Thinking

郭想

逻辑回归的假设条件是怎样的？

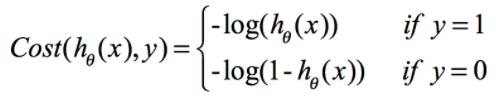
主要有两个假设条件：

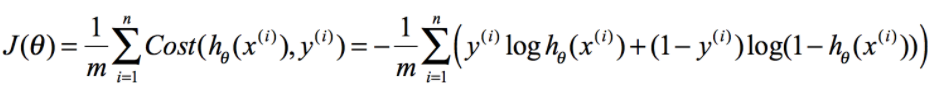
假设1：数据服从伯努利分布

伯努利分布有一个简单的例子是抛硬币，抛中为正面的概率是p ，抛中为负面的概率是1-p。

假设2：正类的概率由sigmoid函数计算

逻辑回归的损失函数是怎样的？





逻辑回归如何进行分类？

设定一个阈值，判断正类概率是否大于该阈值，一般阈值是0.5，所以只用判断正类概率是否大于0.5即可。 当然，根据具体情况这个阈值可以改变。

为什么在训练中需要将高度相关的特征去掉？

（1）去掉高度相关的特征会让模型的可解释性更好。

（2）可以大大提高训练的速度。如果模型当中有很多特征高度相关的话，就算损失函数本身收敛了，但实际上参数是没有收敛的，这样会拉低训练的速度。其次是特征多了，本身就会增大训练的时间。