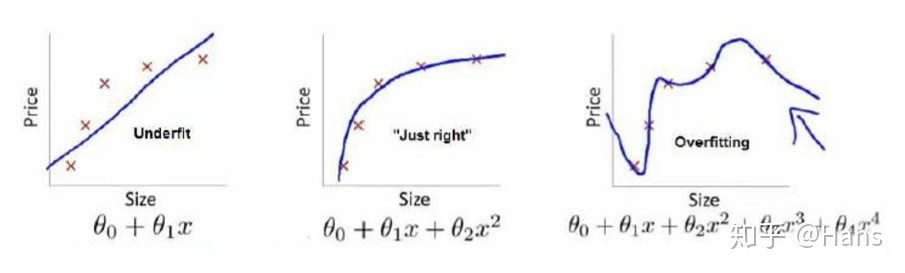
**1.过拟合的问题（Over-fitting）**

如果我们有非常多的特征，我们通过学习得到的假设可能能够非常好地适应训练集，但是可能会不能推广到新的数据。

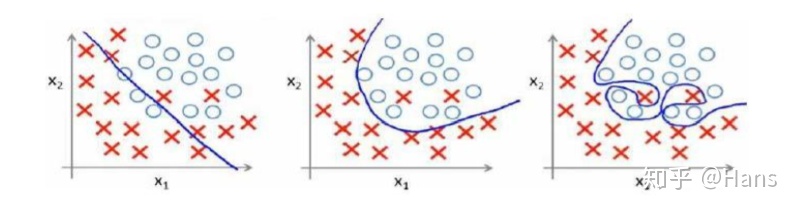
（1）下图是一个回归问题的例子：



* 第一个模型是一个线性模型，欠拟合，不能很好地适应我们的训练集
* 第三个模型是一个四次方的模型，过于强调拟合原始数据，而丢失了算法的本质：预测新数据。

**我们可以看出，若给出一个新的值使之预测，它将表现的很差，这叫作过拟合。虽然能非常好地适应我们的训练集，但是新输入变量进行预测时，效果会表现的不好。**

（2）分类问题中也存在这样的问题：



判断方法：模型在验证集合上和训练集合上表现都很好，而在测试集合上变现很差。

**解决过拟合的方式：**

1. **减少特征值的数量**

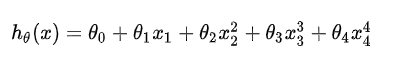
* **人为的保留一些重要的特征值**
* **用特征选择算法进行特征的选择（PCA、层次分析法）**

1. **正则化**

* **保留所有的特征，但是减少参数的大小。**

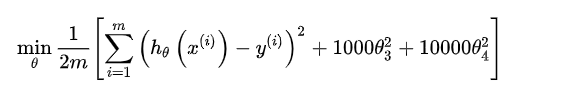
**2.代价函数（Cost Function）**

上面的回归问题中如果我们的模型是：



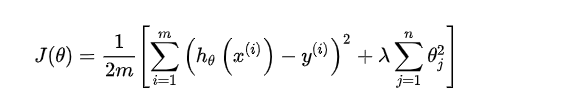
**我们可以从之前的例子中看出来，正是由于高次项导致的过拟合的发生，所以如果我们能够让那些高次项的系数接近于0的话，我们就能够很好的拟合数据集。**

所以我们要做的就是在一定程度上减小这些参数 的值，这就是正则化的基本方法。我们决定要减少 theta3 和 thea4的大小，我们要做的便是修改代价函数，在其中 theta3 和 theta4 设置一点惩罚。这样做的话，我们在尝试**最小化代价时也需要将这个惩罚纳入考虑中，并最终导致选择较小一些的 theta3 和 theta4 。**

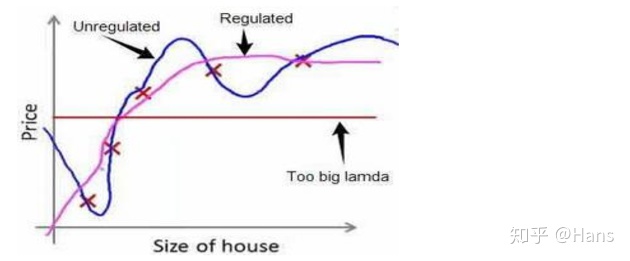
修改后的代价函数： 

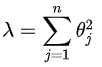
**通过这样的代价函数选择出的theta3和theta4 对预测结果的影响就比之前要小许多。**

假如我们有非常多的特征，我们并不知道其中哪些特征我们要惩罚，我们将对所有的特征进行惩罚，并且让代价函数最优化的软件来选择这些惩罚的程度。这样的结果是得到了一个较为简单的能防止过拟合问题的假设：

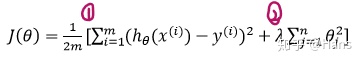


其中  又被称为正则化参数（Regularization Parameter）。注：按照惯例我们对 theta0 不进行惩罚。经过正则化处理的模型与原模型的可能对比如下图所示：



**那为什么增加的一项****可以使 的值减小呢？**

: 可以控制两个不同目标之间的取舍。



目标①：去更好地拟合数据集（训练集）

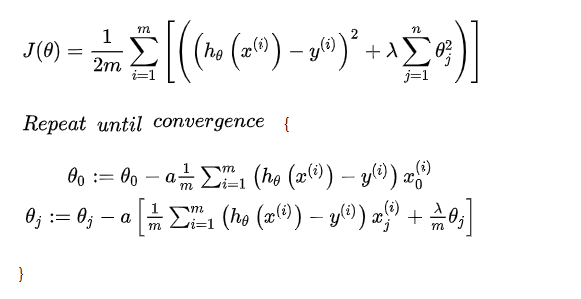
目标②：使参数大小变小，从而使假设函数变得更“简单”。

**两者相互平衡，从而达到一种相互制约的关系，最终找到一个平衡点，从而更好地拟合训练集并且具有良好的泛化能力。**

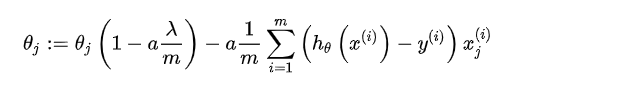
因为如果我们令  的值很大的话，为了使Cost Function 尽可能的小，所有的 的值（不包括 theta0 ）都会在一定程度上减小。如果 取值非常大，那么除了之外的所有参数都趋近于0，模型成为一条直线。

**3.正则化线性回归**

正则化线性回归的代价函数为：



对上面的算法中 = 1,2,..., 时的更新式子进行调整可得：

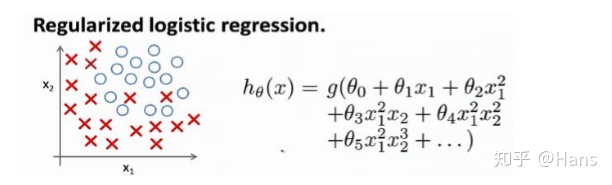


可以看出，正则化线性回归的梯度下降算法的变化在于，每次都在原有算法更新规则的基础上令值减少了一个额外的值。

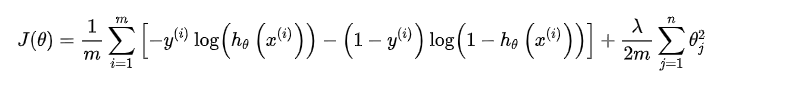
https://www.jianshu.com/p/7058302063ce

https://github.com/Damon0626/My-Projects/blob/b339dccd466b90aacfef276441768aeac4f1bf0a/09-Regularized%20Linear%20Regression%20and%20Bias-Variance/01-Regularized%20Linear%20Regression%20and%20Bias-Variance.py#L62

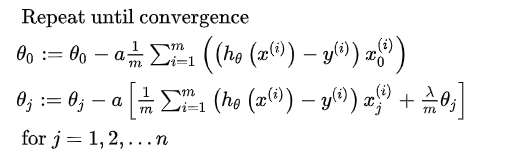
**4.正则化的逻辑回归模型**



自己计算导数同样对于逻辑回归，我们也给代价函数增加一个正则化的表达式，得到代价函数：



要最小化该代价函数，通过求导，得出梯度下降算法为：



**5.如何更好地拟合数据**

1. 去收集更多的数据
2. 尝试更少的特征值
3. 尝试更多的特征值
4. 增加多项式
5. 增大\减小正则化参数

**小数据集只能用来拟合简单的模型，复杂的模型必须用大数据集来拟合。正则化的作用，简而言之就是在用小数据集拟合模型的过程中，强行加入人为偏转，强烈地引用简单的模型，从而控制拟合出来的模型的复杂度，避免过拟合。如果样本数跟参数数量相比很多的话，正则化一般来说是多余的。**