竞赛一

我采用的是逻辑回归模型

步骤：

以HTRU\_2\_train.csv为训练集，以HTRU\_2\_test.csv为验证集

第一步：将训练集中的数据和真实的类别标签分别用dataIn，target来进行存放

第二步：利用梯度下降的方法，来求代价函数的最小值，并返回一个(width1,1)的矩阵。

第三步：利用返回的矩阵来求概率，大于0.5的为一类，小于0.5的为一类。

第四步：利用预测的类别来求正确率。

所用到的函数：

1. sigmoid：用来计算概率
2. init\_data：用来对训练集的数据进行提取分成两个属性和真实的类别标签（dataIn，target）。注：其中dataIn中的数据在所有的行中第0列都填1，为了是计算系数中的偏量。

data：将从HTRU\_2\_train.csv中读取的数据存到data里面

line,width:用来记录data的行数和宽数

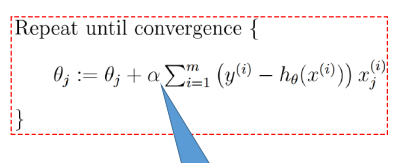
target:是用来存放训练集的真实标签的

data\_1:是将data里面的所有属性存到里面，不包括真实的标签

dataIn:将data\_1中所有行的第1列都添加1，用来方便计算偏量。

返回dataIn和target

1. grad\_descent：用来进行学习的，利用下面的公式（梯度下降）来进行求代价函数的最小值：



dataMatrix:是dataIn的矩阵化

labelMat：是target的矩阵化的转置

line1,width1：用来纪律dataMatrix的行数和宽数

weights：用来放各个属性的系数的矩阵，所以矩阵的大小为(width1,1)

alpha：是学习率

maxCycle：是循环的次数

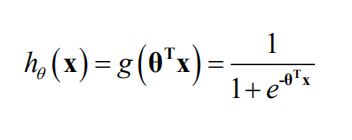
h：用来存放每个数据的概率大小为（line，1），因为每回的weights都不一样所以每回的概率都不一样。

weights：是进行梯度下降来求代价函数的最小值。

返回weights,h

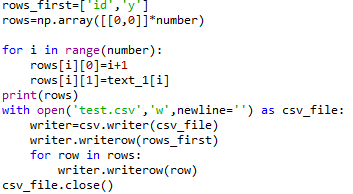
下面就是对测试集的预测：

先求概率，再判断概率是否大于0.5，如果概率大于0.5就将其预测结果为1，反之预测结果为0。

求概率的公式：

最后，得到预测结果并把结果保存到一个数组里面（text\_1）

最后，利用一个写入文件的代码：



将预测的值写入一个CSV文件中，并提交。