

山东大学机器学习课程

实验报告

——实验四：线性判别函数

姓名：陈竞帆

学院：软件学院

班级：软件八班

学号：201500150180

**一、实验目的：**

（1）熟悉python实验软件及相关函数

（2）学习以基本梯度下降法和牛顿梯度下降法进行准则函数求极值的思想与算法

（3）根据已给数据，求得最优权值能够对数据进行分类

**二、实验环境：**

（1）硬件环境：

英特尔® 酷睿™ i5-7500U 处理器

256 GB PCIe® NVMe™ M.2 SSD

8 GB LPDDR3-1866 SDRAM

（2）软件环境：

Windows10家庭版64位操作系统

Python3.6

**三、实验内容**

**3.1基本梯度方法概述**

我们针对数据中分类错的数据进行批处理，假设这些错分的数据为y，a为权值，那么我们可以将问题变为，定义一个准则函数J（a），当a是解向量，即最优向量时，J（a）最小，这样就将问题简化为一个标量函数的极小化问题—通常可以用梯度法去解决。梯度下降法的原理非常简单，首先从一个随意选择的权向量a（1）开始，计算其梯度向量，下一个值则由a（1）向下降最陡的方向移一段距离而得到，即沿梯度的负方向：



其中

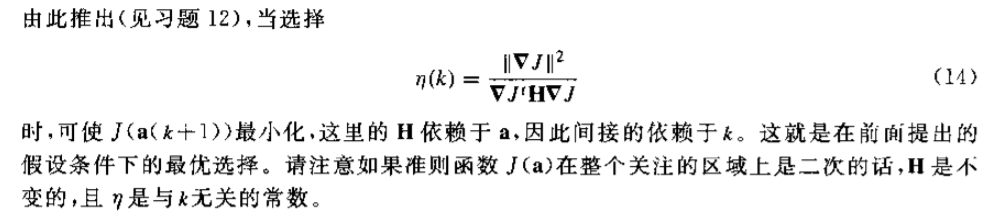
但是梯度下降法存在一定的问题，比如学习率的选择，如果学习率过大，那么有可能会发散，如果太小，那么收敛会非常缓慢。

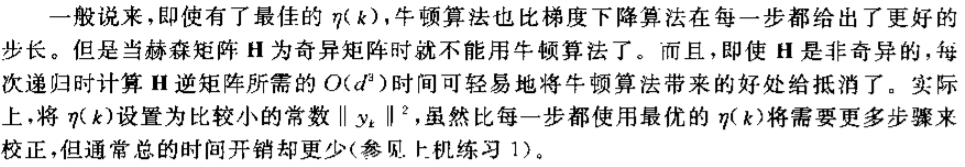
因此我们提出了牛顿梯度下降法

假设我们可以求得准则函数的在a（k）附近的二阶展开：



这里的H是荷森矩阵：

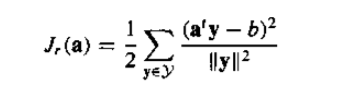


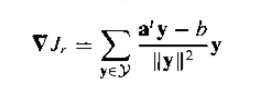


**在实验中，我们也发现，在采用最小误差方法的时候，荷森矩阵在求解过程大概率会变成奇异矩阵，导致算法无法继续下去。因此我们对牛顿梯度下降法采用的准则函数为松弛算法**

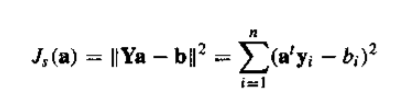
**采用的准则函数：**

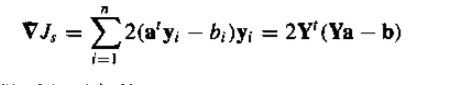
**松弛算法：**





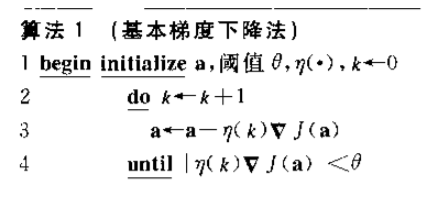
**最小误差平方法：**

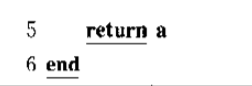


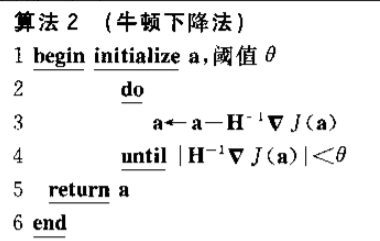


**3.2两种无参估计的算法**

设计程序的伪代码，实现图1所示：

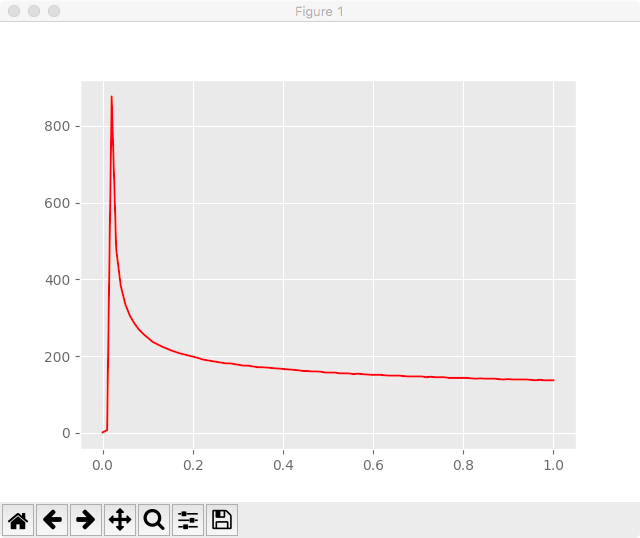




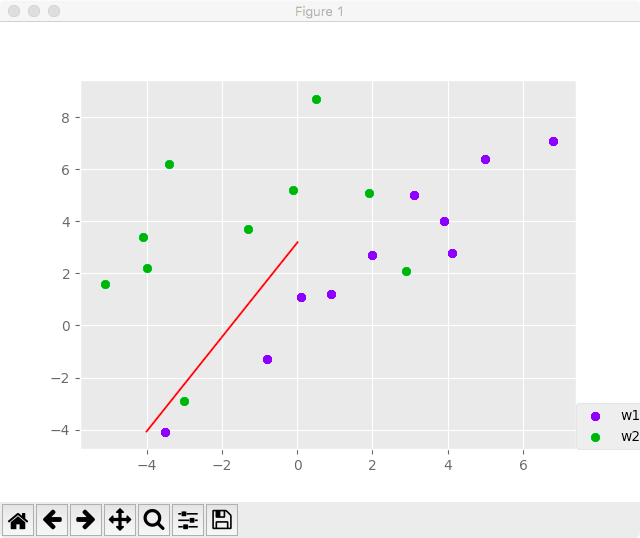


**图1.两种梯度下降法的伪代码**

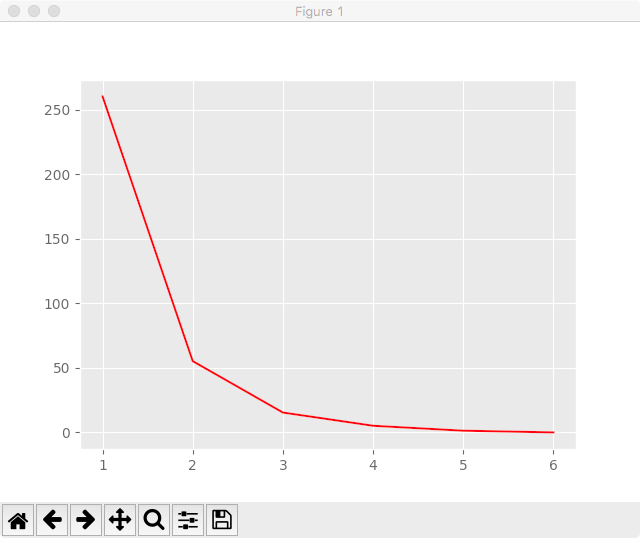
**四、实验结果**



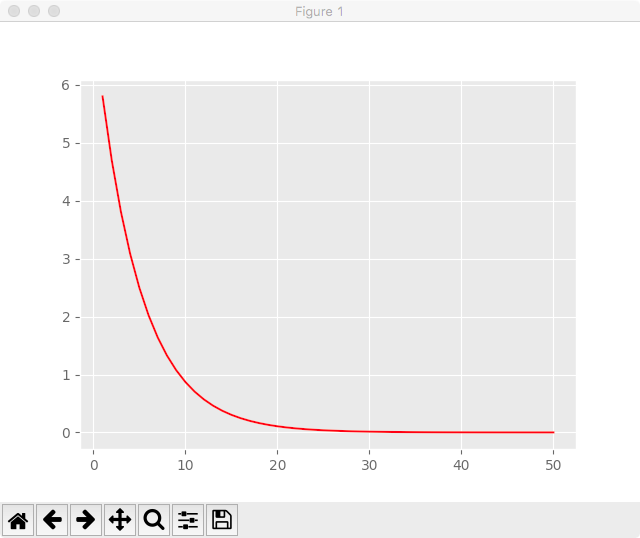
**图2. 基本梯度下降法 的 学习率-收敛时间曲线**



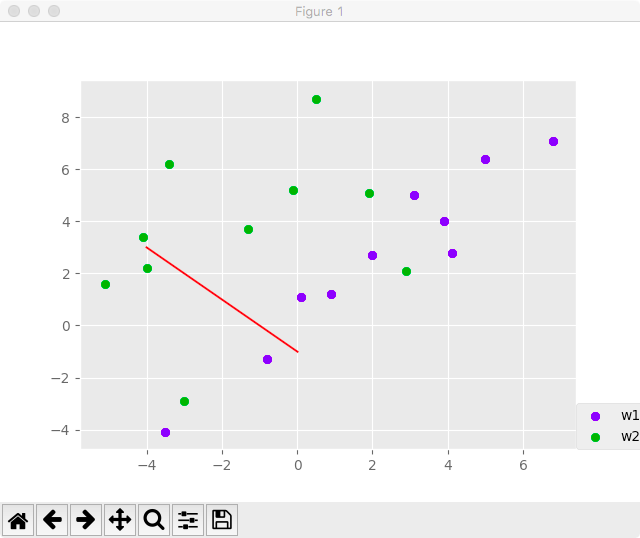
**图3. 基本梯度下降法 学习率 = 0.01 b=1 准则函数为最小平方误差**



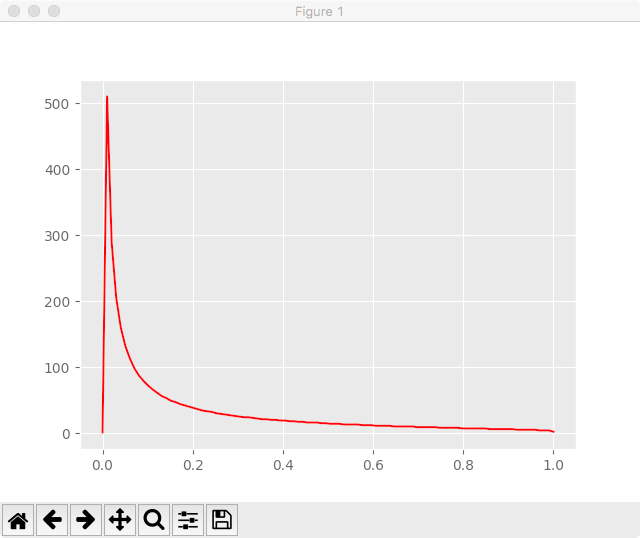
**图4. 基本梯度下降法的 迭代次数-准则函数曲线**



**图5. 牛顿梯度下降法 迭代次数-准则函数**



**图6. 牛顿梯度下降法 学习率为 0.1 b=0 准则函数采用松弛算法**



**图7. 牛顿梯度下降法的 学习率-收敛时间曲线**

***运算量估计***：从图2和图7的对比可以看到，牛顿梯度下降法的迭代次数少于基本梯度下降法，但是牛顿梯度下降法对荷森矩阵的逆矩阵的求解需要花费O（d^3）的时间复杂度，而基本梯度下降法，在计算时至少有两次循环，一次循环是错分的样本数量n，还有一次是迭代次数t，时间复杂度为（n\*t），并且随着n的增加，t也会增加，我们可以近似看作为O（n ^2）的时间复杂度。然而，基本梯度下降法当学习率设置为合适的值时，通常需要的开销更小。

***无法收敛的最小学习率*：**基本梯度下降法我们从图2可以看到为0.03附近

而牛顿梯度下降法则为0.05附近。

**五、总结与归纳**

**可以观察发现这次实验的数据都是线性不可分的，因此我们不能采用简单的感知器算法。因此我们根据题目要求采用了梯度下降算法。**

**首先对于基本梯度算法，我采用了松弛算法的准则函数以及最小平方误差方法的准则函数，经过实验发现最小平方误差方法的准则函数效果更好，分类结果更加准确。从图3可以看出。并且我设置的参数b=1，threshold= 0.0001， 当b=10-100时，分类效果更好。题目中要求学习率为0.1，但是当采用最小平方误差算法时，学习率为0.1为导致无法收敛。从图2中我们可以观察发现，当学习率在0.05-0.1附近时，迭代次数剧烈上升，可以认为是无法收敛。**

**对于牛顿梯度下降算法，缺点是很有可能出现奇异矩阵，当我采用最小平方误差方法的准则函数时，会大概率出现奇异矩阵，因此我采用了松弛方法的准则函数，并且使学习率为0.1，但是分类的结果非常不好，我分析认为是牛顿梯度下降法本身的缺陷，以及准则函数的选取问题，并且调整学习率并不能改善分类的结果。**

**因此在本实验中，采用基本梯度下降法会更好。**