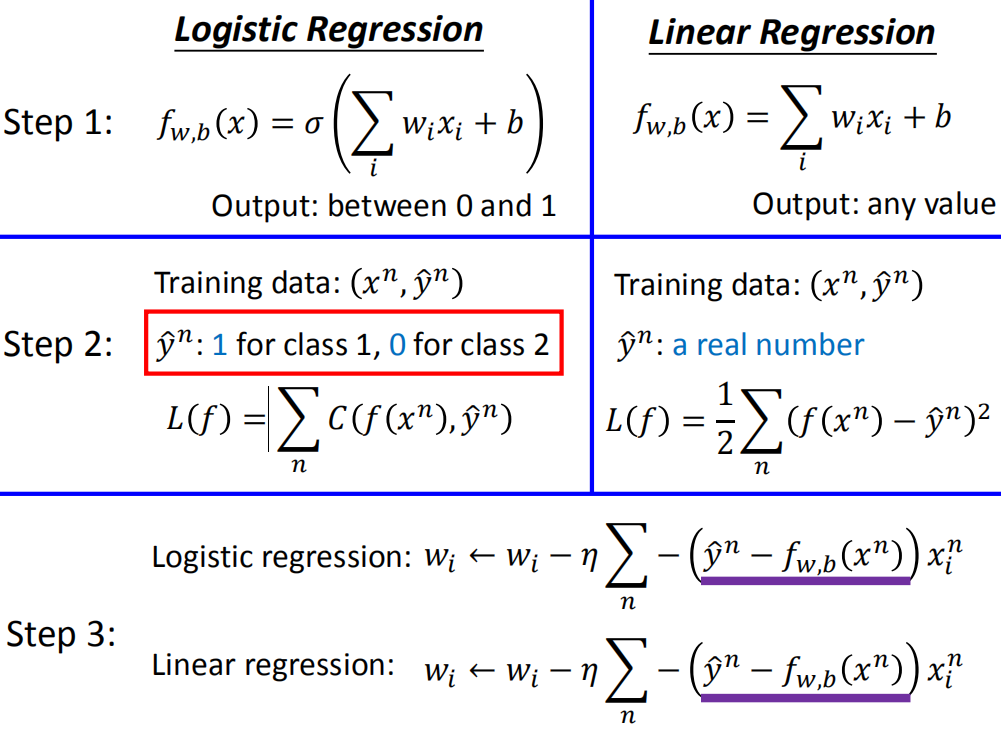
P10-概率分类模型

## 逻辑回归和线性回归的对比

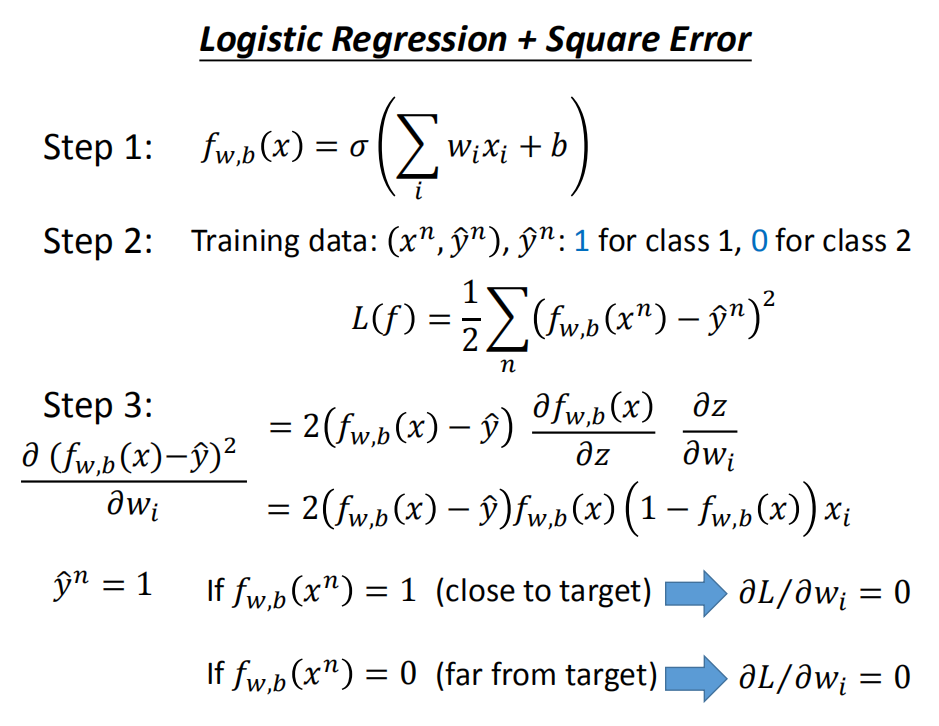
首先如果不学这一章，知道逻辑回顾是分类模型，而线性回归是回归模型



Step1也就是模型或者称为函数集合上来看，逻辑回归加上了sigmoid，step2两者的损失函数完全不同，而在step3利用梯度下降求解参数的时候，可见两者是几乎相同的，但其实因为模型的函数形式不同，事实上他们的梯度下降经过化简之后也是不同的

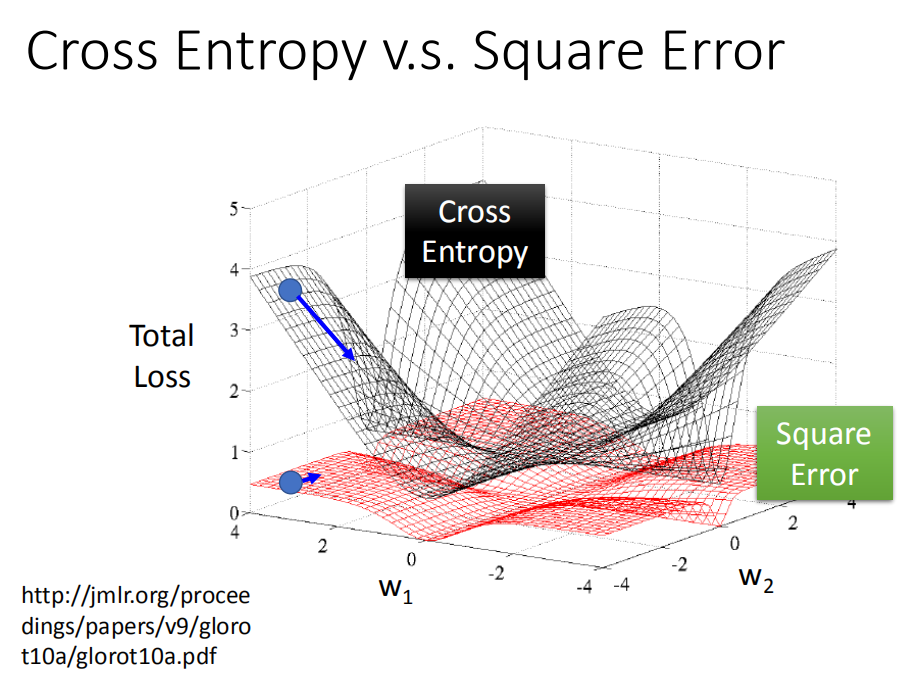
## 损失函数：为什么不学线性回归用平方误差？

其实一直就是认为MSE只是回归问题的损失函数，不可能作为分类损失函数，那么如果真的用了会怎么样呢？



也就是说如果在分类问题中用MSE作为损失函数，不管是接近最优点还是远离最优点，其实他们的梯度都是0

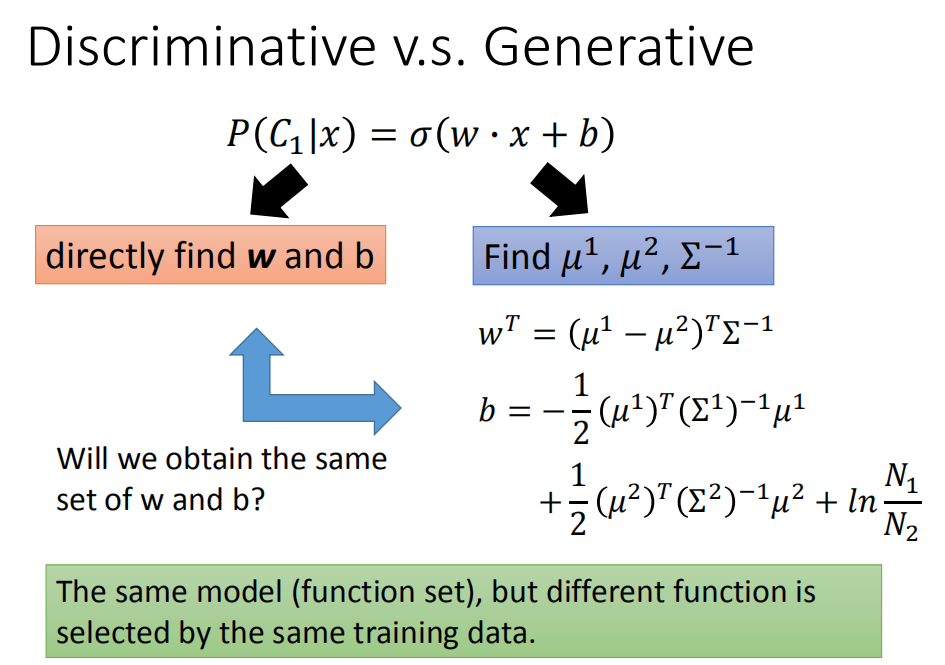
可以通过对比损失函数来看：



## 判别模型v.s. 生成模型

### 3.1 基本原理上的不同

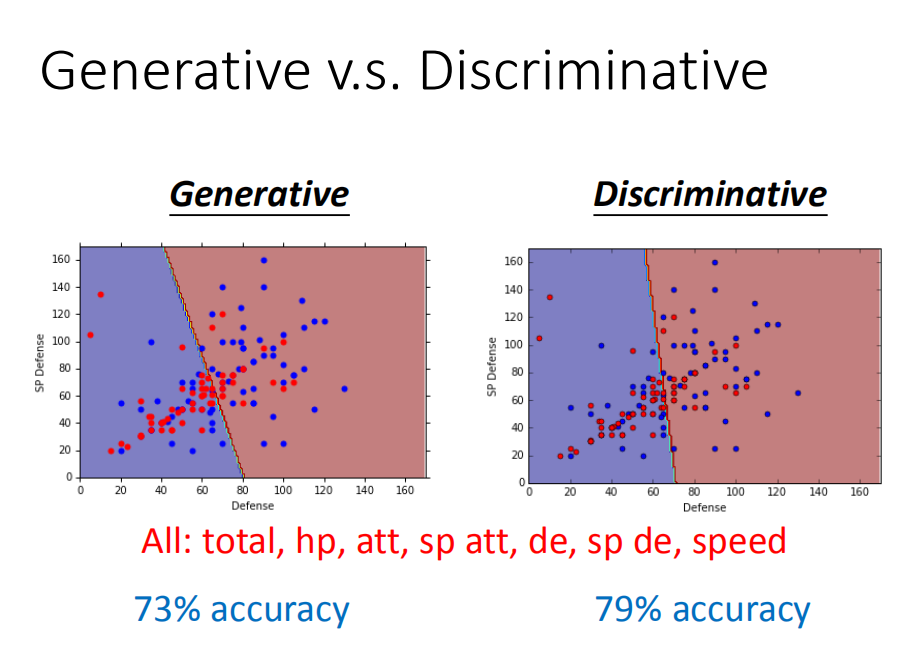
逻辑回归的方法称为Discriminative（判别） 方法；上一篇中用高斯来描述后验概率，称为 Generative（生成） 方法。它们的函数集都是一样的：



如果是逻辑回归，就可以直接用梯度下降法找出w和b；如果是概率生成模型，像上篇那样求出 μ^1,μ^2μ1,μ2 ，协方差矩阵的逆，然后就能算出w和b。

用逻辑回归和概率生成模型找出来的w和b是不一样的。

### 3.2 效果上的差异（一般判别比生成模型好，因为生成模型有很多的假设）



### 3.3 判别方法不一定比生成方法好

生成方法的优势：

（1）训练集数据量很小的情况；因为判别方法没有做任何假设，就是看着训练集来计算，训练集数量越来越大的时候，error会越小。而生成方法会自己脑补，受到数据量的影响比较小。因为生成模型有时候会忽略数据，基于自己的假设来考虑问题，那么对于数据的要求就会低一点。

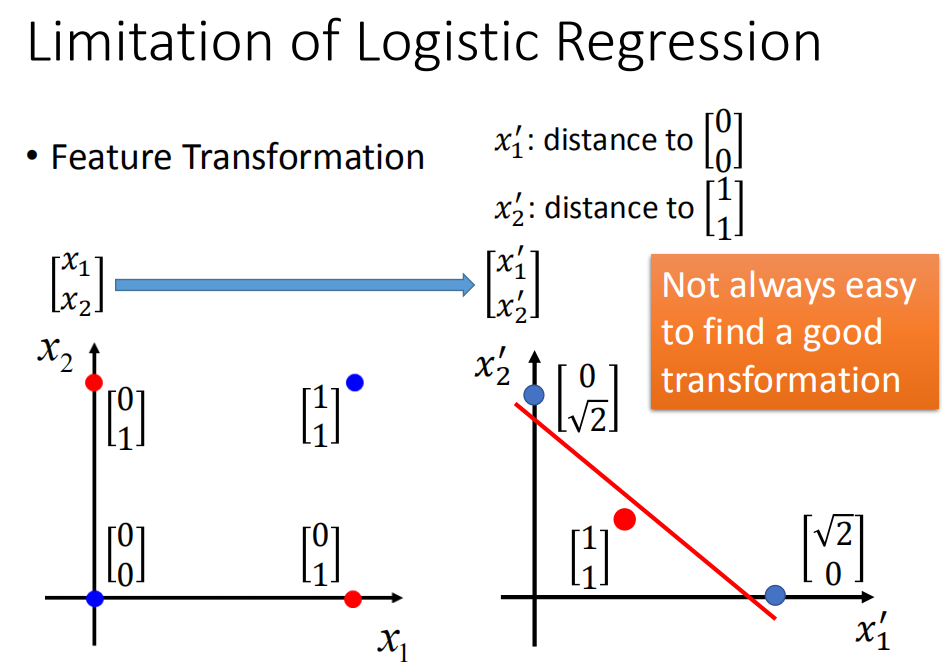
（2）对于噪声数据有更好的鲁棒性（robust）。因为生成模型有时候会忽略数据，基于自己的假设来考虑问题，也能够避免噪声数据的影响

（3）先验和类相关的概率可以从不同的来源估计。比如语音识别，可能直观会认为现在的语音识别大都使用神经网络来进行处理，是判别方法，但事实上整个语音识别是 Generative 的方法，DNN只是其中的一块而已；因为还是需要算一个先验概率，就是某句话被说出来的概率，而估计某句话被说出来的概率不需要声音数据，只需要爬很多的句子，就能计算某句话出现的几率。也就是说计算生成模型的概率其实可以不仅从训练集中获取，还可以从其他地方获取。

## 4. 逻辑回归的限制

### 4.1 限制：不能解决亦或问题的分类

### 4.2 对亦或问题可以使用特征转换，但是实际中并不是总能轻易的找到好的特征转换的方法。



### 4.3 级联逻辑回归模型，就像多层感知机，因此可以推导出深度学习

