实验报告实验一

15231100 曹强港

1 原始数据集

1.1 来源

共采集了 2 组数据，分别为 2 维数据和 3 维数据。

2 维数据根据直线 $2x_1 + 3x_2 - 4 = 0$ 创建，数据集中包含了 200 组数据，对于任意一组数据点 (x_1, x_2) 位于直线两侧，距离直线范围为 $(0, 10]$。数据集在记录时，添加了一维数据，成为 $(x_1, x_2, (-1|1))$，1 表示位于直线上方，-1 表示位于直线下方。数据集记录在"dataset1.csv"

2 维数据根据平面 $2x_1 + 3x_2 + 4x_3 - 5 = 0$ 创建，数据集中包含了 200 组数据，对于任意一组数据点 $(x_1, x_2, x3)$ 位于平面两侧，距离直线范围为 $(0, 10]$。数据集在记录时，添加了一维数据，成为 $(x_1, x_2, x_3, (-1|1))$，1 表示位于直线上方，-1 表示位于直线下方。数据集记录在"dataset2.csv"

2 数据预处理

为了感知机算法的简化性，将二维数据中和三维数据中位于下方的数据点 $(x_1, x_2, -1)$ 和 $(y_1, y_2, y_3, -1)$ 转换为 $(-x_1, -x_2, 1)$ 和 $(-y_1, -y_2, -y_3, 1)$

3 实验步骤

1. 初始化 weight vector $\hat{W}(0)$。二维数据的实验中，初始化为 (0,5,20)；三维数据的实验中，初始化为 (0,5,20,45)

2. 利用感知机算法进行迭代，每一次迭代中每一个错误分类的数据点均会对 $\hat{W}$ 产生影响。每隔 10 或 100 迭代次数，记录一次当前迭代中错误分类的数据点个数。

3. 设置不同的学习率为 0.01,0.1,0.2,0.3,...,0.9,0.99，分别记录步骤二中的数据，并记录不同学习率下收敛时的迭代次数和得到的直线方程 (r $\hat{W}$)。

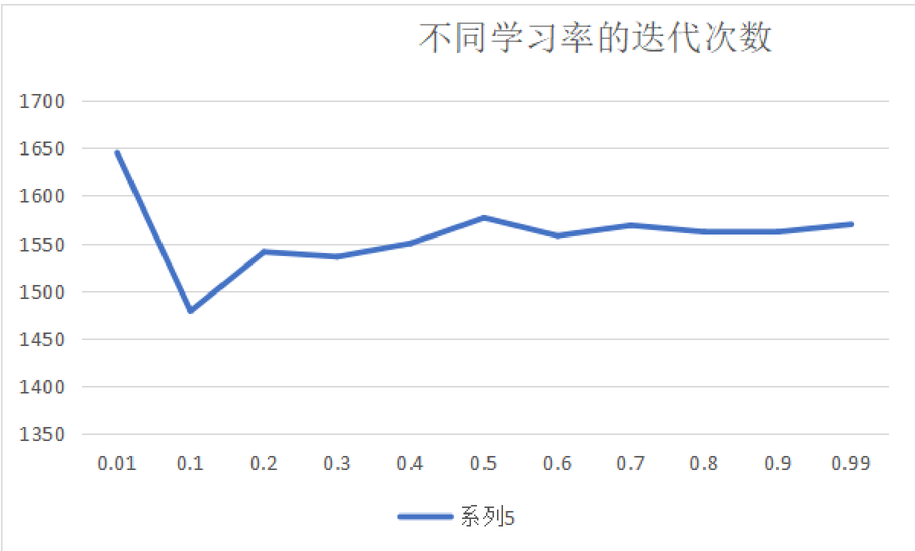
4. 对二维数据集和三维数据集，都进行步骤 1 到步骤 3 的操作，记录并分析数据。

4 实验数据

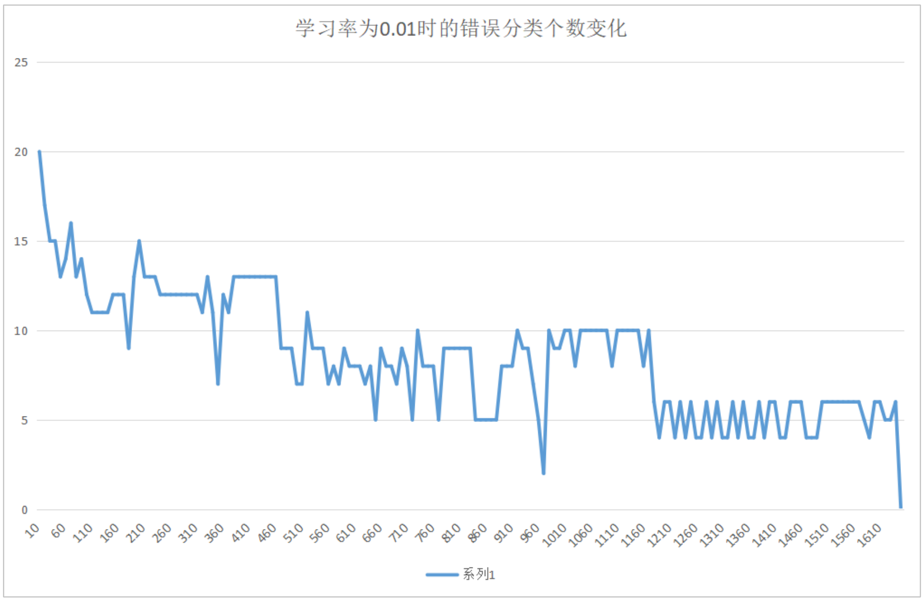
4.1 数据集 1

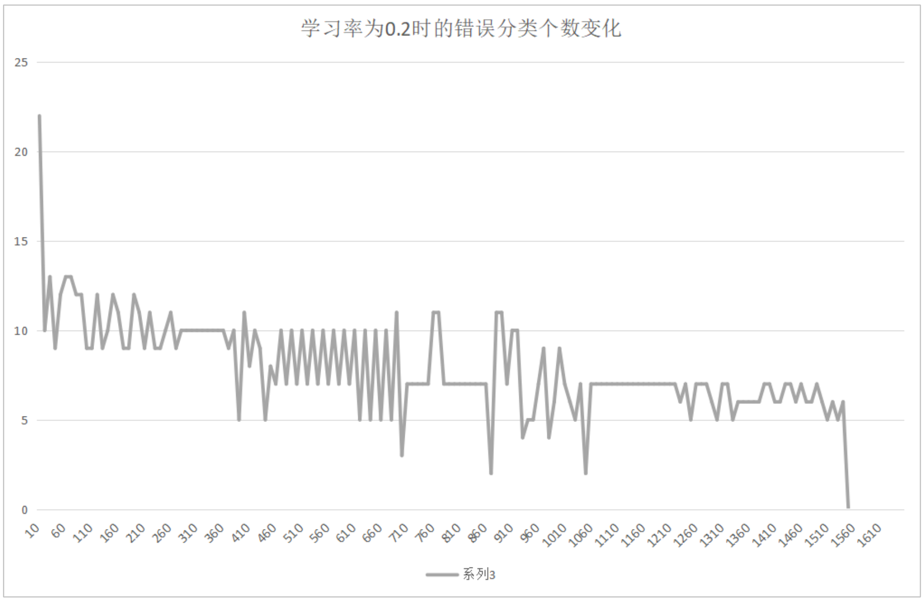
迭代次数

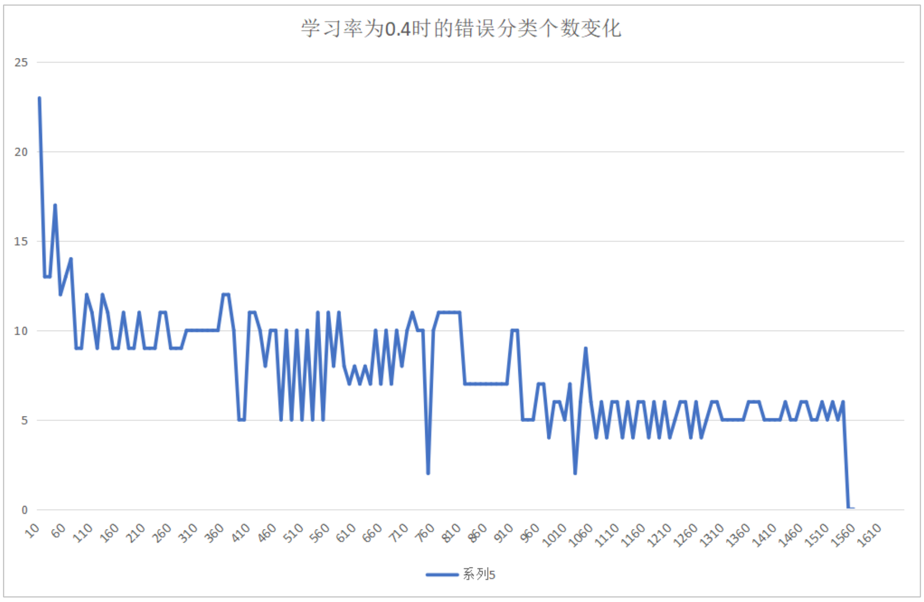
学习率	迭代次数	向量 (已处理)
0.01	1645	(-2.244471848,-3.381655343,4)
0.1	1479	(-2.251570102,-3.393645026,4)
0.2	1541	(-2.24924311,-3.389998459,4)
0.3	1536	(-2.24983023,-3.391170643,4)
0.4	1550	(-2.24555108,-3.38459983,4)
0.5	1577	(-2.250315653,-3.391966728,4)
0.6	1558	(-2.248698272,-3.389340623,4)
0.7	1569	(-2.247384684,-3.387281908,4)
0.8	1562	(-2.249579796,-3.39060821,4)
0.9	1562	(-2.249351384,-3.390192555,4)
0.99	1550	(-2.242391559,-3.379749671,4)

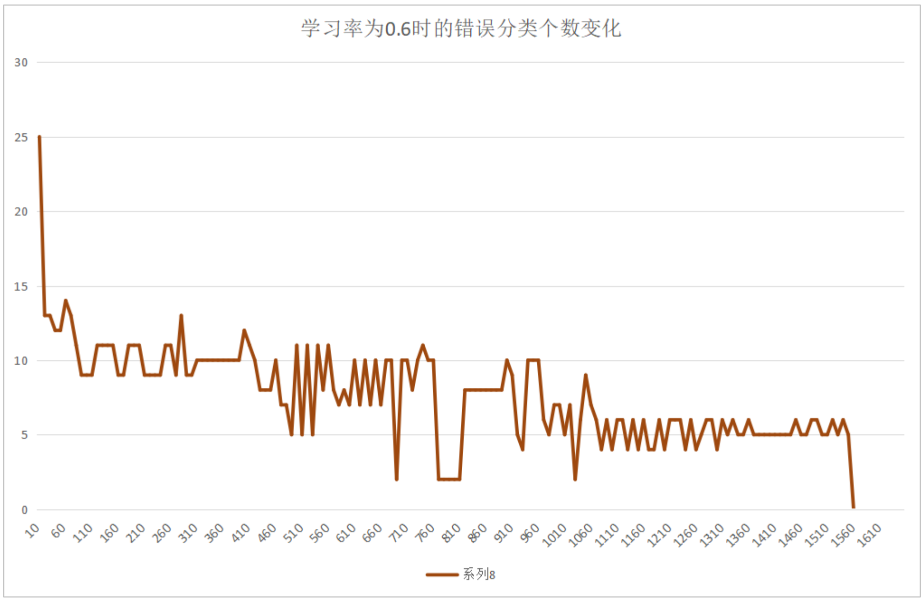


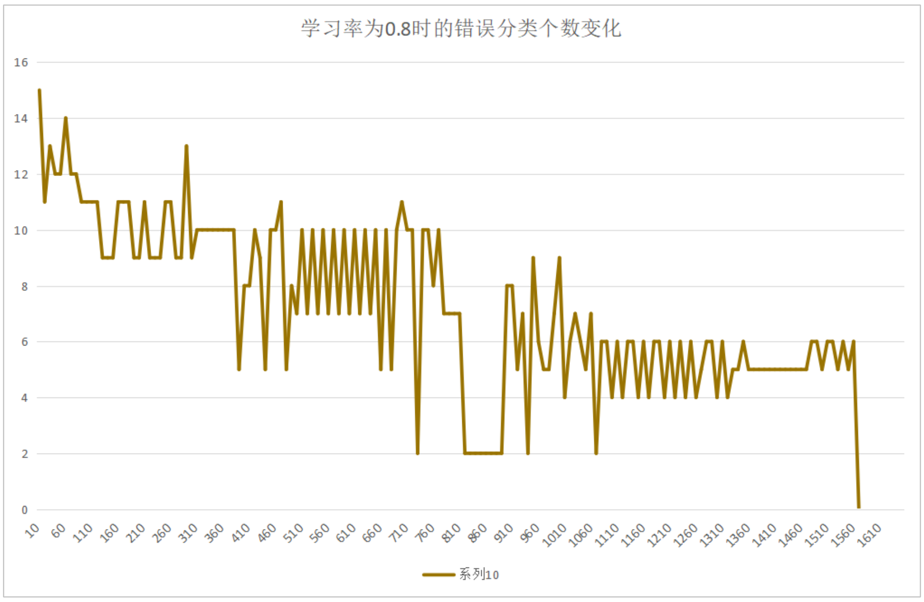
错误分类个数随迭代次数的变化

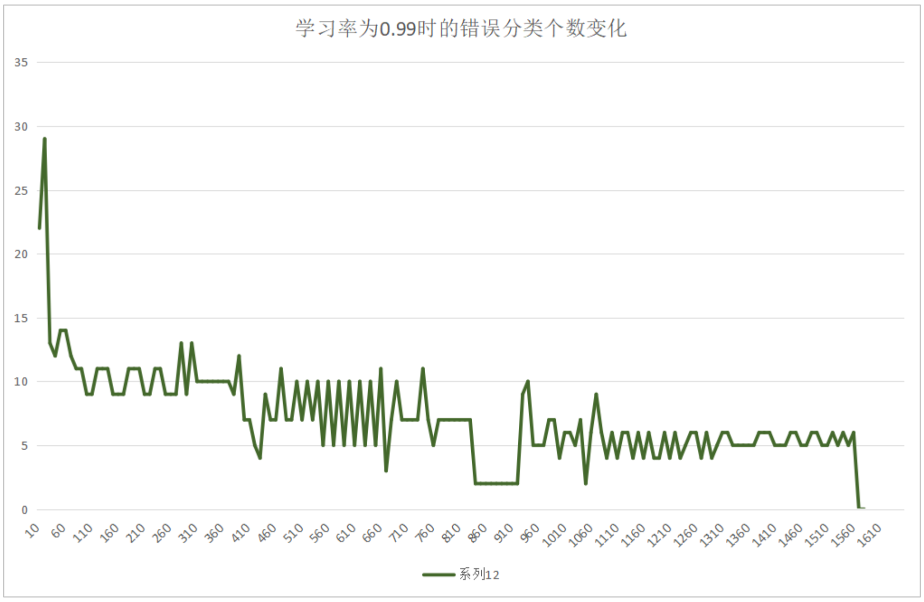

学习率为0.01时的错误分类个数变化
25
20
15
10
5
0
10 60 110 160 210 260 310 360 410 460 510 560 610 660 710 760 810 860 910 960 1010 1060 1110 1160 1210 1260 1310 1360 1410 1460 1510 1560 1610
系列1


学习率为0.2时的错误分类个数变化
25
20
15
10
5
0
10 60 110 160 210 260 310 360 410 460 510 560 610 660 710 760 810 860 910 960 1010 1060 1110 1160 1210 1260 1310 1360 1410 1460 1510 1560 1610
系列3


学习率为0.4时的错误分类个数变化
25
20
15
10
5
0
10 60 110 160 210 260 310 360 410 460 510 560 610 660 710 760 810 860 910 960 1010 1060 1110 1160 1210 1260 1310 1360 1410 1460 1510 1560 1610
系列5


学习率为0.6时的错误分类个数变化
30
25
20
15
10
5
0
10 60 110 160 210 260 310 360 410 460 510 560 610 660 710 760 810 860 910 960 1010 1060 1110 1160 1210 1260 1310 1360 1410 1460 1510 1560 1610
系列8

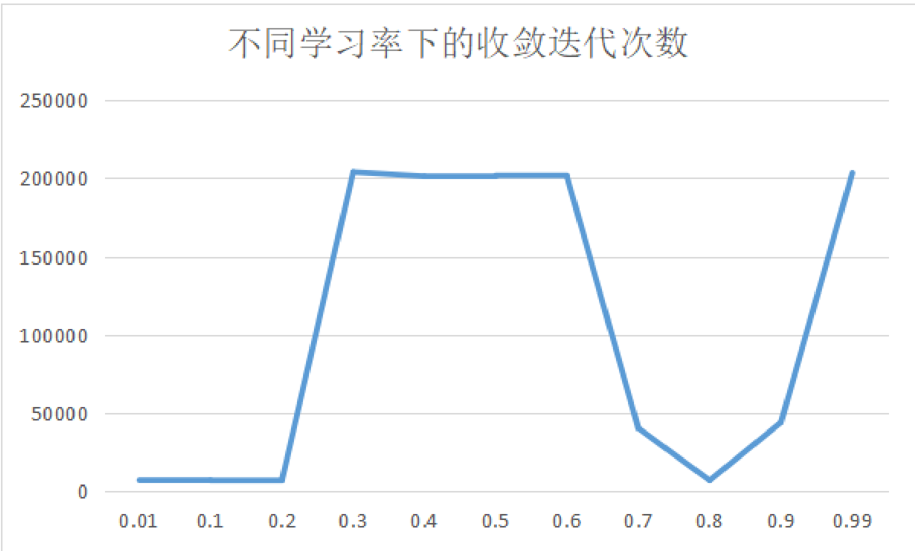

学习率为0.8时的错误分类个数变化
系列10


学习率为0.99时的错误分类个数变化
系列12

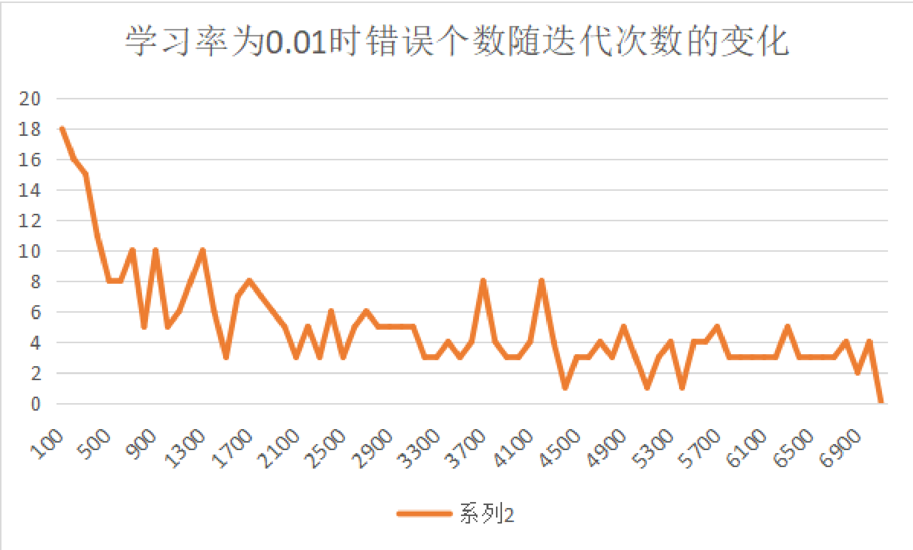
4.2 数据集 2

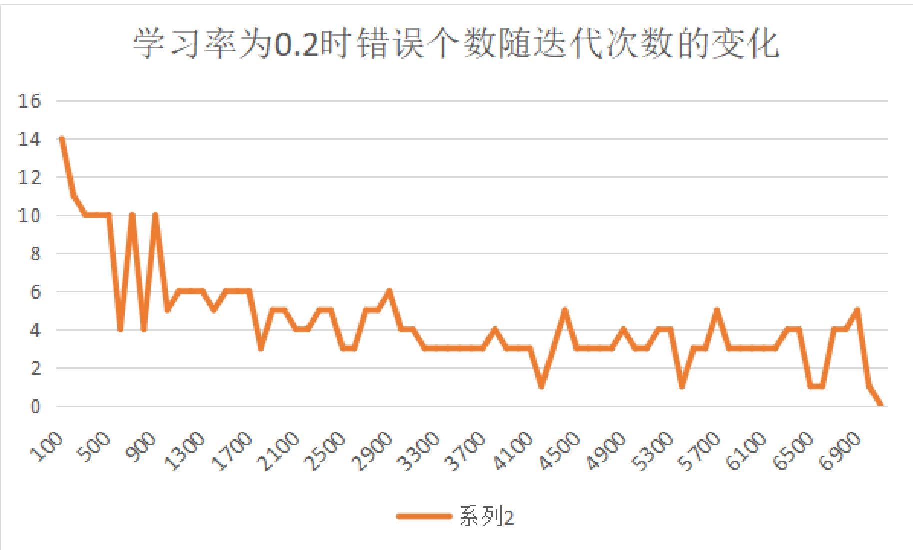
迭代次数

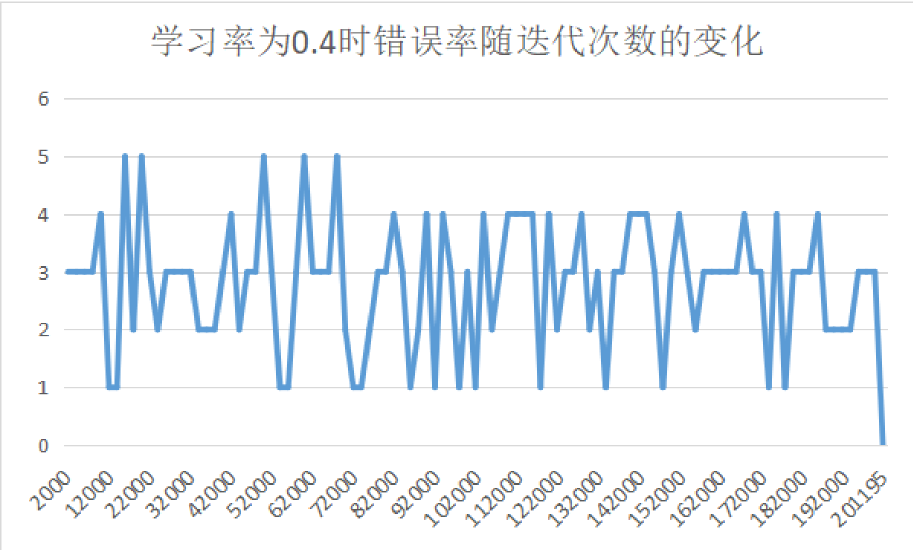
学习率	迭代次数	向量 (已处理)
0.01	7044	(1.942167763,2.991252798,3.919304133,-5)
0.1	6899	(1.935799178,2.978850037,3.905652123,-5)
0.2	7007	(1.94221003,2.990710071,3.919053093,-5)
0.3	203890	(1.994767722,3.031921564,3.999931218,-5)
0.4	201195	(1.995396729,3.033231016,4.001340657,-5)
0.5	201700	(1.995074649,3.032444268,4.000563569,-5)
0.6	200947	(1.995453479,3.033217345,4.00139915,-5)
0.7	40252	(2.005733815,3.033692721,4.016678797,-5)
0.8	6971	(1.937408701,2.987559465,3.911967239,-5)
0.9	44002	(2.006672689,3.029927748,4.016781729,-5)
0.99	203375	(1.995493868,3.033289313,4.001494139,-5)

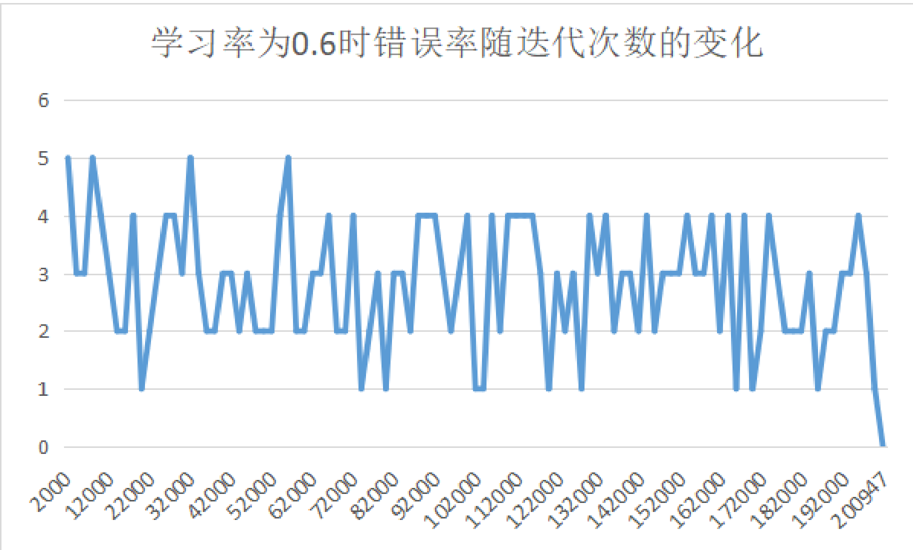


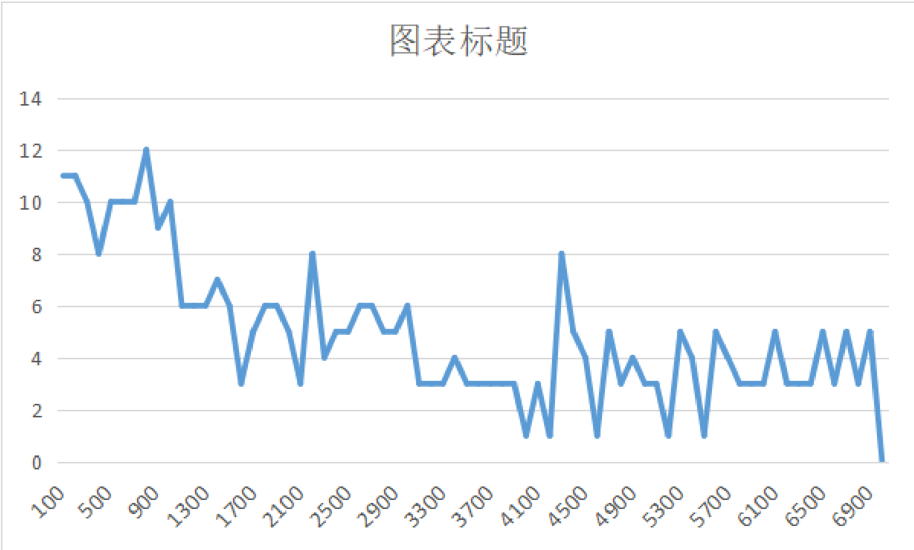
错误分类个数随迭代次数的变化

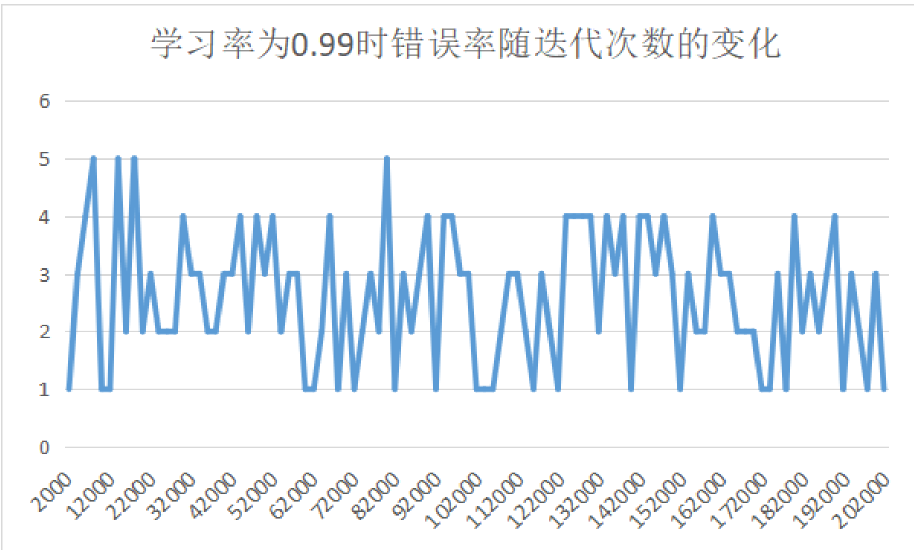

学习率为0.01时错误个数随迭代次数的变化
20
18
16
14
12
10
8
6
4
2
0
100
500
900
1300
1700
2100
2500
2900
3300
3700
4100
4500
4900
5300
5700
6100
6500
6900
系列2


学习率为0.2时错误个数随迭代次数的变化
16
14
12
10
8
6
4
2
0
100
500
900
1300
1700
2100
2500
2900
3300
3700
4100
4500
4900
5300
5700
6100
6500
6900
系列2


学习率为0.4时错误率随迭代次数的变化
6
5
4
3
2
1
0
2000
12000
22000
32000
42000
52000
62000
72000
82000
92000
102000
112000
122000
132000
142000
152000
162000
172000
182000
192000
201195


学习率为0.6时错误率随迭代次数的变化
6
5
4
3
2
1
0
2000
12000
22000
32000
42000
52000
62000
72000
82000
92000
102000
112000
122000
132000
142000
152000
162000
172000
182000
192000
200947


图表标题
14
12
10
8
6
4
2
0
100
500
900
1300
1700
2100
2500
2900
3300
3700
4100
4500
4900
5300
5700
6100
6500
6900


学习率为0.99时错误率随迭代次数的变化
6
5
4
3
2
1
0
2000
12000
22000
32000
42000
52000
62000
72000
82000
92000
102000
112000
122000
132000
142000
152000
162000
172000
182000
192000
202000

5 实验分析

5.1 现象和结论

1. 观察可知，当学习率逐渐由小增大时，感知机收敛时的迭代次数会随之先减后增

2. 在某些较大的学习率的情况下，迭代次数会突然达到较高的数量级。此时梯度下降的值较大，很难正好达到收敛。

3. 支持向量的初始值对总的迭代次数有影响，特别的，当初始值为零向量时，迭代次数与学习率无关

4. 错误分类的个数最初会随着迭代次数的增加而减少，学习率越高时减少越明显；然后可能会出现震荡（忽高忽低）, 学习率越高时震荡的幅度越大，越难达到收敛

5. 学习率越低或者越高并不一定能取得最优解，因为损失函数只是判别是否有数据点被分错，不关心所有分类的点是否与分类面相距较远 (“分得越开”)