# 概述

Logstic回归(逻辑回归)的因变量常为二元分类变量，其自变量既可以是分类变量也可以是连续变量。

注：多元线性回归（Ridge、Lasso、ElasticNet）是做回归预测的

逻辑回归（Logistic Regression）是做分类任务的

# 回归原理

Logistic回归通过Logit转换将取值为正负无穷的线性方程的值域转化为(0,1)，正好与概率的取值范围一致。

说明：线性回归是一种预测方法，适用于预测连续型变量，而Logistic回归适合预测分类变量，而且其预测的是一个区间在0到1的概率。

# 极大似然估计

线性回归采用的是最小二乘法进行参数的估计，逻辑回归采用极大自然法进行参数估计。

极大似然估计是一种找出与样本的分布最接近的概率分布模型，“似然”表示的是已知结果的前提下，随机变量分布的最大可能。

# 损失函数

Q：做回归预测损失函数是什么？

A：平方均值损失函数MSE

Q：做分类损失函数是什么？

A：做分类损失函数是交叉熵！

# 熵

Q：什么是熵？

A：熵是一种测量分子不稳定性的指标，分子运动越不稳定，熵就越大，来自热力学

熵是一种测量信息量的单位，信息熵，包含的信息越多，熵就越大，来自信息论，香农

熵是一种测量不确定性的单位，不确定性越大，概率越小，熵就越大！

Q：熵和概率是什么一个关系？

A：随着概率的减小，熵会增大

逻辑回归

Q：为什么本质是多元线性回归？

A：1，公式，首先应用了多元线性回归的公式，其次才是把多元线性回归的结果，交给sigmoid函数去进行缩放

2，导函数，逻辑回归的损失函数推导的导函数，整个形式上和多元线性回归基本一致，

只是y\_hat求解公式包含了一个sigmoid过程而已

Q：逻辑回归的损失函数是什么？

A：交叉熵，做分类就用交叉熵，-y\*logP，因为逻辑回归是二分类，所以

loss func = (-y\*logP + -(1-y)\*log(1-P))，也就是说我们期望这个损失最小然后找到最优解

事实上，我们就可以利用前面学过的梯度下降法来求解最优解了

Q：逻辑回归为什么阈值是0.5？

A：因为线性回归区间是负无穷到正无穷的，所以区间可以按照0来分成两部分，所以带到sigmoid公式里面去

z=0的话，y就等于0.5

Q：逻辑回归做多分类？

A：逻辑回归做多分类，把多分类的问题，转化成多个二分类的问题，如果假如要分三个类别，就需要同时训练三个互相不影响的模型，比如我们n个维度，那么三分类，w参数的个数就会是(n+1)\*3个参数

上面所谓的互不影响，指的是模型在梯度下降的时候，分别去训练，分别去下降，三个模型互相不需要传递数据，也不需要等待收敛

Q：文字本身是几维的数据？音乐本身是几维的数据？图片本身是几维的数据？视频本身是几维的数据？

A：看什么类型的数据，文字一维的，音乐是单声道的音乐，音乐是一维的数据，如果音乐是双声道的，

就是二维的数据，图片如果看成是张图片，就是个平面二维的数据，视频一张张图片按时间顺序码放的，

那就是三维的数据

但我们做机器学习的时候，真的只会这样考虑吗？

文章是由不同的词组成的，词的种类越多，事实上考虑的维度就越多

图片如果是彩色的图片，图片可以有R、G、B、alpha

音乐可以有不同的频率，每个频率如果看成是一个维度，那么就可以N多个维度

图片也可以有不同的频率，每个频率如果看成是一个维度，那么就可以N多个维度