**逻辑回归原理简单介绍**

假设一个具有个样本的数据集中有个特征，则线性回归为：

 (1)

逻辑回归就是在线性回归模型上加了一个sigmoid变换：

 (2)

最终逻辑回归模型为：

 (3)

上面一个式子可以等价变换为：

 (4)

若将视为样本作为正例的可能性，则是其反例的可能性，并将视为类后验概率估计，则上式可变为：

 (5)

显然有：



通过极大似然法来估计和，即最大化对数似然：

 (8)

即令每个样本属于其真实标记的概率越大越好。为便于讨论，将考虑进一起讨论，令则可简写为。再令，，则式(8)中似然项可重写为：

 (9)

将式(9)带入式(8)，并根据式(6)和(7)可知，式(8)可变为：

  (10)

注意



所以式(10)可重写为：

 (11)

最大化式(8)等价于最小化：

 (12)

上式是关于的高阶可导连续凸函数。用梯度下降法可求得其最优解，于是得到：

 (13)

式(12)对求一阶导数得：

 (14)

所以，其第t+1轮迭代解的更新公式为：



**代码实现**

分别利用腾讯数据集和ipinyou数据集各实现了一个简单的逻辑回归例子，在腾讯数据集上是调用sklearn库函数实现的，在ipinyou数据集是自己写函数实现的，建议都看一看。

需安装的库：

numpy

scipy

pandas

sklearn

operator

等等，缺什么装什么

安装方法举例： pip install numpy

对于后缀名为.ipynb的代码文件，可以直接用pycharm打开，也可以安装jupyter库，在项目更目录下终端输入jupyter notebook命令进行查看，更多使用方法请百度。