



北京航空航天大学
B E I H A N G U N I V E R S I T Y

模式识别实验报告

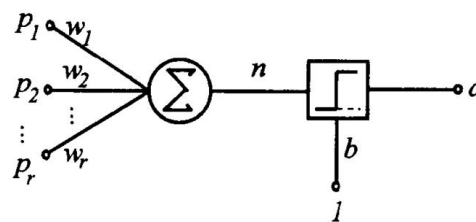
实验一、感知器学习

院（系）名称	自动化科学与电气工程学院
专 业 名 称	自动化
学 生 学 号	15031117
学 生 姓 名	柳天宇
指 导 教 师	

感知器是一种针对二分类问题非常简单有效的线性分类器，通过输入样本的特征，依据输出结果与期望值的差距不断调整权值的大小，可以在数据的特征空间找到一个超平面完成对数据的分类，对于线性可分的数据，训练结果将会收敛到一个能够对所有样本做出正确分类的权向量，从而达到非常好的效果。本次实验通过建立感知器模型对于具有二维特征的数据样本进行分类，验证了感知器学习在简单模式分类问题中的适用性。并研究了初始权值、学习率等参数的设置对于收敛速度和分类结果的影响。

1.感知器算法

如图所示，感知器是以数据的特征向量作为输入，计算输入向量的线性组合，与某个阈值进行比较决定最终的输出为 1 或 -1。



感知器计算法则如下：

$$o(x_1, x_2, \dots, x_n) = \begin{cases} 1 & (w_0 + w_1 * x_1 + w_2 * x_2 + \dots + w_n * x_n > 0) \\ -1 & (\text{otherwise}) \end{cases}$$

感知器的决策面是 n 维特征空间中的超平面，对于处于超平面一侧的数据，输出结果为 1，对于处于另一侧的数据输出 -1，作为一种线性分类器，其决策面的方程可表示为 $\vec{w} \bullet \vec{x} = 0$ 。训练感知器的目的就是通过调整权值的大小获得决策面，从而使其能够对输入向量达到期望的划分效果。

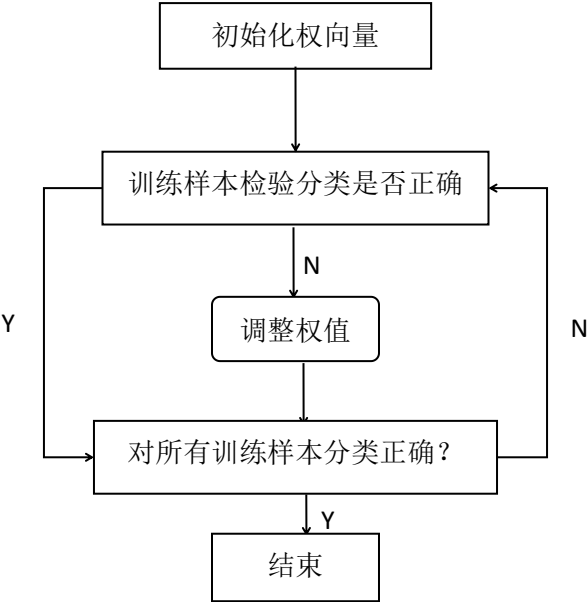
感知器的训练法则：

$$w(t+1) = w(t) + \frac{1}{2} \eta \{ (x(t) - x(t) \text{sgn}[\vec{w}(t) \bullet \vec{x}(t)]) \} = \begin{cases} w(t) & (\vec{w}(t) \bullet \vec{x} > 0) \\ w(t) + \eta x(t) & (\vec{w}(t) \bullet \vec{x} < 0) \end{cases}$$

若当前输出与期望值一致，则证明分类正确，不更新权值；若输出与期望不一致，则朝着使输出接近期望值的方向对权值进行调整，使结果更接近于正确的分类。学习率的作用是

控制每一次权值的调整程度。只要样本是线性可分的，在一个适宜的学习率下，经过有限的迭代次数总可以找到使分类结果完全正确的超平面。

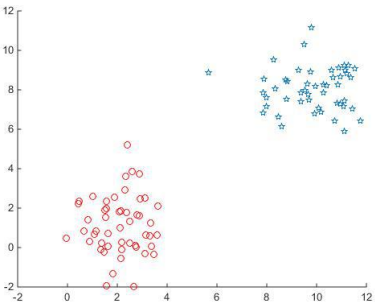
2. 算法流程图



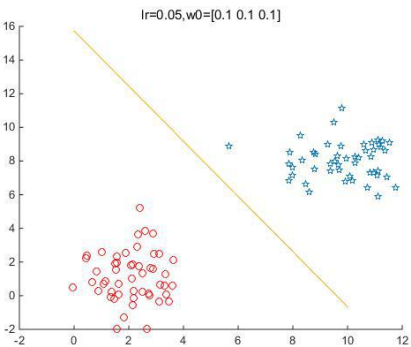
3. 实验内容及结果分析

3.1 编写感知器程序对线性可分的数据集进行二分类

样本如下图所示，两类样本用不同的颜色和形状表示，其中右上方蓝色样本为正样本，左下方红色样本为负样本：

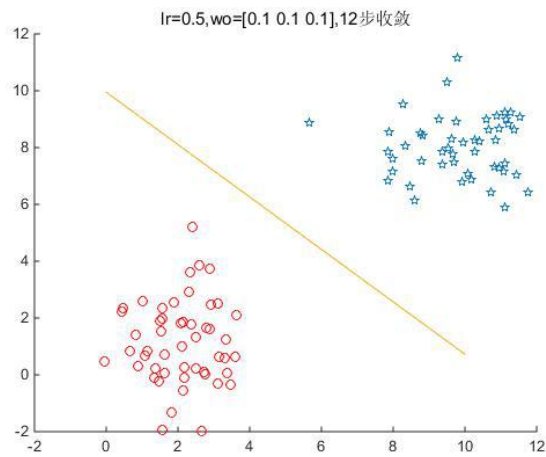


设置初始权值向量为[0.1 0.1 0.1]，选择学习速率 $lr=0.05$ ，运行后经过 14 步迭代后达到收敛，可以完成对所有训练样本的正确分类得到的分界面如下：

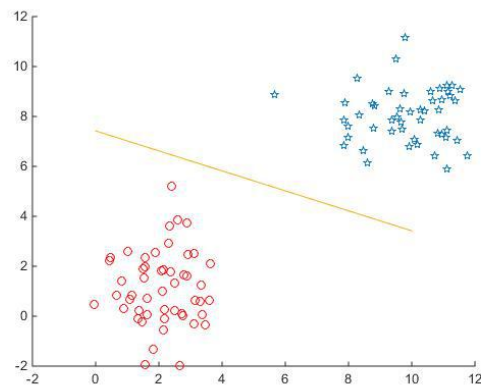


3.2 改变学习率，观察分类效果

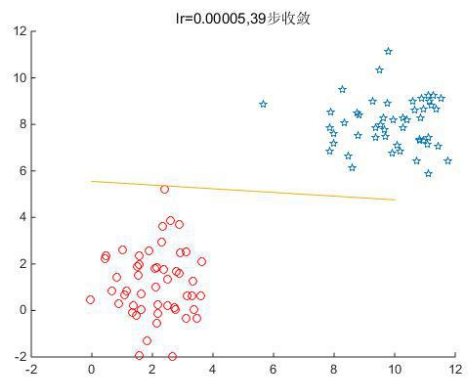
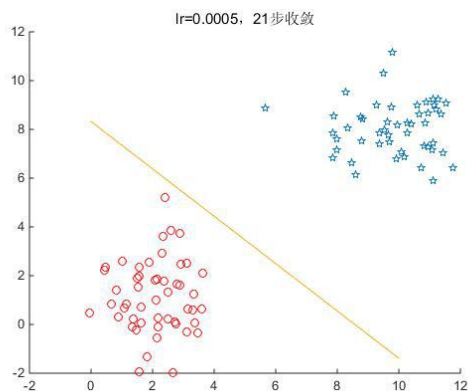
(1) 保持初始权向量不变，增大学习率，取 $lr=0.5$ ，得到如下结果，经过 12 步收敛



(2) 保持初始权向量不变，增大学习率，取 $lr=1$ ，得到如下结果，经过 13 步收敛



(3) 保持初始权值向量不变，减小学习速率，分别选取 $lr=0.0005$ ， 0.00005 ，得到如下结果：



不同学习率下达到收敛所需迭代次数（初始权向量设为[0.1 0.1 0.1]）：

学习率	1	0.5	0.05	0.005	0.0005	0.00005
达到收敛所需迭代次数	13	12	14	15	21	39

由运行结果可以看出，在选取 $lr=1$ 、 0.5 、 0.0005 、 0.00005 的情况下，模型均可达到收敛，完成对所有训练样本的正确分类。但不同学习率下完成分类的效率和效果是不同的。

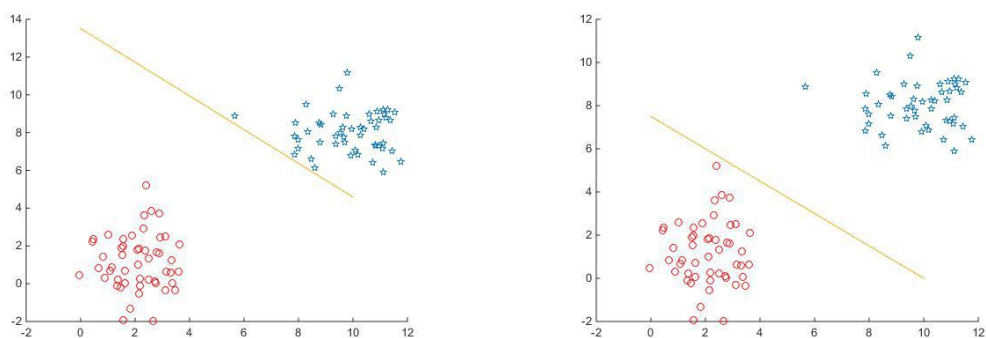
首先，从完成分类的效率看，当 $lr=0.5$ 时，需要 12 步迭代， $lr=0.05$ 时，需要经过 14 步迭代，随着 lr 的减小，达到收敛所需要的迭代次数渐渐增加，到 $lr=0.00005$ 时，迭代次数增加到 39 次。由此可见学习率过小会导致收敛速度变慢。

其次，尽管不同学习率下模型均能完成对于所有训练样本的正确分类，但不同学习率下解出的分界面是不同的，这也就意味着它们在训练集下表现相同，但如果使用新的数据对分界面的分类效果进行测试，得到的分类结果的正确率可能是不同的。从以上三幅图的决策边界可见，对于 $lr=0.5$ 、 0.05 和 0.0005 时，边界距离两类样本点的距离都较大，这样就保证了很好的容错性，使用这条决策边界对于新的数据进行类别的判断时，即使新样本点的特征与训练样本差异稍大也可以给出正确的类别判断，而对于第三种情况（ $lr=0.00005$ ）求解出的边界与两类训练样本之间的距离都较近，尤其是对于左下方的样本，边界甚至紧靠着其中一个样本点，这样的边界在使用测试集进行数据泛化时很容易产生错分。就是说，不同学习率下求出的分界面对于已知类别的样本均能正确分类，但它们对于未知类别的新样本类别的预测能力是不同的。

这组实验中选取较大的学习率在经过少量次数的迭代即可达到收敛，同时相比于小的学习率具有更好的分类性能。对于这组实验，由于两类样本之间差异性非常显著，很容易进行分类，因此学习率可以适当选大一些。对于这组样本的分类，如果在初始权值保持为[0.1 0.1 0.1]时，选取 $lr=0.5$ 时，只需 12 步就可达到收敛，说明 0.5 是对于该组数据比较适合的学习率。这并不意味着我们在选取学习率时需要一成不变地遵循某种规律，学习率的大小需要根据分类样本自身的差异性来确定。但以上的运行结果的不同说明了学习率对于分类器的性能的一个非常关键的参数，针对不同的分类样本，选取适宜的学习率对于满足算法实时性的要求以及分类精度的要求是至关重要的。

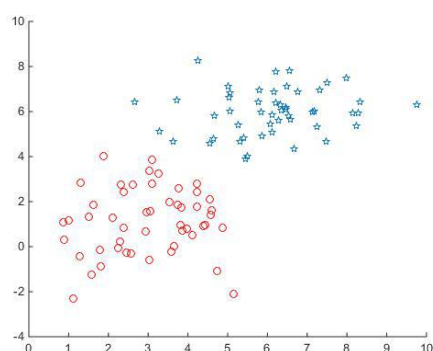
3.3 改变初始权向量

初始权向量的选择决定了超平面起始位置与目标位置之间的距离远近,也会直接影响感知器分类的速度,取 $w_0=[2\ 5\ -3]$, $lr=0.05$ 运行结果如下左图,经过 7 步达到收敛,极端地,如果取 $w_0=[3\ 4\ -32]$,运行 1 步就达到收敛(如下右图),因为最初选取的权向量就可完成对所有训练样本的正确分类,因此无需再调整权值。

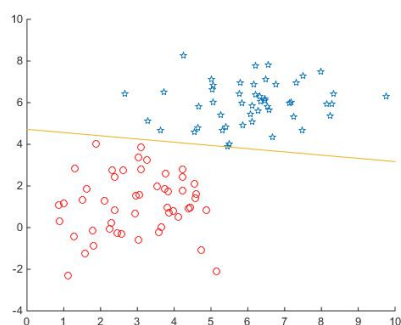


3.4 更换不同的数据集重复实验

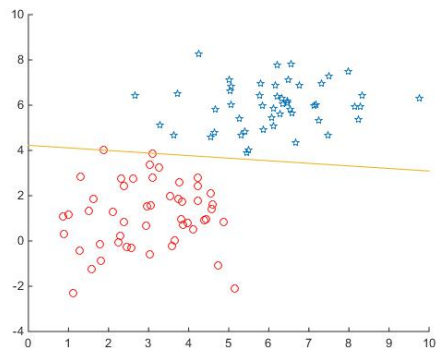
使用如图所示的数据样本对感知器模型进行训练,相比于步骤一中的样本,这个训练集的类型距离更小,从直观上看进行分类的难度相比步骤一更大。



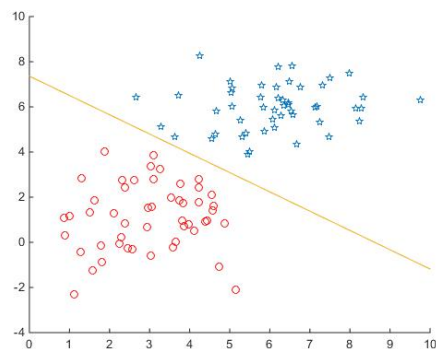
(1) 依旧设置初始权值向量为 $[0.1\ 0.1\ 0.1]$, $lr=0.05$ 时运行结果如图,需要 36 步迭代达到收敛



(2) $lr=0.5$ 时运行结果如图，需要 45 步迭代达到收敛



(3) $lr=0.01$ 时运行结果如图，需要 60 步迭代达到收敛



学习率	0.5	0.1	0.05	0.001	0.0005	0.00005
达到收敛所需迭代次数	45	46	36	60	43	79

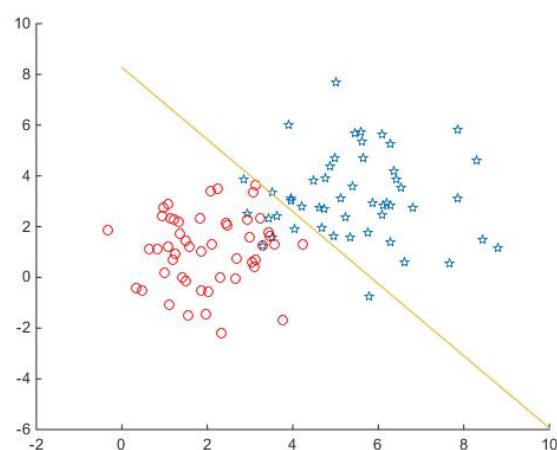
在该组样本下进行实验，可见完成对于所有样本的正确分类所需的迭代次数普遍多于第一次的实验。实验结果表明由于样本类间的差异性减小，可分性降低，对于感知器模型的训练就需要更长的时间（需要更多的迭代次数来修正权值）。同时，第一组样本在 $lr=0.5$ 时所需迭代次数最少，而该组实验中 $lr=0.05$ 时对应最快的收敛速度，说明了对于不同的数据集进行分类，适宜的学习率是不同的。对比两组实验，可以发现对于类间距离较小、类间差异性不太突出的数据我们应当适当减小学习率，而对于第一组实验中那样差异非常显著，很容易区分的数据，可以适当增大学习率。

在第二组实验中，调整学习率对于收敛速度的影响相比于第一组更为显著，这也就意味着我们对于可分性很强的样本分类时，在不同的学习率下往往都能获得较快的收敛速度；而

当我们面临的分类样本较为复杂，差异性较小时，学习率这个参数的设置对于边界的求解就显得非常重要了。

若采用线性不可分的数据集，无论怎样调整学习率、初始权值，都无法完成对所有样本的正确分类，这是感知器模型本身的缺陷，但有时可以通过最小错误法则来减少错误率，做到对大多数样本的正确分类

在线性不可分数据集下感知器分类的表现（经过 2000 次迭代）：



4.总结

感知器学习是处理线性可分的二分类问题的一种有监督学习的方法，本次实验根据感知器学习法则建立感知器模型针对于不同的数据集进行分类，求取特征空间上的超平面，获得了对已知类别样本正确的分类结果。实验过程中通过改变学习率、初始权值的大小研究了这些参数对于分类器性能的影响，使用不同样本分布对分类器的表现进行测试与评价，加深了对于感知器模型的理解，也发现了这种模型在处理线性不可分问题时的显著缺陷。

Github 网址:<https://github.com/ECHOLIuty/PR-report>