**中山大学数据科学与计算机学院**

**移动信息工程专业-人工智能**

**本科生实验报告**

**（2016学年秋季学期）**

课程名称：Artificial Intelligence

|  |  |  |  |
| --- | --- | --- | --- |
| 教学班级 | **M2** | 专业（方向） | **移动互联网** |
| 学号 | **14353205** | 姓名 | **刘万里** |

# 实验题目

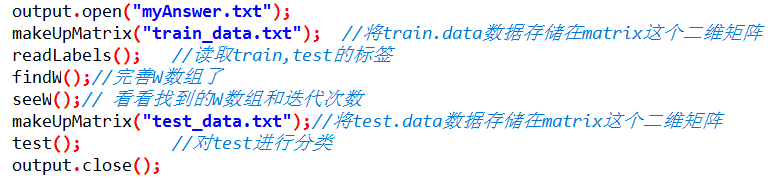
感知机算法实现

# 实验内容

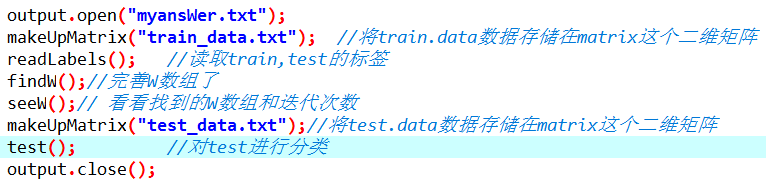
1. 算法原理

对于二元分类问题，向量的每一维对应一个权重，我们尽可能地找出这样的一个权重向量，使得可以完美地（至少也是错误率最低的）划分二元分类，通过迭代的算法每次更新权重向量，直至每一个训练样本都可以正确划分（至少也是错误率最低），再使用该权重向量去划分测试机得到分类。

1. 伪代码
   1. Init-PLA



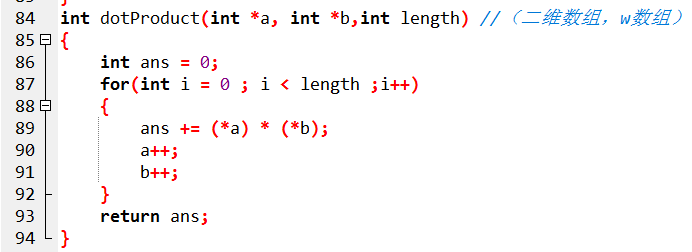
* 1. Pocket-PLA



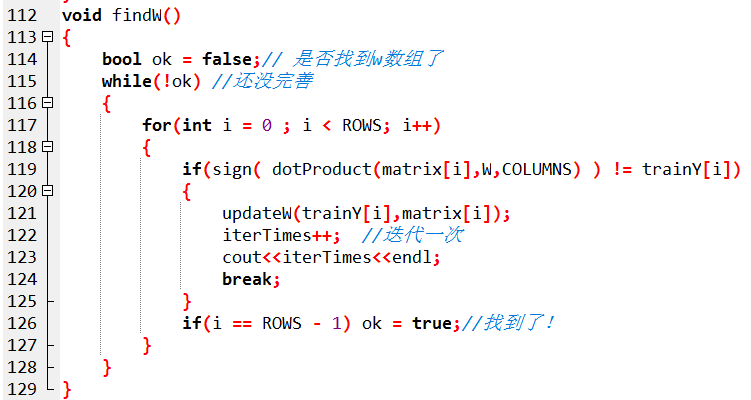
* 1. 优化（如果有）

1. 关键代码截图（带注释）
   1. Init-PLA

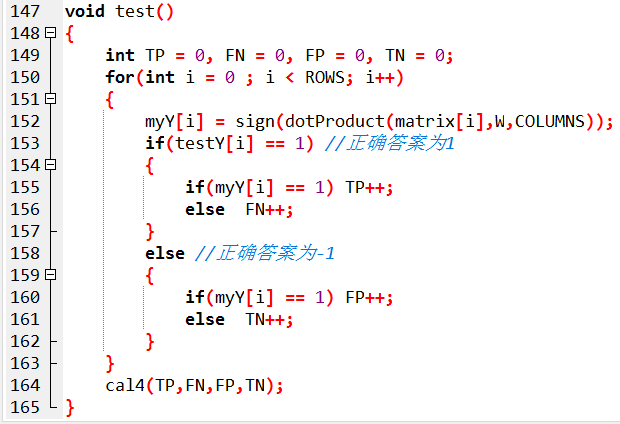
封装好的便于使用的点乘函数：



找到正确划分的W向量：



对测试集进行划分：

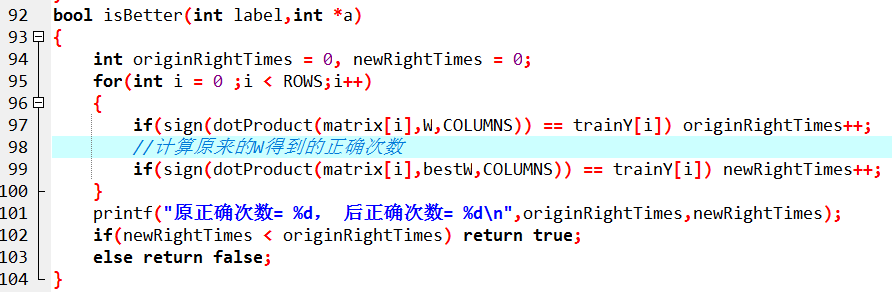


* 1. Pocket-PLA

都和原始的PLA是一样的，不过寻找W数组的方法不一样：bestW就是找到的最好的权重向量



其中比较函数isBetter:



1. 创新点&优化（如果有）

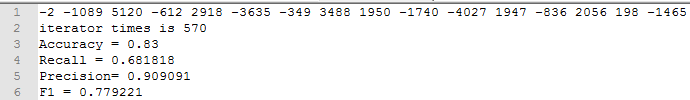
# 实验结果及分析

1. 实验结果展示示例（可图可表可文字，尽量可视化）

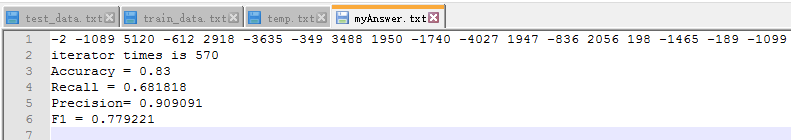
(原始算法结果)

\\请截图你们的运行结果，包含四种指标

* 1. Init-PLA



* 1. Pocket-PLA



**2. 评测指标展示即分析**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Accuracy | Precision | Recall | F-1 |
| Init | 0.83 | 0.909091 | 0.681818 | 0.779221 |
| pocket | 0.83 | 0.909091 | 0.681818 | 0.779221 |

**|-----------如有优化，请重复1，2，分析优化后的算法结果-----------------------|**

# 回答问题

1. 请查询相关资料，记录这四种评测指标有什么意义，尤其是在信息检索中

Accuracy准确度 is the difference between the measured value and the true value of a tested material.

Precision精确度 is the repeatability of successive measurements under the same conditions.

测量的准确度高，是指系统误差较小，这时测量数据的平均值偏离真值较少，但数据分散的情况，即偶然误差的大小不明确。

测量精确度（也常简称精度）高，是指偶然误差与系统误差都比较小，这时测量数据比较集中在真值附近。

**召回率(recall)**的公式是,它计算的是所有"正确被检索的item(TP)"占所有"应该检索到的item(TP+FN)"的比例。

F1值就是精确值precision和召回率recall的调和均值

通俗地说没，准确率就是**找得对**，召回率就是**找得全**

**准确率和召回率是互相影响的，理想情况下肯定是做到两者都高，但是一般情况下准确率高、召回率就低，召回率低、准确率高**

**2. 思考迭代数对于原始PLA算法有什么影响**

如果train的迭代次数增加，有助于提高正确率，对于线性不可分的数据也能找到一个较为理想的犯错误最少的权重w。

**3. 有什么其他的手段可以解决数据集非线性可分的问题？**

PS：如果不喜欢我们提供的模板，可以自行重新设计，唯一的要求就是这个模板里面提到的内容大纲不能更改，另，本学期的实验不需要写实验感想。

命名格式为：学号\_姓名拼音.pdf