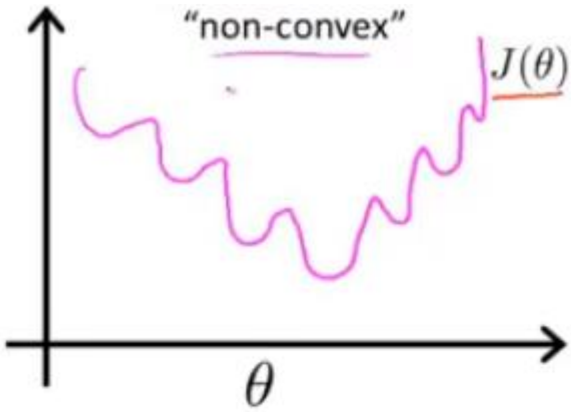
1、原版代价函数

在线性回归拟合中，我们利用代价函数求偏导的思想，对θ参数进行逐步优化，使得最后的代价差为局部最优值。同样，我们来看，这样的一个思想能否也利用在 logistic 模型优化中呢？

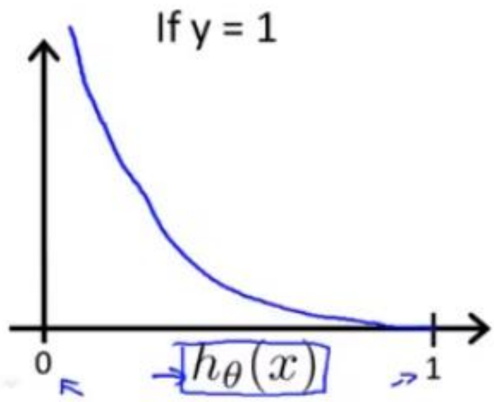
Logistic 模型，$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$，若用$J_{(\theta)}$进行优化，得到的图形是这样的，它是一个非凸函数，因此我们无法通过原来的代价函数进行梯度下降得到想要的θ值。

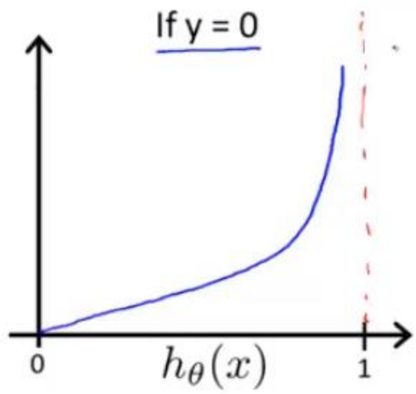


2、新版代价函数

针对新的模型，应用新的参数求解过程。直接看定义：

$$Cost(h_\theta(x), y) = \begin{cases} -log\left(h_\theta(x)\right) & \text{if } y = 1 \\ -log\left(1 - h_\theta(x)\right) & \text{if } y = 0 \end{cases}$$





在左图中，当样本点属于 1 结果时，此时计算它的代价函数为$-log\left(h_\theta(x)\right)$，若$h_\theta(x)$预测值为 1 时，代价为 0，反之趋于无穷，这就限制住了对样本值为 1 的点预测它趋于 0 的可能性；右图同理，若$h_\theta(x)$预测值为 0 时，代价为 0，反之趋于无穷，这就限制住了对样本值为 0 的点预测它趋于 1 的可能性。

到目前为止，我们学会了用 logistic 模型拟合分类训练集，对训练集拟合的好坏，给出了新的代价函数方程。接下来，通过梯度下降算法来给出一个较好的θ。

3、梯度下降算法——优化θ

之前用方程组的形式表示出当 y 为 0、1 时的代价函数，现在用一个式子来表示：

$Cost(h_\theta(x),y) = -ylog\left(h_\theta(x)\right) + (y-1)log\left(1-h_\theta(x)\right)$ ， y = 0/1

这是对单个样本进行代价值计算，现在扩展到所有训练集中，得出总的代价函数为：

$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)}log\left(h_\theta\left(x^{(i)}\right)\right) + (1-y^{(i)})log\left(1-h_\theta\left(x^{(i)}\right)\right)]$，值得庆幸的是，这个代价函数是一个凸函数，因此可以用梯度下降算法进行θ值的迭代更新：

$$\begin{cases} \theta_j = \theta_j - \alpha\frac{1}{m}\frac{\partial}{\partial\theta_j}J(\theta) \\ \\ \theta_j = \theta_j - \alpha\,\frac{1}{m}\sum_{i=1}^{m}(h_\theta\left(x^{(i)}\right) - y^{(i)})x_j^{(i)} \end{cases}$$

可以看到，这一个对θ值的更新，与之前讲的线性回归梯度下降算法所用公式是一样的，但其实本质是不同的，因为对预测模型的定义完全不同。同样，向量化运算和特征值缩放也适用于 logistic 模型。

参数值更新：$\theta = \theta - \alpha\frac{1}{m}\mathrm{X}' * (\mathrm{y_{pred}} - \mathrm{y})$，

代价函数计算：$\mathrm{J} = -\frac{1}{m}sum(y.* \log\left(y_{pred}\right) + (1-y).* \log\left(1-y_{pred}\right))$