# 机器学习

*所谓机器学习,就是在大量数据的运行下,使得计算机可以进行归纳,预测*

机器学习分为三类:监督学习,无监督学习,强化学习

抛开强化学习不讲,这里的监督学习与无监督学习的根本区别在于:有无数据的标记(即y值)

我们将输入的数据称之为数据的特征,一组特征为一个样本,需要求得的结果为标签

例:

1. 有一组数据,格式为(身高,体重,BMI)

这里使用监督学习的效果为: 输入许多样本的特征值(身高,体重)与对应的标签(BMI),让计算机得出一个可以得出 Y=AX+b 这样的函数(称之为模型),然后对后续的样本值,可以根据模型计算得出标记

1. 一组数据,格式为(身高,体重) 即无y值

无监督学习: 根据身高,体重将他们聚类,如高瘦的一组,低矮的一组

对监督学习与非监督有一个大致的概念了,下面开始详细介绍监督学习:

## 监督学习

监督学习分为:分类,回归

分类,就是指y为离散数据, 如通过 特征(尺寸,logo)判别 标签(手机品牌)

回归则是指y为连续数据,如上述通过 特征(身高,体重)判别 标签(bmi)

### 线性回归

以上面的身高,体重 BMI为例

(但是由于三维图不方便绘画,且认为BMI与身高相关,由二维开始理解原理)

如图1-1所示,我们可以很明显的看出来有一条虚线可以拟合出来身高与BMI的关系,但是如何根据样本数据求得该直线

下面说一下机器学习的步骤:

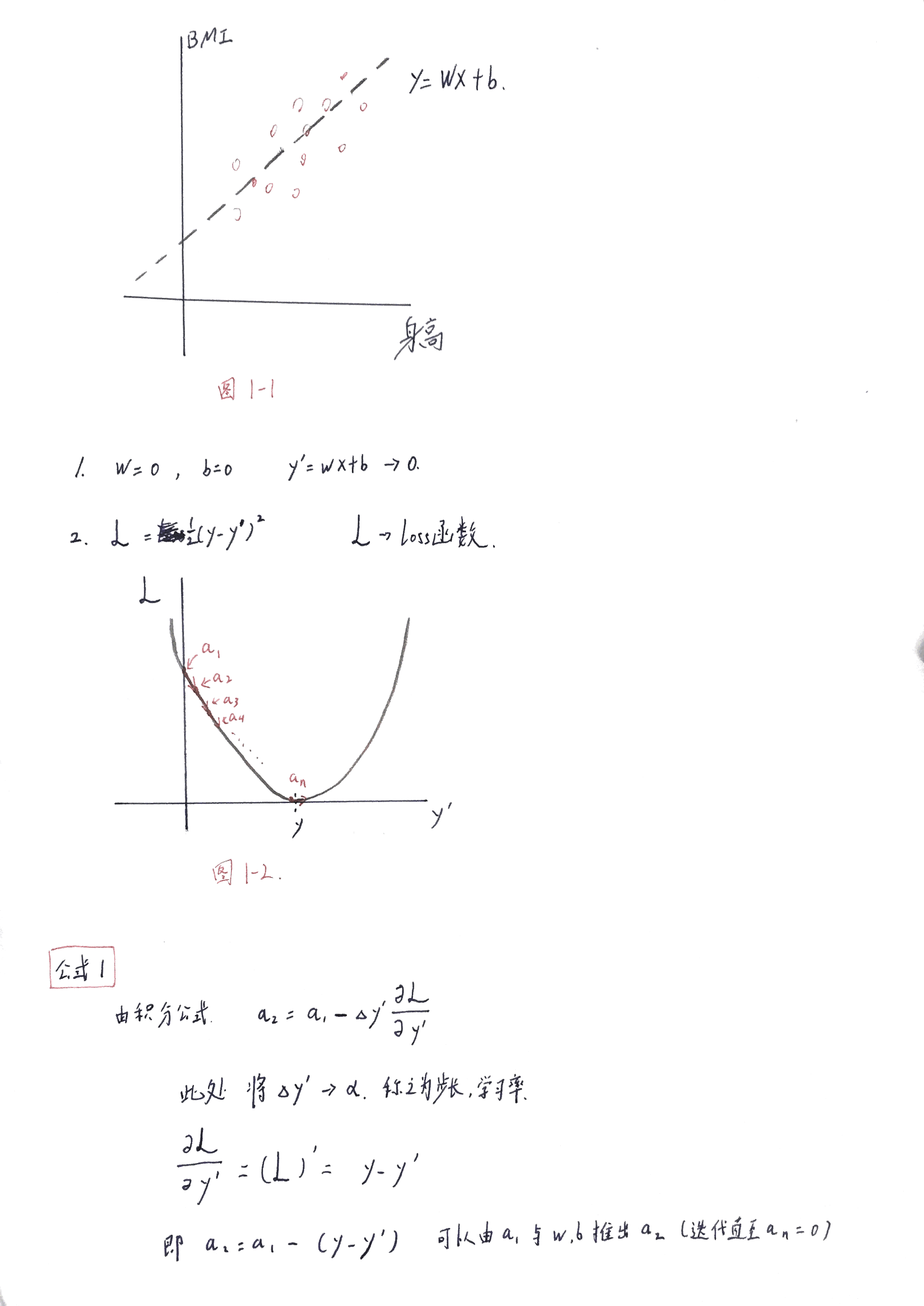
W指权重,b指偏差值,y1为预测的输出值,y为真实值 y1 = wx+b

1. 我们不妨设w,b初始都为0,此时求得的所有y1也为0
2. Loss函数(损失函数)指的是**预测得出的y1与真实的y的平方差**,(此处加上1/2是为了计算方便),由图1-2可以看出,loss函数为凸函数,在导数(即斜率)为0时,loss函数值最小,即此时的w与b最佳,但由于很难做到预测的y1完全等于y,所以在此处我们需要得出可以使得(y1-y)2最小的w与b

我们在开始y1=0时,位于a1点,由图1-2可以得出,始终沿着斜率下降的方向,我们总会走到最低点附近,此时的w与b满足loss函数最小,则y1=wx+b为我们所求模型

具体请看公式1

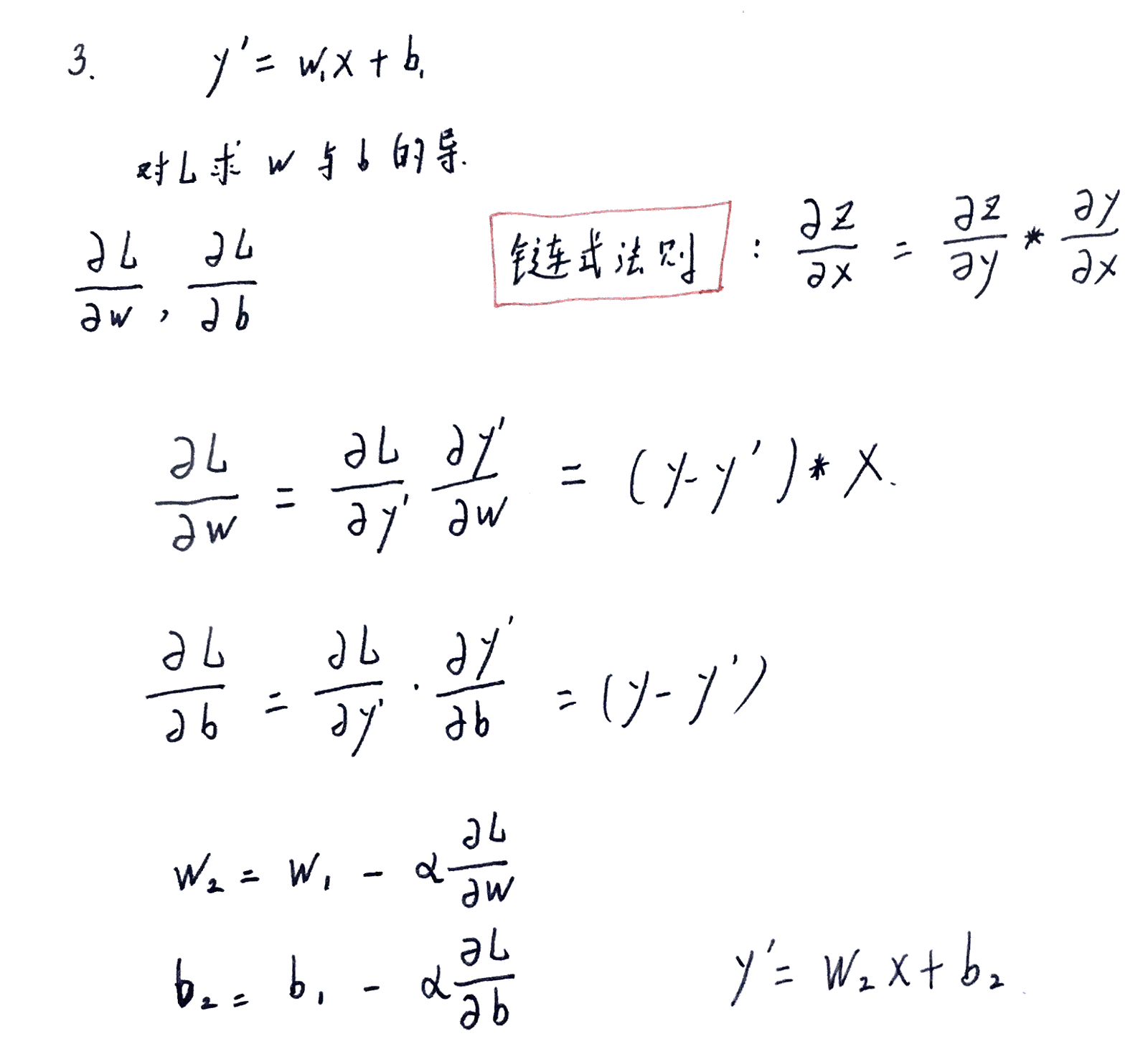
根据斜率逐渐趋近于0来使得loss损失函数最小的方式称之为**梯度下降(Gradient Descent)**



1. 对于y1,y1=wx+b, 因此上面的损失函数对y1求导可以解析为L对w,b求导

又根据链式法则(图中公式),可以求出此处L对w与b的导数,然后根据求得结果重新定义w,b作为y1的模型

根据先定义的w,b求得y1,再根据y1反推w,b并重构造,称作**反向传播**



1. 接下来就很好理解了,我们将样本不断输入,使得w,b不断地重构造,最终会得出结果使得Loss函数位于最低点附近,此时的模型就符合我们的需求

对于多个特征值来说,y1=w0x0+w1x1+w2x2…+b 同理

### 逻辑回归

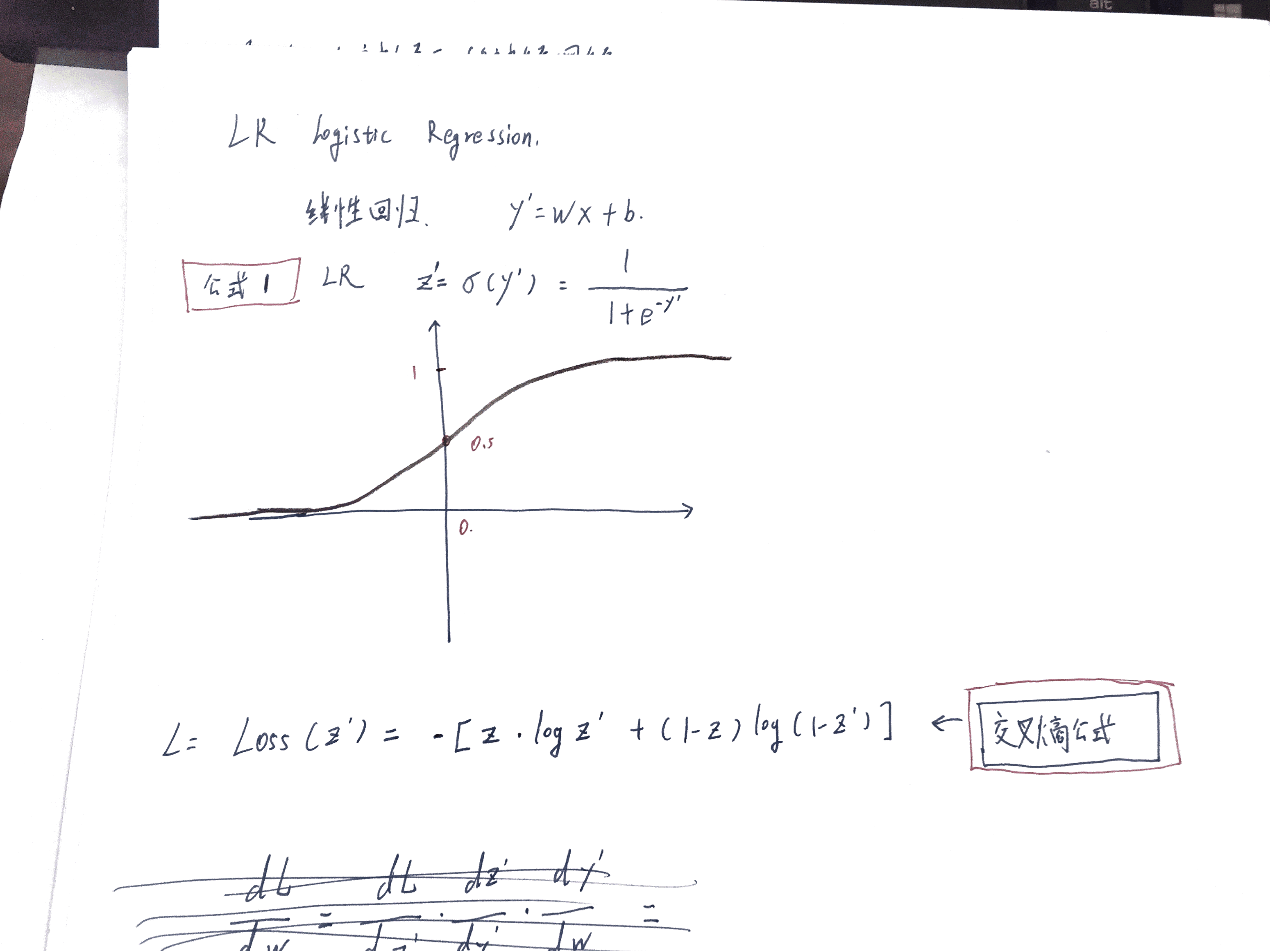
线性回归只可以拟合普通的线性函数,但是对于不满足线性条件无法做到,因此我们引入逻辑回归,构造非线性函数

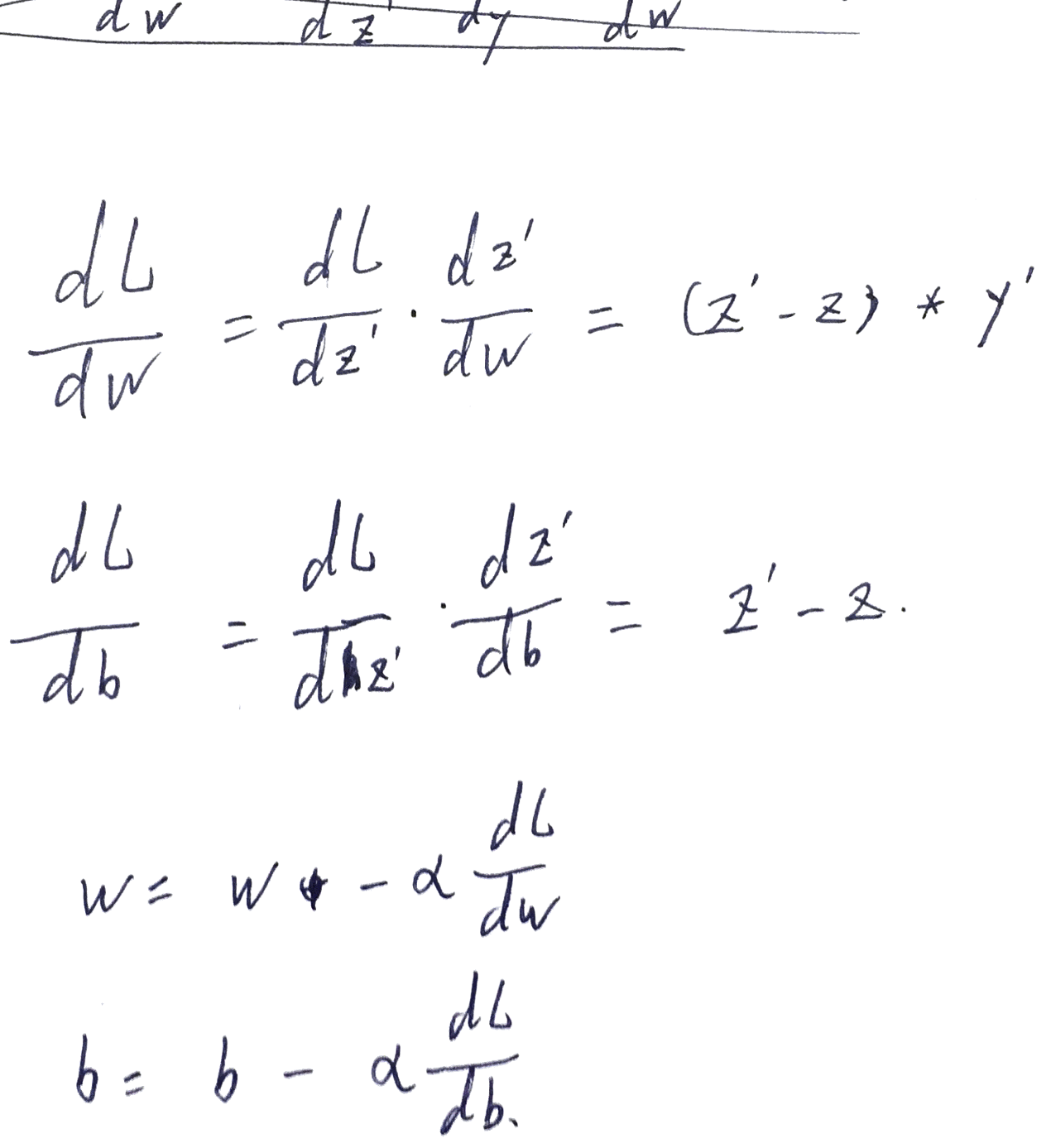
逻辑回归公式,如图公式1(在线性回归基础上做了一个转换)

在逻辑回归中,损失函数也与线性函数不同,使用交叉熵公式

同线性回归所论,对其进行**梯度下降**与**反向传播,**求出使得损失函数最小的w,b

逻辑回归是非线性回归





# 神经网络

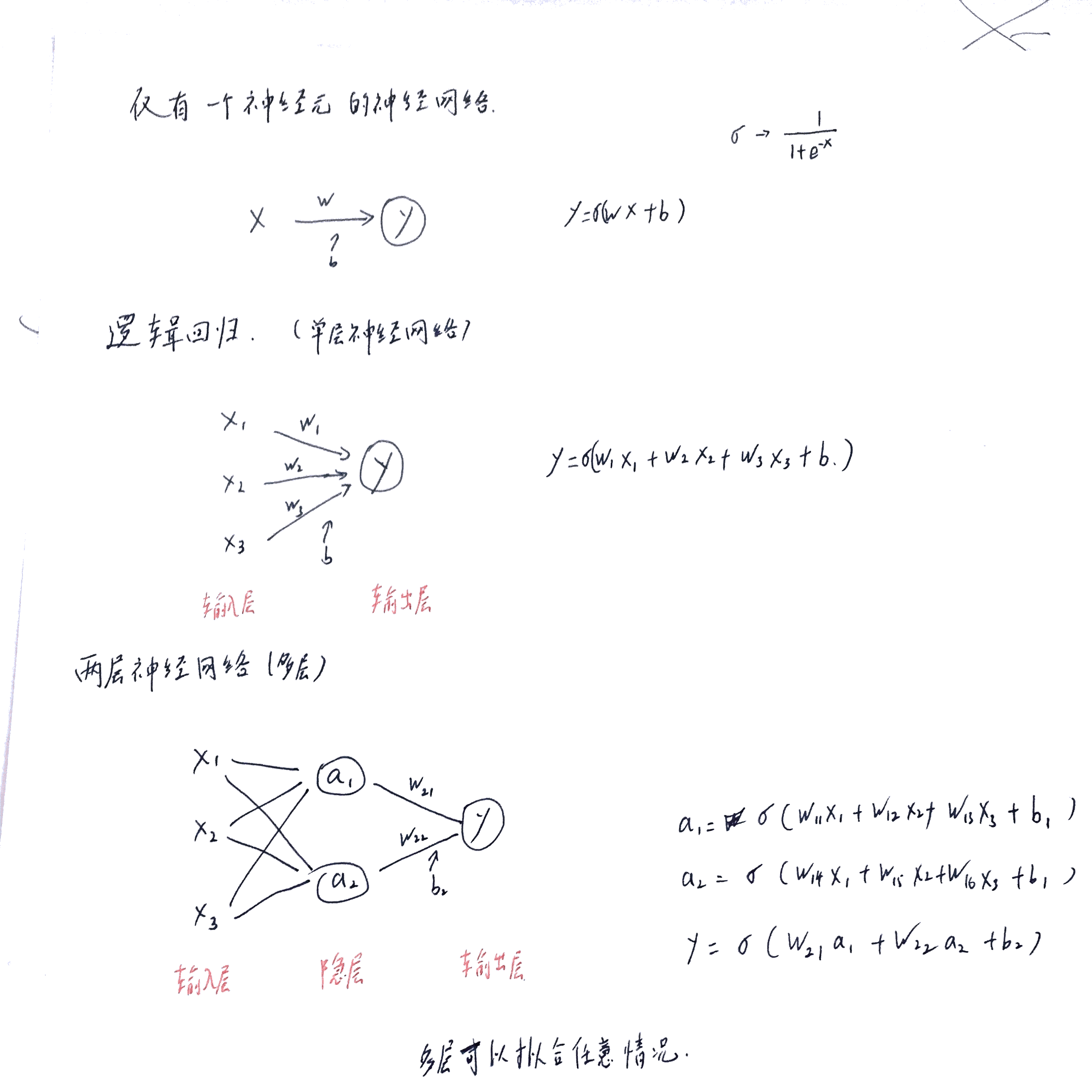
首先需要知道几个知识:

神经网络包含:**输入层,隐层,输出层**(对应输入样本特征,不可见的计算过程,输出的标签)

逻辑回归可以看做是最简单的神经网络

在计算神经网络层数时,往往不考虑输入层

逻辑回归LR于简单神经网络NN区别如图



神经网络的原理同逻辑回归 (梯度下降,反向传播)

神经网络,我个人理解 就是一个**找出最佳拟合函数模型的实现方式,**根据样本的持续输入,由误差函数反向传播,推倒出w,b,训练出合适的模型

神经网络NN分为

深层网络 DNN 多个隐层的NN

卷积神经网络 CNN 图像识别

循环神经网络 RNN 自然语言处理,语义识别,时间序列预测

混合神经网络

### RNN 循环神经网络