

**课程报告**

徽标

描述已自动生成

报告题目： Lecture 2线性模型实验报告

学生姓名: 翁解语

学生学号: 522031910417

专 业: 信息安全

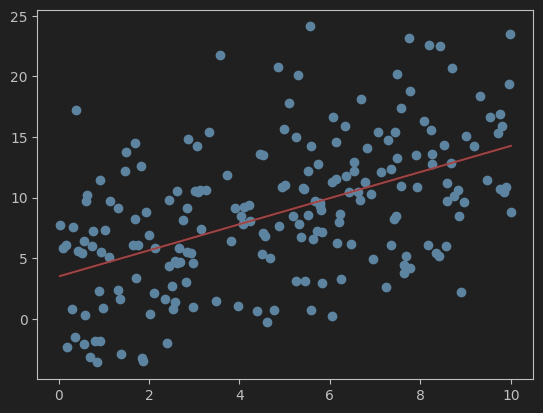
授课教师: 王士林

学 院: 网络空间安全学院

1. **线性模型**
   1. 最小二乘法

使用如下公式计算：

模型的拟合结果如下：



* 1. 梯度下降法

使用如下公式计算梯度并更新：

* + 1. 不同更新步长的影响:

在不同步长和迭代次数下的实验结果如表1.2.1所示，可以发现learning rate=0.01, iteration=1000时效果最好，主要是因为其他情况下更新不充分，欠拟合了。而最优配置适中的学习率（步长）确保梯度下降不会跳过最优解，充分的迭代次数确保模型能收敛到合理范围。

|  |  |  |
| --- | --- | --- |
| **更新步长** | **迭代次数** | **拟合结果** |
| 0.01 | 100 | weight: [[1.46317043]]  bias: [0.93041579]  14.295495550273472 |
| 0.01 | 1000 | weight: [[1.11925605]]  bias: [3.21861654]  13.27745900517356 |
| 0.001 | 100 | weight: [[1.51705025]]  bias: [0.23116278]  14.94729719686484 |
| 0.001 | 1000 | weight: [[1.45299078]]  bias: [0.9981451]  14.246029343351399 |

表1.2.1梯度下降拟合线性模型过程中不同步长和迭代次数的影响

* + 1. 不同初始化方式的影响:

取learning rate=0.01, iteration=1000，在不同初始化方式下的实验结果如表所示。初始均值过大 会导致初始参数偏离最优值，优化收敛慢，容易陷入局部最优解。初始方差过大 会导致梯度爆炸，训练不稳定，甚至计算溢出（NaN）。合理的初始化（如 0 均值, 0.1 方差）可以提高训练稳定性，加速收敛并提高模型的拟合效果。

|  |  |  |
| --- | --- | --- |
| **初始平均** | **初始方差** | **拟合结果** |
| 100 | 0.1 | weight: [[0.02689115]]  bias: [10.48655979]  18.97223931938023 |
| 0 | 1000 | weight: [[-1.04501744]]  bias: [17.61839895]  37.762199560848195 |
| 0 | 0.1 | weight: [[1.120591]]  bias: [3.20973454]  13.278808160117205 |
| 100 | 1000 | weight: [[nan]]  bias: [nan]  nan |

* + 1. （扩展）尝试添加不同的中止方式:

对于线性模型，还维度不高，在有解析解的情况下当然是二分法求最好；不同的中止方法意义不大，因为线性模型不会像很深的模型震荡或者过拟合。而且线性模型收敛太快，所以一半限定轮数或者根据loss变化早停就够用了。

|  |  |
| --- | --- |
| **方法&参数** | **拟合结果** |
| 基于损失变化的提前终止  tolerance = 1e-13 | Early stopping at iteration 71 with loss: 23.94224224367715  weight: [[1.38973363]]  bias: [1.4039071]  13.975589081657178 |
| 设定最大无改进次数  patience=10 | Stopping at iteration 14 due to no improvement  weight: [[1.18561605]]  bias: [1.10929151]  15.296326476382106 |

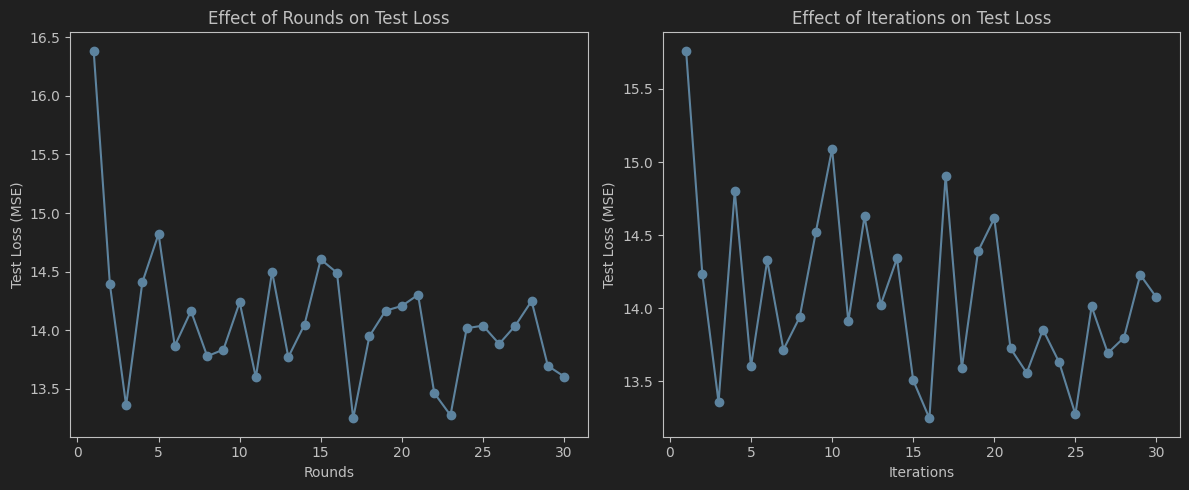
* 1. 模拟退火法

线性回归这种只有全局最优解的情况，真的有必要用模拟退化法吗？

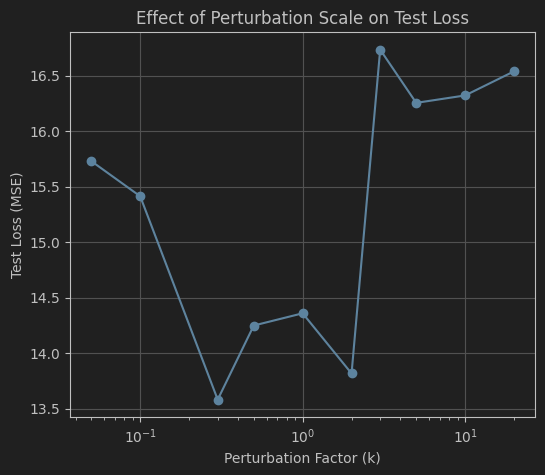
* + 1. 不同的迭代次数的影响：

随着iteration和round的提高，loss逐渐趋于稳定。

|  |  |  |
| --- | --- | --- |
| **iterations** | **round** | **拟合结果** |
| 10 | 10 | 14.275877658450272 |
| 3 | 10 | 15.489085876940644 |
| 10 | 3 | 16.053949911626262 |
| 3 | 3 | 19.366172191640423 |



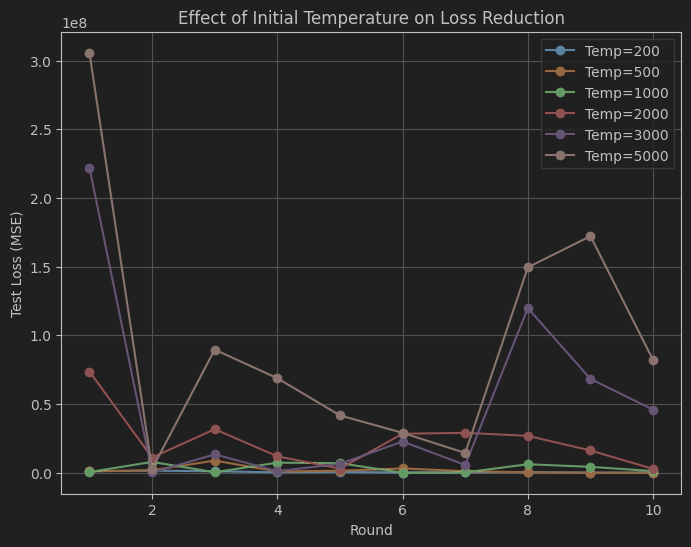
* + 1. 不同的扰动幅度的影响：



过小的扰动不足以有效探索新的解空间，而过大的扰动会破坏已经优化好的参数，导Loss都较高。

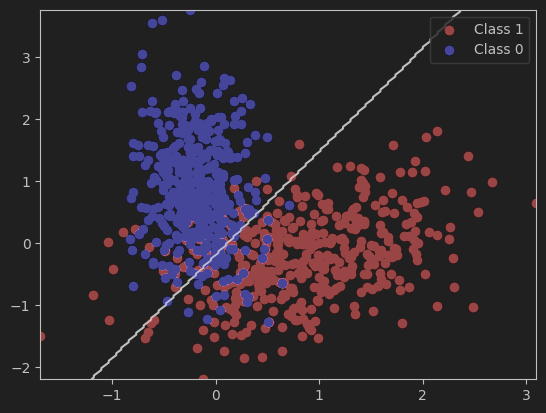
* + 1. （扩展）尝试不同的温度下降过程：

图中的temp代表初始温度系数。温度越高，下降过程越不稳定，loss越不稳定。



1. 逻辑回归
   1. 使用sklearn的实现

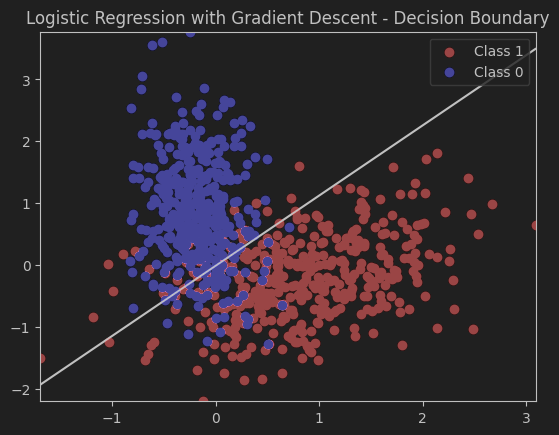
观察拟合情况，对比分类准确率



Accuracy: 0.91

* 1. 使用梯度下降法实现

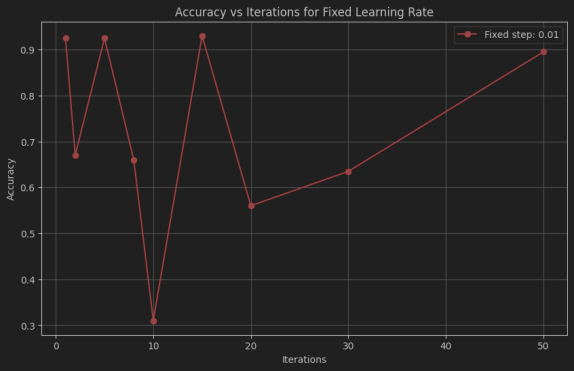
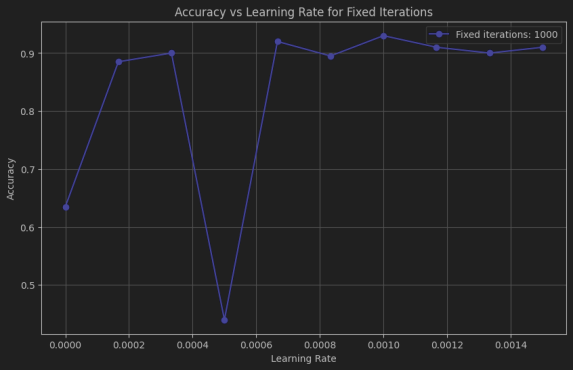
取step=0.01，iter=100



Accuracy: 0.925

* + 1. 尝试不同的初始化

|  |  |  |
| --- | --- | --- |
| **初始平均** | **初始方差** | **拟合结果** |
| 100 | 0.1 | Accuracy: 0.45 |
| 0 | 1000 | Accuracy: 0.17  （指数爆炸） |
| 0 | 0.1 | Accuracy: 0.92 |
| 100 | 1000 | Accuracy: 0.915  (指数爆炸） |

* + 1. 尝试不同的迭代次数和步长
  1. 固定分别固定step和iteration，得到的测试集上准确率如下图所示：
  2. 在合理的范围内，模型随着训练时间增长逐渐稳定。
  3. 
  4. 使用模拟退火法优化
     1. 尝试不同的迭代次数和扰动幅度

1. 高斯判别分析

对比不同分析方式的结果

* 1. 高斯线性判别分析

在线性情况下，对于两个类别使用相同的协方差，得到如下的分类边界图，这主要是因为

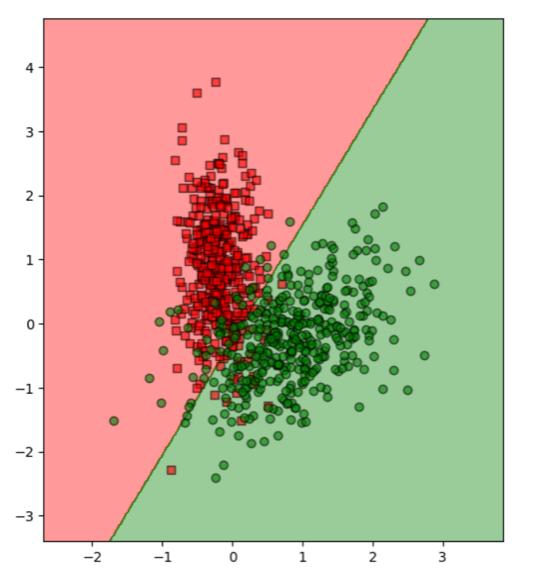


图3.1.1线性判别分析的分类边界图

* 1. 高斯二次判别分析

高斯二次判别(QDA)允许有不同的协方差，因此得到的分类边界是二次曲 线或者曲面，适合于分类非线性数据。

1. 感知器

参考之前的内容，实现一个单层的感知器模型。尝试使用梯度下降法来学习模型参数并实现分类任务。