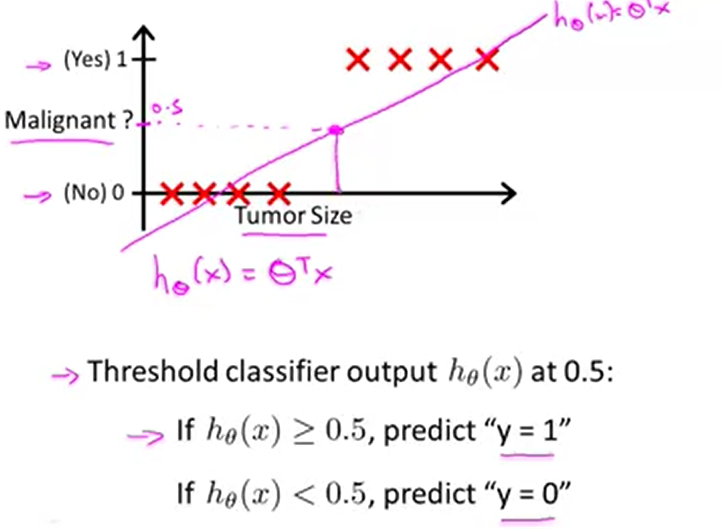
# 逻辑回归

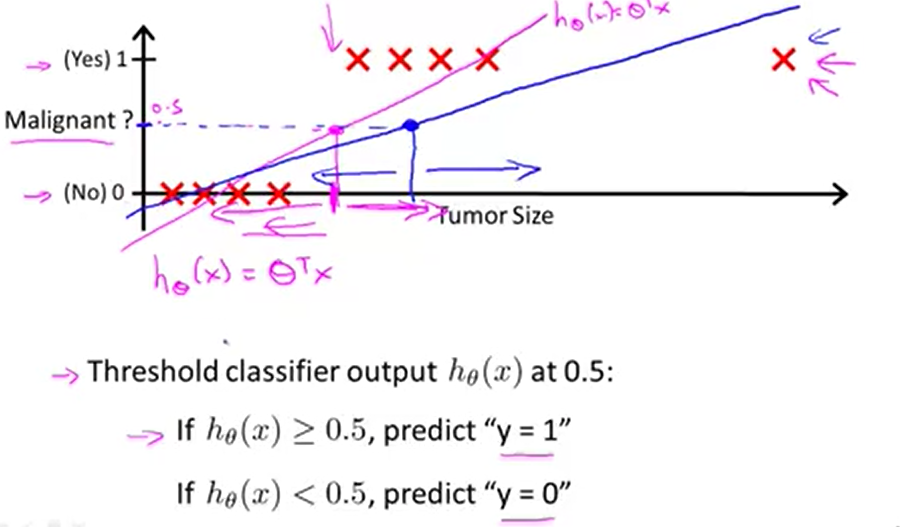
分类算法与回归算法的不同在于因变量是否是连续变量，逻辑回归可用于分类，本文的内容包括：（一）线性回归分类算法的弊端，（二）逻辑回归函数，（三）、正则化。

## 线性回归分类算法的弊端

线性回归算法预测值为连续的，线性回归分类算法受到训练样本的影响，如下图：

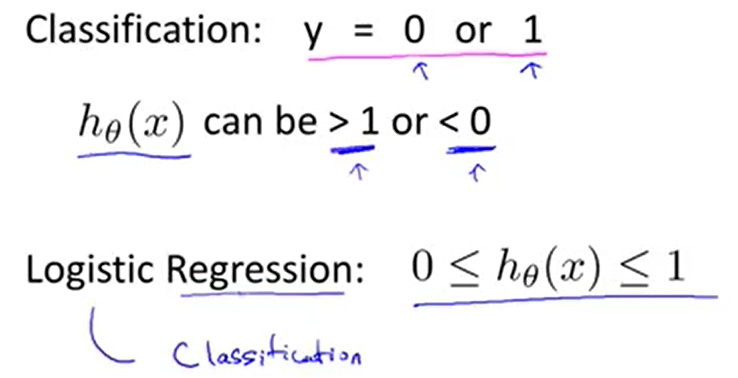


（一）对训练样本进行建模，当阈值大于等于0.5时则判断为恶性肿瘤，反之为良性。对于该训练样本的分布，线性回归分类算法能够正确分类，但假如一个特征分布偏离分布区域的训练样本时，则导致线性建模公式发生了改变，如下图蓝色曲线：



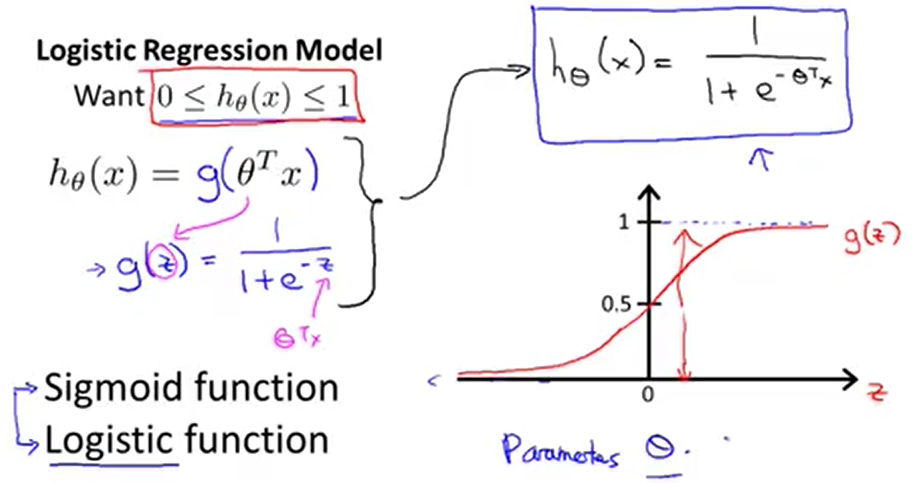
线性建模算法因为加入了一个异常的样本而发生了改变，导致了分类错误。

（二）、分类算法的分类结果（二分类）只有0和1两种情况，但是线性回归算法的取值结果不可控，可能远大于1或者小于0，这对分类算法来说非常奇怪（引用andrew老师的话），逻辑回归的算法限制结果范围在【0,1】区间。

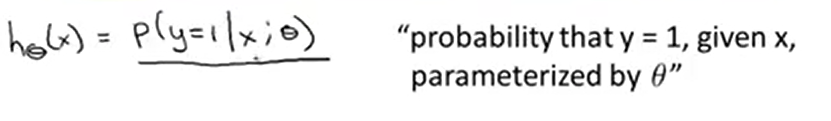


## 逻辑回归函数

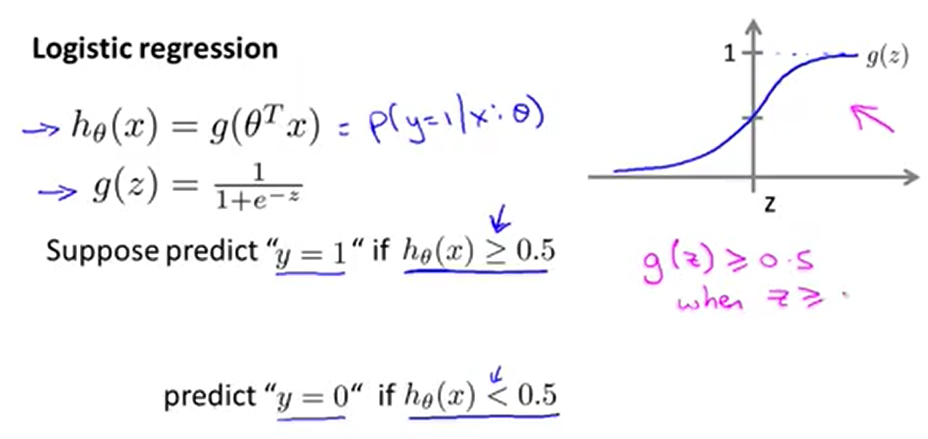
逻辑回归函数的表达式：



逻辑回归函数的意义：对于特定的输入特征，其逻辑回归的值代表了为1的概率，如下图：

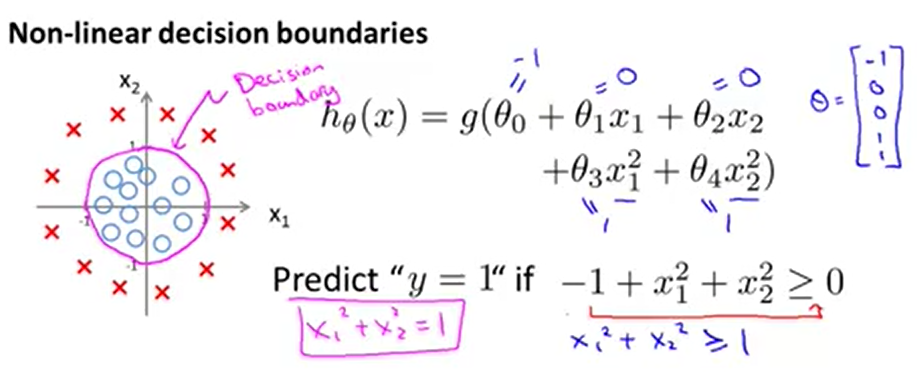


### 2.1 决策边界



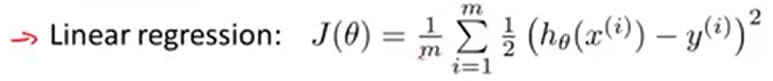
如上图所示，决策边界可以定为，若值大于等于零，则预测y=1的概率大于0.5，因此分类为1，反之分类为0。

分类边界可以应用到非线性分类方向，如下图：

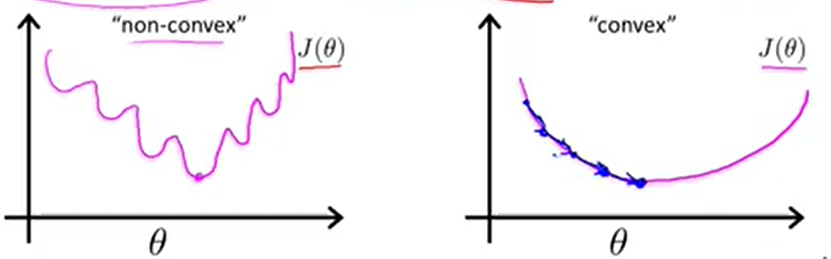


### 2.2 损失函数

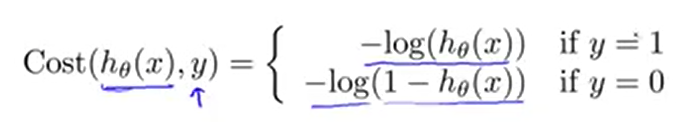
线性回归的损失函数：



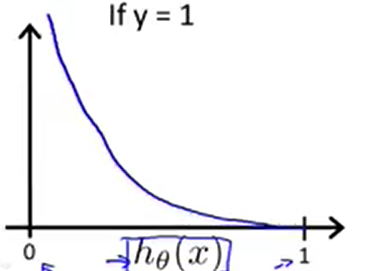
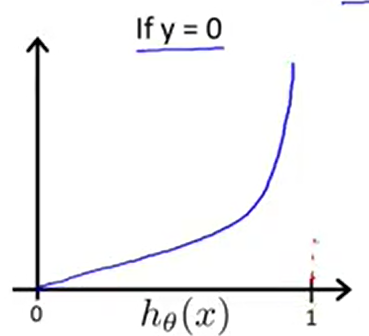
若逻辑回归也用该损失函数表达式，则会发现损失函数为非凸函数，有很多局部极值点，运用梯度下降法不能够定为损失函数的全局最小值，因此损失函数最好写成凸函数的表达式。



定义逻辑回归的损失函数：

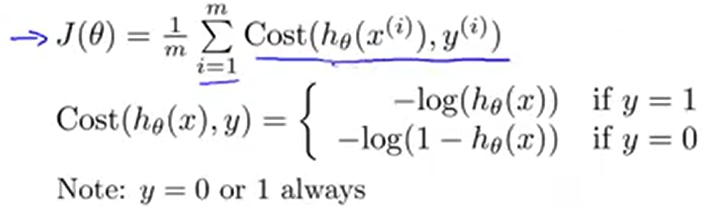


当训练样本值为1时，如果模型函数值为1，则无损失；若模型值为0，则损失函数值为无穷大；训练样本值为0的情况类似。

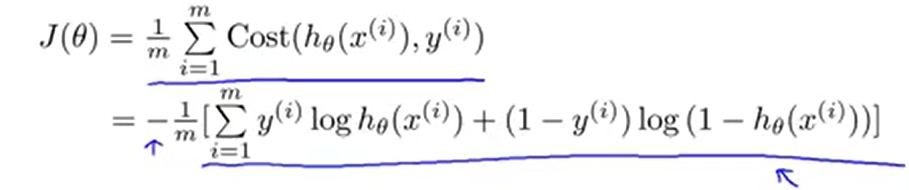
 

### 2.3 损失函数最小化

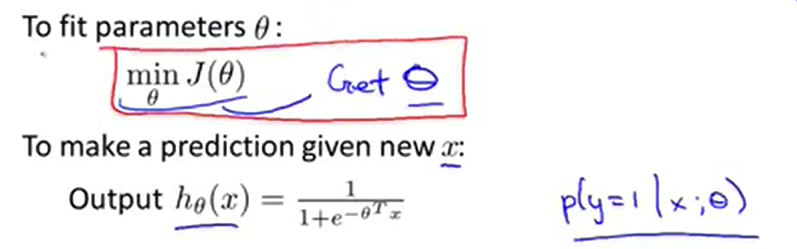
对m个训练样本，逻辑回归损失函数的表达式：



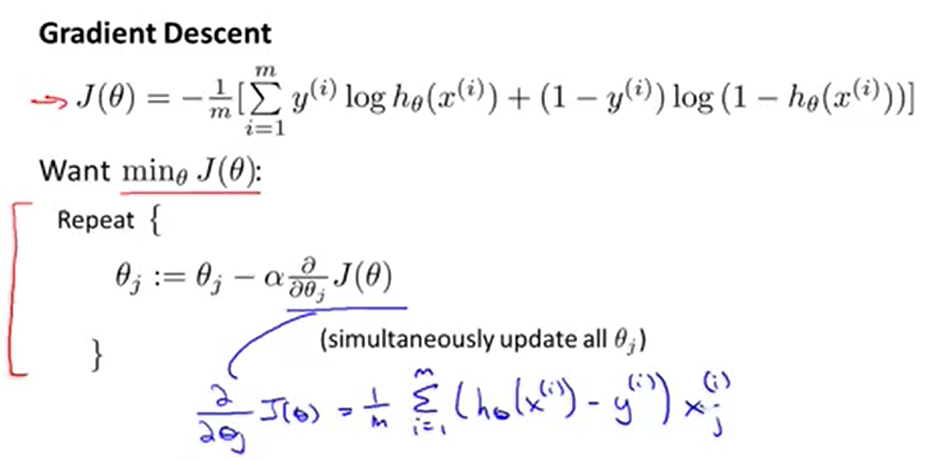
对于不同类的样本，逻辑回归损失函数可以合并为：



最后，最小化损失函数，得到参数θ，构建模型，逻辑回归的结果是输出为1的概率。

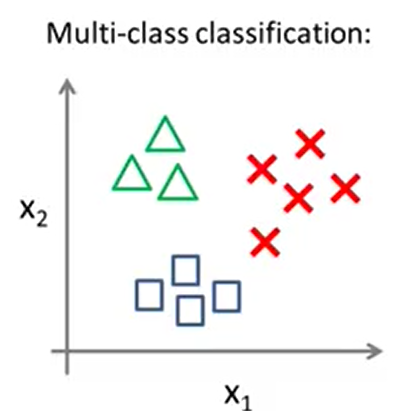


运用梯度下降法来求参数θ



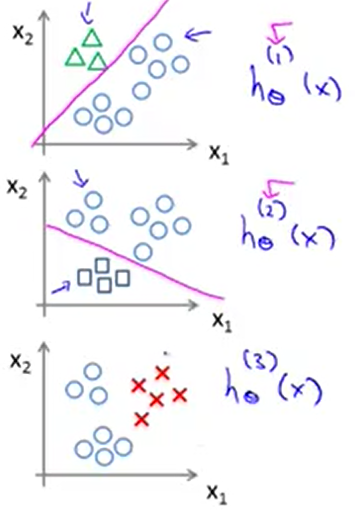
### 2.4 多类别分类

当样本判别类别大于两个时，则为多类别分类。

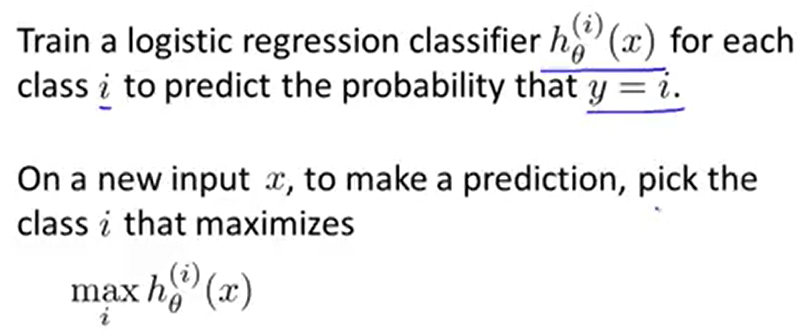


多分类问题可以用一对多的算法来构建分类模型，具体步骤如下（考虑三分类情况）：

1. 选择其中一个类，其余两个类合并为一个类，问题转化为二分类问题，对这二分类构建分类模型；
2. 重复第一个步骤两次，一共构建三次特定类的分类模型；



1. 输入测试数据，运用逻辑回归分类模型得的输出值，选择最大输出值所对应的分类模型，就是该测试分类结果；



## 正则化

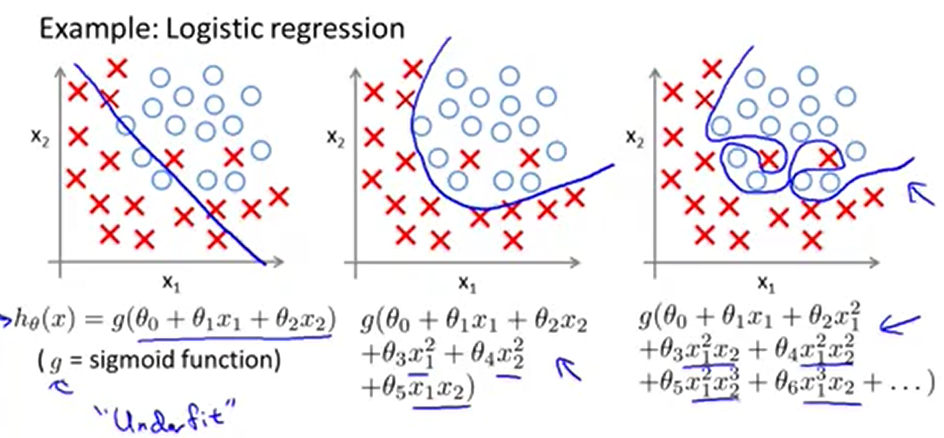
### 3.1 过拟合问题

如下图利用房子大小来预测房子价格的模型中，左一用了线性回归的方法，但是拟合效果很差，与给定数值的结果相差较大，因此称为欠拟合或高偏差；右一运用高阶多项式回归拟合样本数据，损失函数为零，但是泛化效果差，称为过拟合或高方差；中间图形的拟合效果最好。



**过拟合**：如果利用很多特征来构建模型，训练算法可能很好的拟合训练数据，但是在泛化新的测试数据时，效果很差。

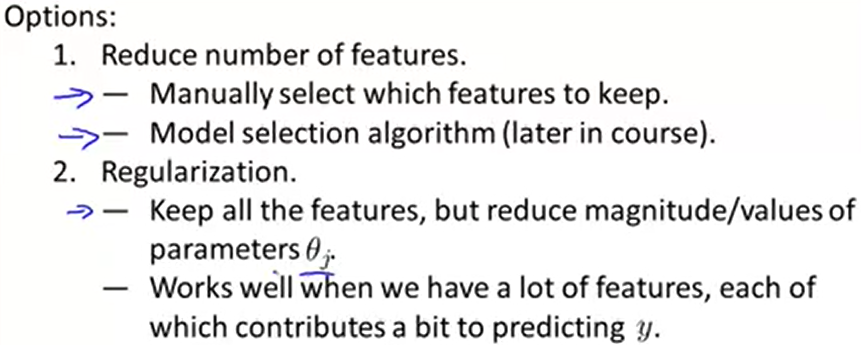
逻辑回归的欠拟合和过拟合问题：



当特征较多且训练样本数据较少时，就会出现过拟合问题。我们有两种方案来解决过拟合问题：

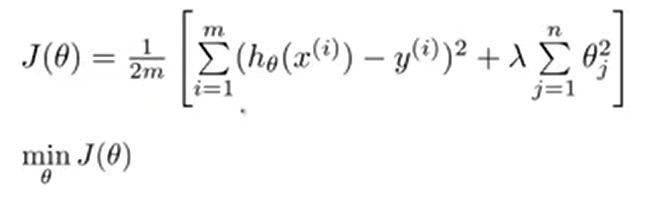
（1）、减少特征数量，i）人工选择保留特征，ii）模型选择算法，模型自动选择哪些变量可以保留，哪些变量舍弃；优点是减少过拟合情况的发生，缺点是舍弃了一些特征，可能会丢失与问题相关的信息。

（2）、正则化，因为每个特征可能对结果都有影响，保留所有与问题相关的特征，但是减小参数θ。

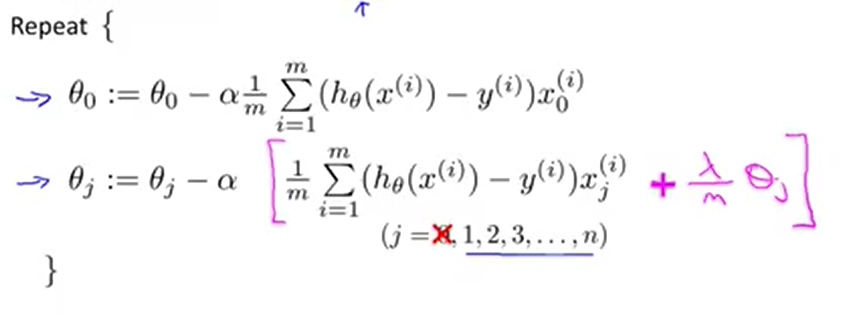


### 3.2 线性回归正则化

λ为正则化参数，通过求正则化方程的全局最小值来求参数θ。

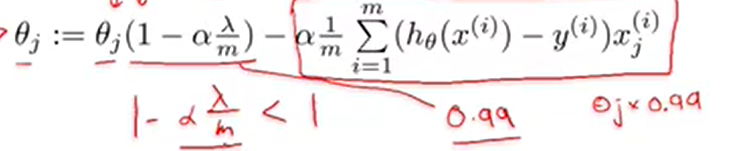


**梯度下降法求正则化全局最小值**：



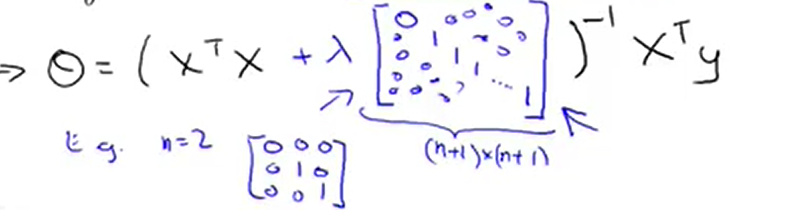
θ0的更新与其他参数分开，因为正则化方程没有包含θ0的惩罚参数。

对θj更新结果约等于0.99\*θj减去无正则化损失函数的偏导数：



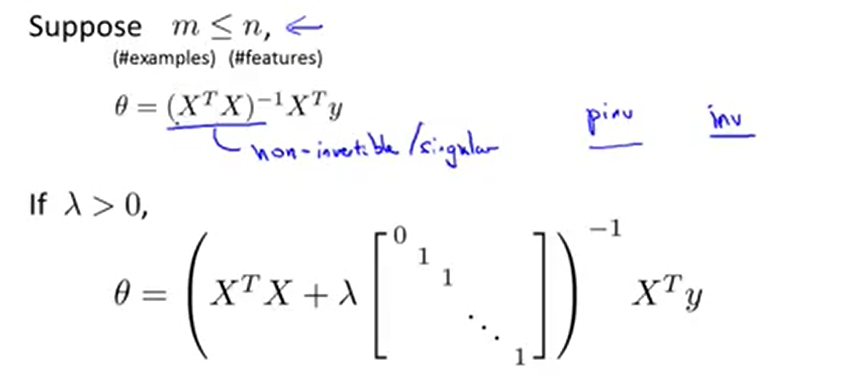
**正规方程求正则化全局最小值：**

损失函数对参数θ的偏导数设置为零，根据这一等式可以推导正规方程：



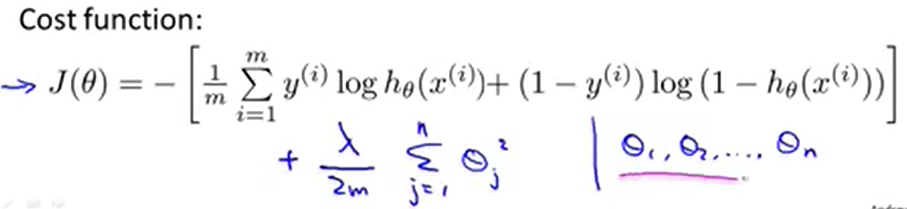
**讨论正规方程矩阵的不可逆性：**

对于样本数小于特征数，则会导致非正则化正规方程中的矩阵的不可逆性（奇异矩阵），但是正则化正规方程矩阵肯定可逆（非奇异矩阵）。



### 3.3 逻辑回归的正则化

逻辑回归的损失函数正则化方程：



运用梯度下降算法求全局最小值，参数θ迭代更新：

