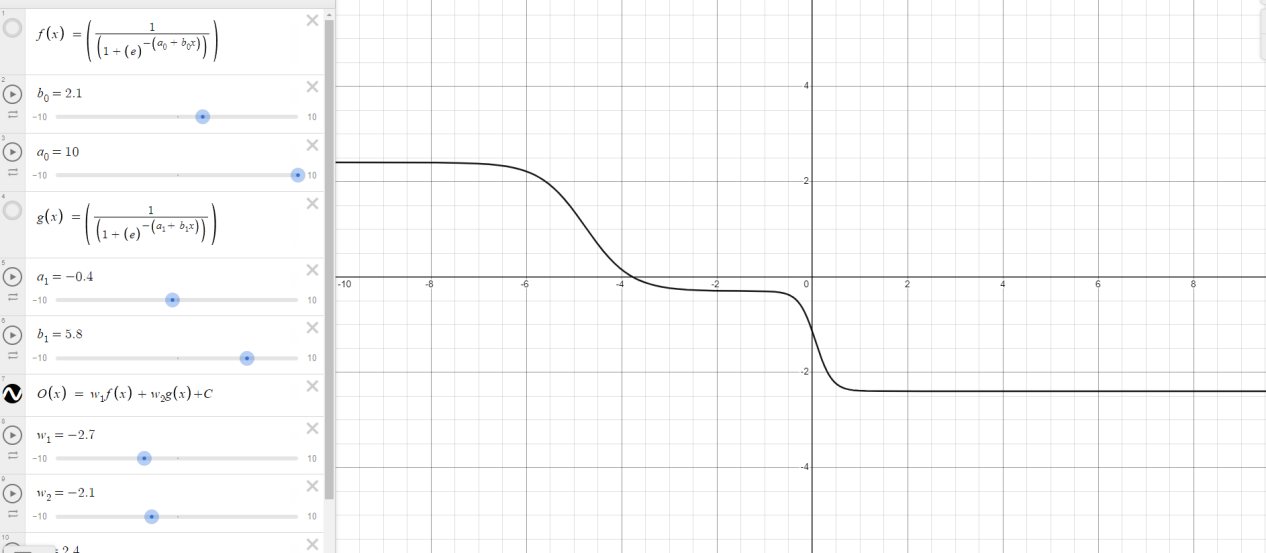
不同的激活函数组合，可以拟合不同的曲线

<https://www.desmos.com/calculator?lang=zh-CN>

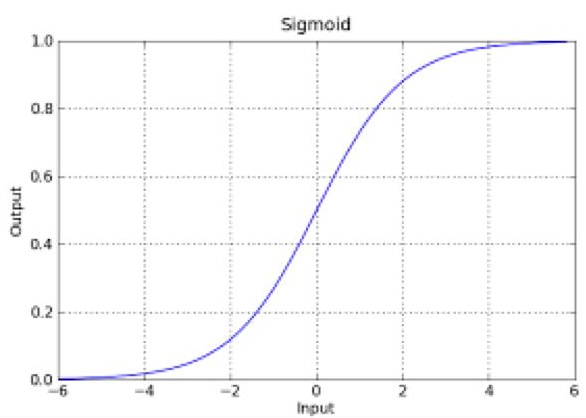
f\left(x\right)\ =\left(\frac{1}{\left(1+\left(e\right)^{-\left(a\_{0}\ +\ b\_{0}x\right)}\right)}\right)

g\left(x\right)\ =\left(\frac{1}{\left(1+\left(e\right)^{-\left(a\_{1}+\ b\_{1}x\right)}\right)}\right)

O\left(x\right)\ =\ w\_{1}f\left(x\right)\ +\ w\_{2}g\left(x\right)+C



有很多种激活函数，我们以常用的sigmoid函数举例

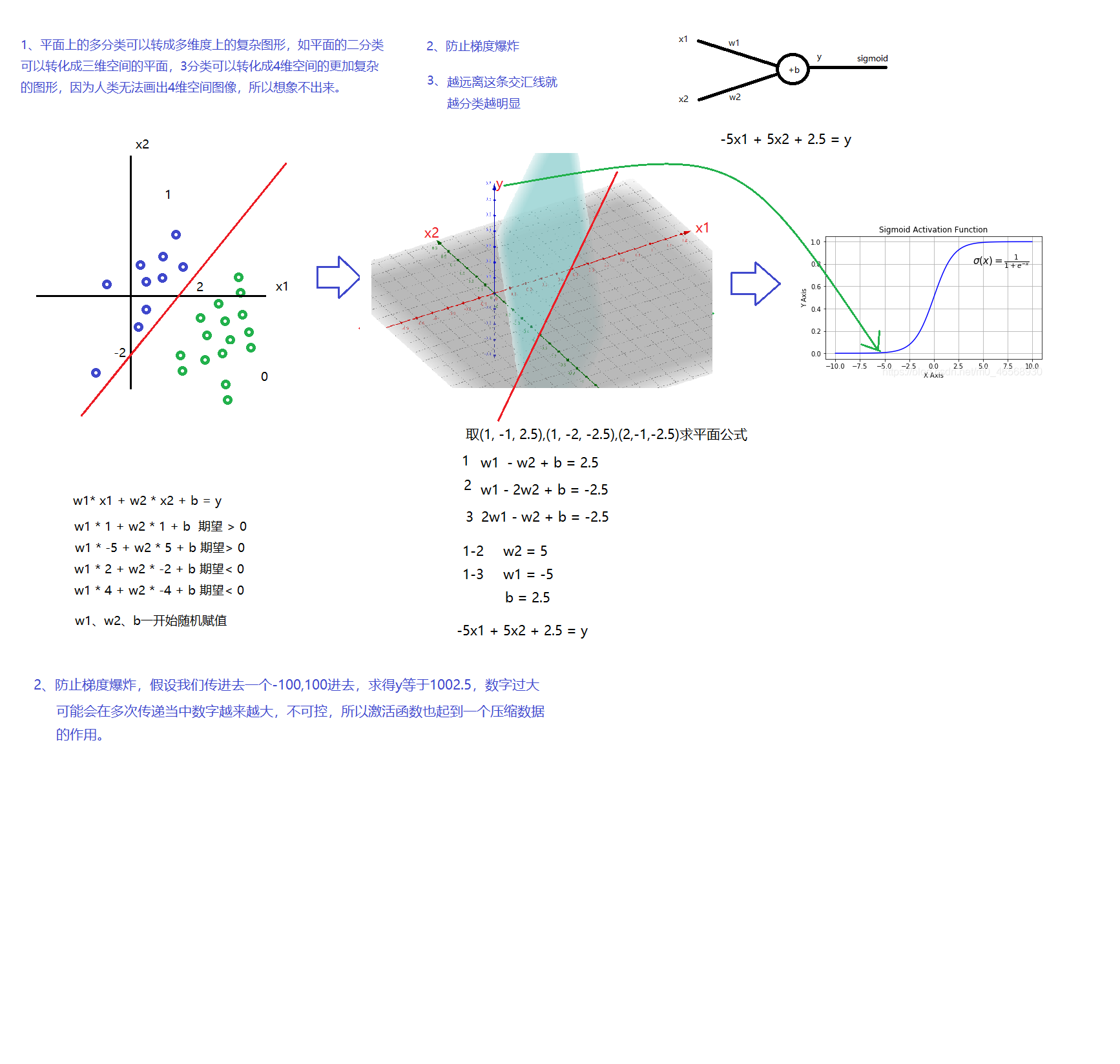


激活函数一般具有一个特征就是突变，还有趋向于一个稳定的值。

激活函数的几何意义，以下我们举例一个二分类来了解激活函数的几何意义。

<https://www.geogebra.org/download>

f: z=-5 x+5 y+2.5



# 导数的作用

我们去调整神经网络的权重，一般不会手动去调整，如果只有很少的神经元，人工调整确实可以实现，当我们有几十层，一层几百上千个神经元的时候，人工调整就不可能了。

一个权重的调整涉及到两个问题，一个是调大还是调小，一个是调整的幅度。

快速理解导数在神经网络上的意义

1. 导数是什么，就是求解某个变量对于结果的影响程度，当变量多的时候，我们就把其他变量当成常数，在求导数的过程中把它略掉。如下

y = 2x1 + 3x2 + 10x3

x1对y的影响程度是2

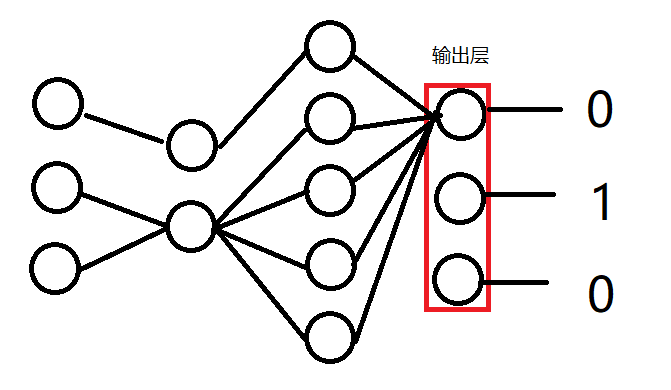
X2对y的影响程度是3

X3对y的影响程度是10

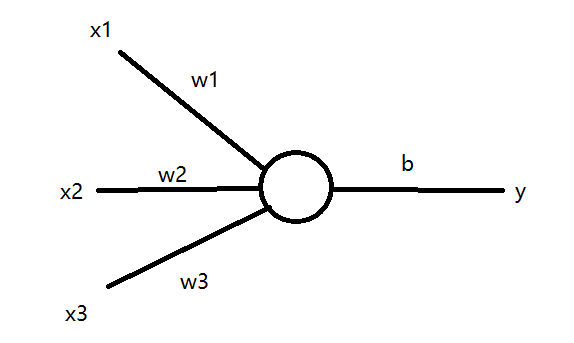
然后调整权重，就是调整2,3,10这些数字，如果数字过大，我们就适当调小一点，例如把10调整成5，越大的数字调整的幅度就应该越大。

凡是变量x存在的，都是求某一个点的变化率

恰好导数的这种意义可以应用到神经网络里面，因为需要求解每一个神经元节点对于最终结果的影响。



下面的输出值y,公式为y = w1\*x1 + w2\*x2 + w3\*x3 + b,我们要求解w1，w2，w3，b对于结果的影响，这就是为什么我们需要用到数学上的导数的原因。



# 简单的导数以及链式求导

导数的简单公式计算，瞬时变化率就是导数。

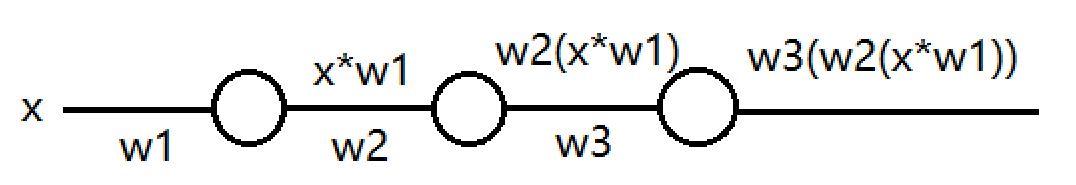
y=3x ,导数是3,是一条直线，所以处处的变化率是一样的。

y=3x^2,导数是3x,x取某一个值，因为x的平方是曲线，不同位置的变化率是不一样。

y=3x^3,导数是3x^2,x取某一个值，因为x的平方是曲线，不同位置的变化率是不一样。

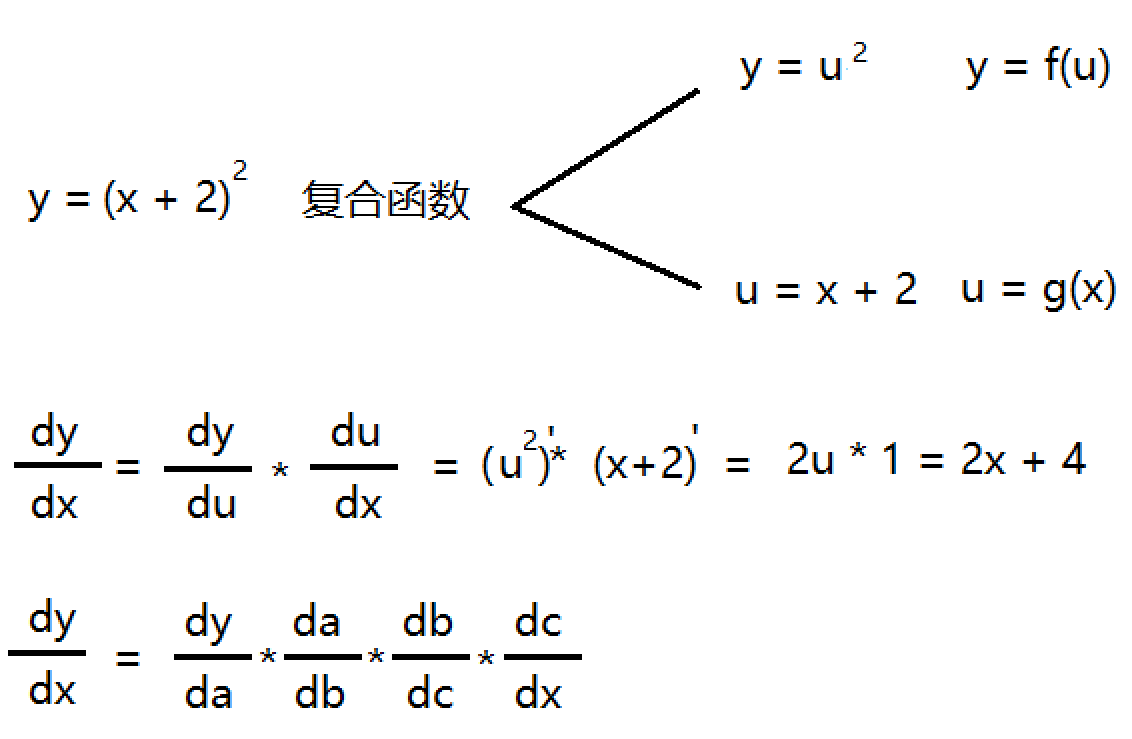
导数的链式求导，为什么神经网络会引入链式求导？首先我们必须知道链式求导是什么？

如下图，我们输入一个x经过多重神经元，会变成这样，这就是一个复合函数：



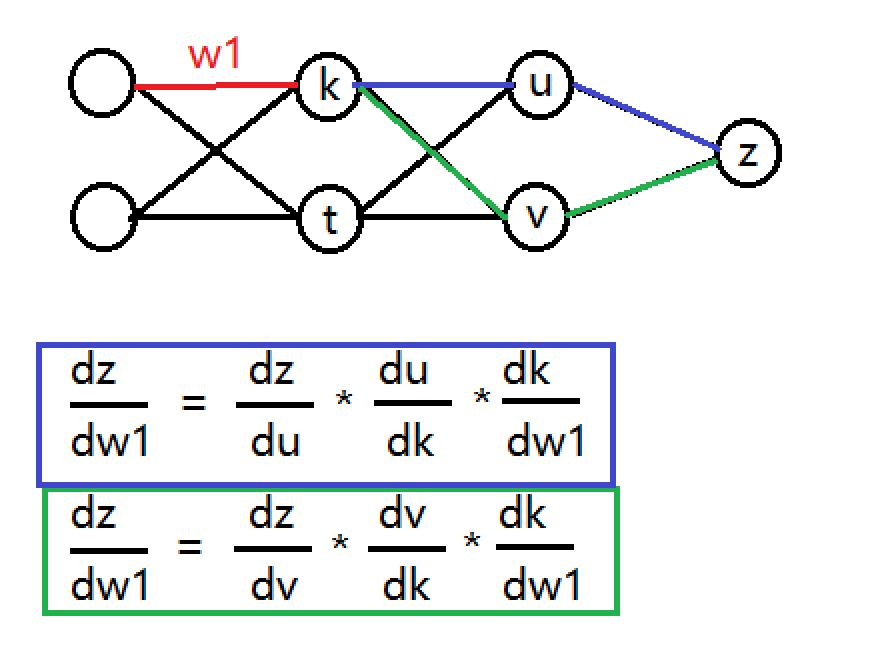
复合函数需要使用链式求导法则进行求解

最简单的链式求导



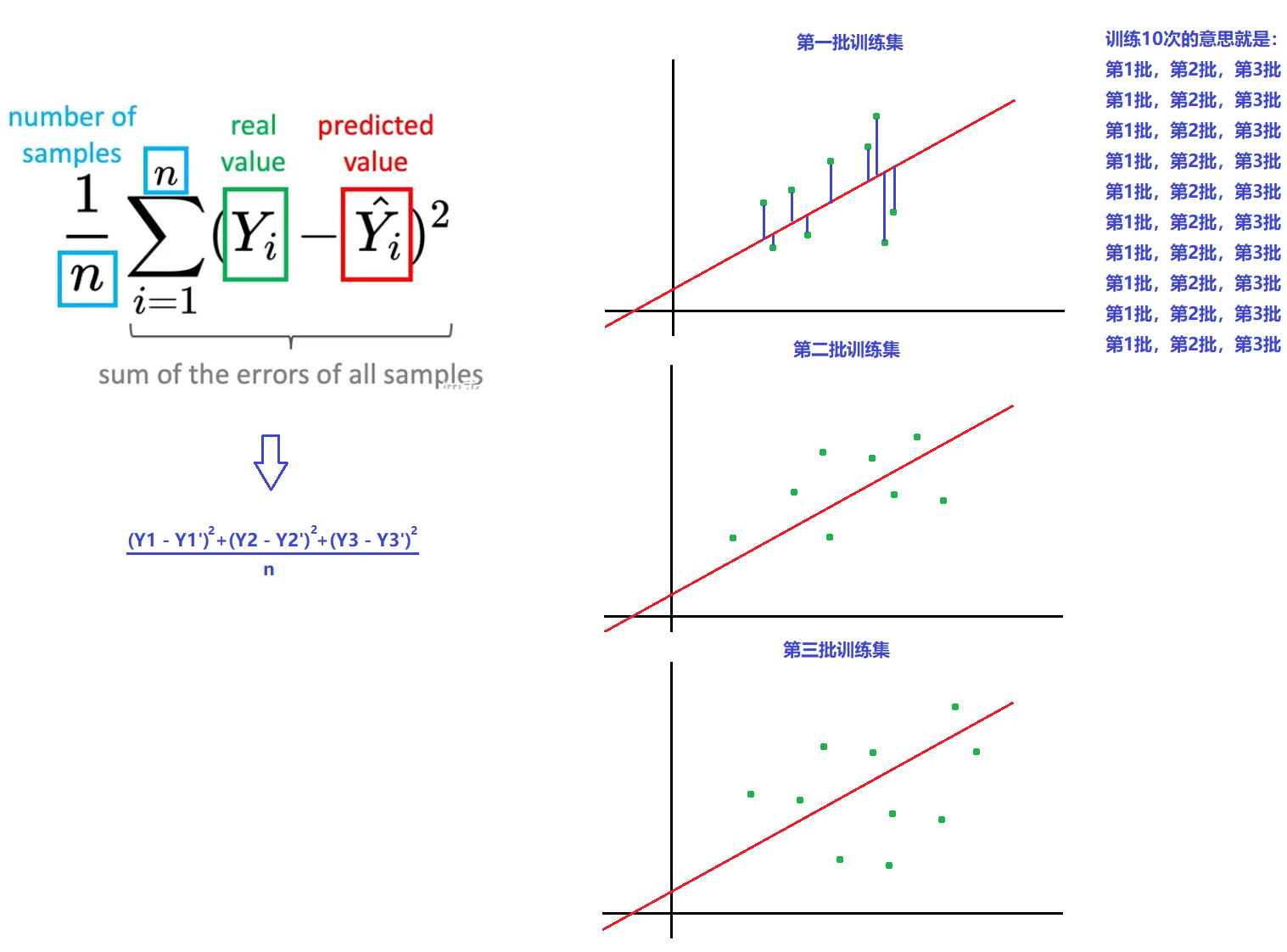
# 链式求导的多链路概念

我们求解z对于w1的导数的时候，有很多链路可以求，一个可靠的方法是将所有的链路求解出来，全部相加起来，就得出了w1的导数。



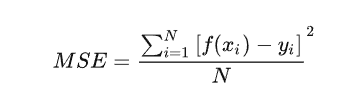
# 误差函数

下图是均方误差

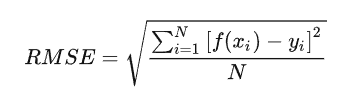


每种误差函数的优点：

均方误差：对误差进行平方，取平方有一个特性，它惩罚更大的错误更多，但是同时异常值对结果的影响也很大。



均方根误差：RMSE 的优点是对较大误差值有较大的惩罚，因为它对差异值进行了平方操作。这可以避免较大误差值对拟合度的影响过大。



不同维度，这个表情会有不同的概念

