逻辑回归的学习和理解————Day 4-5

2019 年 7 月 14 日

1　认识逻辑回归

- **回归任务与分类任务**
  我们按照任务的种类，将任务分成回归任务和分类人物两种。输入变量和输出变量均为连续变量的预测问题就是回归问题，输出变量为有限个离散变量的预测问题是分类问题
  比如我们想通过一个人的饮食情况预测一个人的体重，体重的值可以有无限多个，但预测的结果是一个确定的数，而具体为哪个数有无限多个可能，这时我需要训练出一个模型，传入参数后输出我们想要预测的这个确定的数，这类问题就是回归问题。由于预测的这个变量有无限种可能的值，在数轴上是连续的，所以这种变量为连续变量。
  而又比如我们想要预测一个人的星座，我们知道每个星座对应几个独立的日期，我们预测的结果只有几个值，把每个值当作一类，预测对象到底属于哪一类。这样的问题便是分类问题

- **什么是逻辑回归？**
  逻辑回归被用来处理不同的分类问题，这里的目的是预测当前被观察的对象属于哪一个组。其会给你提供一个离散的二进制输出结果。一个简单的例子就是判断一个人是否会在即将到来的选举中进行投票。

- **逻辑回归与线性回归**
  逻辑回归给出离散的输出结果，然而线性回归给出的是连续的输出结果。

2 逻辑回归的使用和原理

- **工作原理**
  逻辑回归使用基础逻辑函数通过估算概率来测量因变量即我们想预测的变量，和一个或者多个自变量之间的关系。

- **预测结果**
  估算的概率值要转换为二进制数，一边实际中进行预测。这是逻辑函数的任务，也被称为sigmoid函数。

- **Sigmoid函数**
  Sigmoid函数是一个S形曲线，可以实现将任意真实值映射为值域为0-1的值。

$$Y(x) = \frac{1}{1+e^{-x}}$$

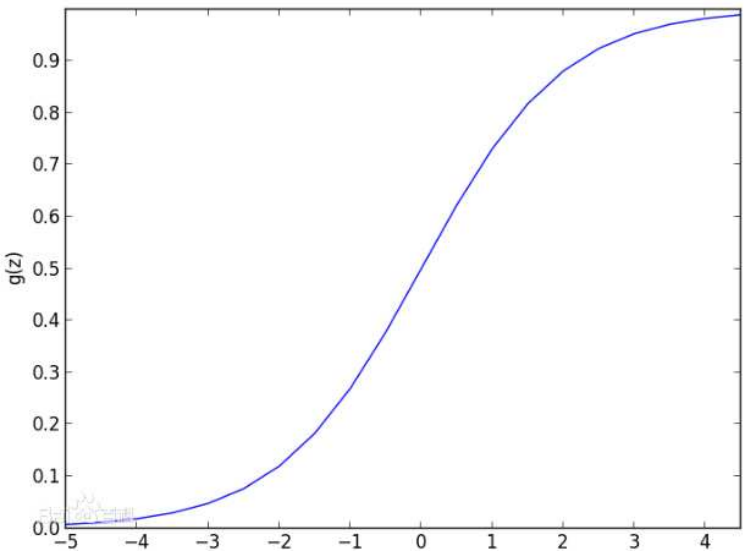

图 1: Sigmoid函数

按照多元线性回归的思路,我们可以先对某个任务进行线性回归,学习出这个事情结果的规律,比如根据人的饮食,作息,工作和生存环境等条件预测一

个人”有”或者”没有”得恶性肿瘤,可以先通过回归任务来预测人体内肿瘤的大小,取一个平均值作为阈值,假如平均值为y,肿瘤大小超过y为恶心肿瘤,无肿瘤或大小小于y 的,为非恶性.这样通过线性回归加设定阈值的办法,就可以完成一个简单的二分类任务。该函数可将函数的输入范围(∞,-∞)映射到了输出的(0,1)之间且具有概率意义。具有概率意义的理解:将一个样本输入到我们学习到的函数中,输出0.7, 意思就是这个样本有70% 的概率是正例,1-70%就是30%的概率为负例。

- **逻辑回归实质**
  发生概率除以没有发生概率再取对数。就是这个不太繁琐的变换改变了取值区间的矛盾和因变量自变量间的曲线关系。究其原因，是发生和未发生的概率成为了比值，这个比值就是一个缓冲，将取值范围扩大，再进行对数变换，整个因变量改变。不仅如此，这种变换往往使得因变量和自变量之间呈线性关系，这是根据大量实践而总结。所以，逻辑回归从根本上解决因变量要不是连续变量怎么办的问题。

明天将进行对逻辑回归的*python*实现