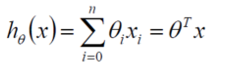
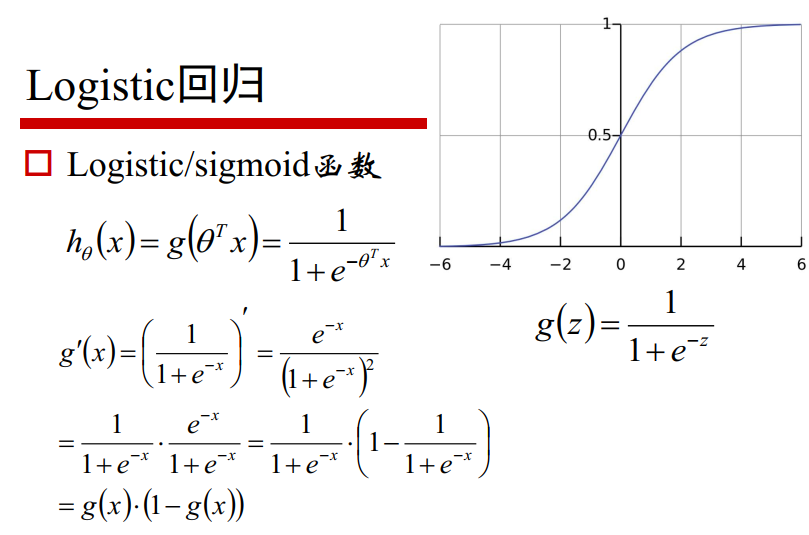
逻辑回归用于解决分类问题，是从线性回归的基础上推出来的，

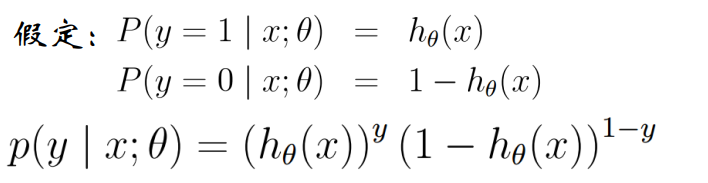


在线性回归中我们通过极大似然最终得到了其的损失函数L，即最小二乘法，但是在分类问题中，比如二分类问题，我们的样本真实值一般只有两个，比如0或者1，直接通过线性组合的结果去和0、1比较显然不太合理，逻辑回归在原来的线性组合基础之上作用一个缩放函数sigmoid函数，将组合结果压缩在0到1的范围之内，逻辑回归一般使用的损失函数为交叉熵等等。同样是通过目标函数学得最优参数，在决策函数中使用最优参数，最终决策也作用sigmoid映射，然后通过概率的形式来判断新样本的正负。

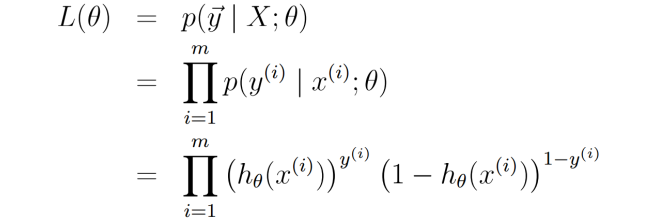


由于二分类问题可以看做是一个二项分布，及伯努利分布，假定样本独立

我们可以给出y的条件概率分布函数



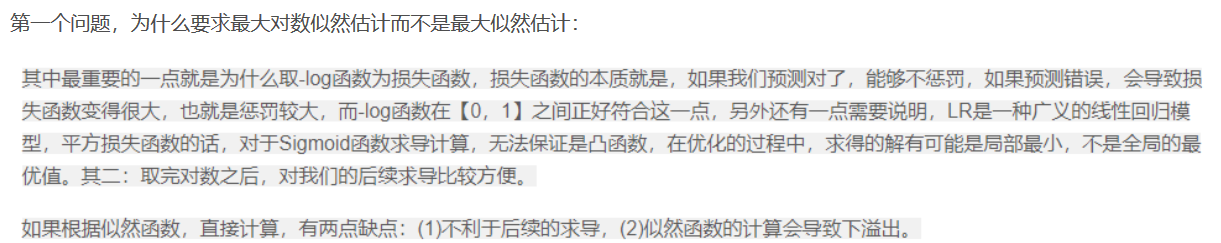
依然我们使用极大似然估计，我们希望p发生的概率最大，

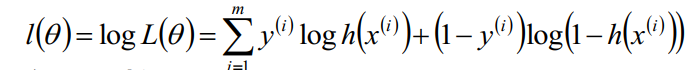


取对数log，取对数的原因：

1求导比较方便

2，如果是最小二乘法基础上作用sigmoid函数，不能保证目标函数为凸函数



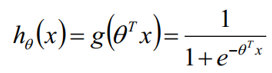


我们希望L越大越好，上述取反即可

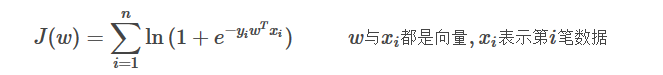


以上就是交叉熵最常见的表达式

y~就是hx



带入之后化解得到逻辑回归的损失函数

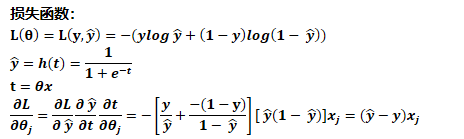


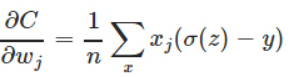
问题：为什么不使用平方损失函数而是利用极大似然？

可以从两个角度理解。

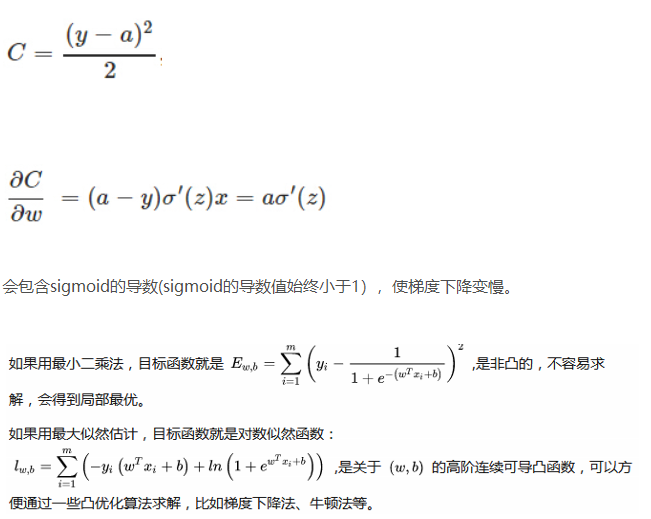
1. 交叉熵损失函数的好处是可以克服方差代价函数更新权重过慢的问题（针对激活函数是sigmoid的情况）。

原因是其梯度里面不在包含对sigmoid函数的导数：因为sigmod一个极大的优点就是，其导数形式

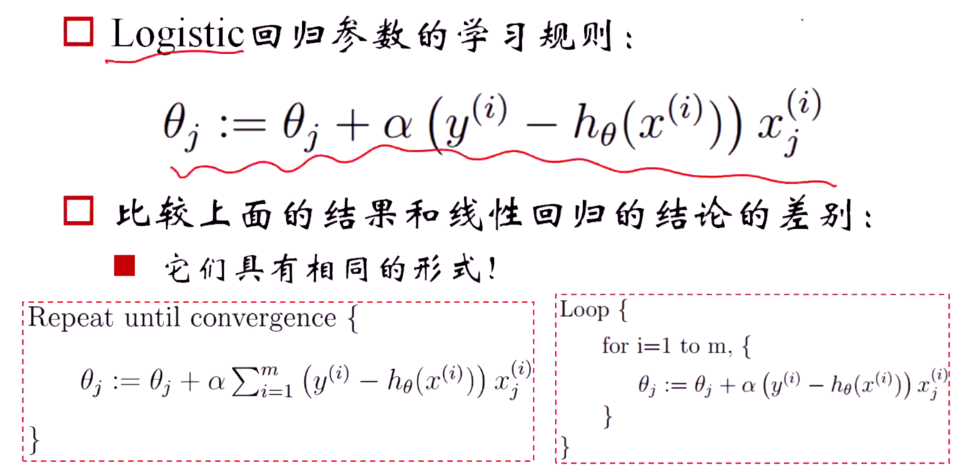




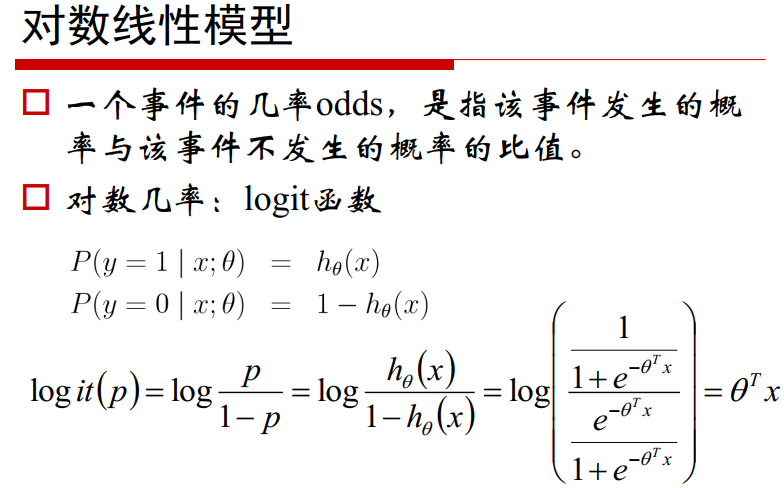
而如果使用的是平方损失函数加sigmoid函数，则计算梯度时：



最后我们看看线性回归和逻辑回归的梯度下降过程



和线性回归一模一样，区别就在于逻辑回归多了一个sigmoid映射，说明无论是从高斯分布的假设还是从二项分布的假设推过来，他们都是属于指数族分布，所以逻辑回归也是一个回归模型，他是一个广义的线性回归模型，他们都是线性模型，具体的说逻辑回归是一个对数线性模型



也就是说，一个事件发生的概率与不发生的概率的比值去对数之后是一个线性组合

我们看逻辑回归的决策边界，和线性回归一样，他还是一个线性模型正常在图上面就是一个直线一个平面而已，下图得到的决策边界其实是

对特征取了高阶运算，比如x平方，立方等等，才可以得到一个二维空间的曲线，就是说决策函数是可以通过特征的高阶操作模拟出来

