《模式识别》第3 次作业

姓名：谷绍伟　　　学号：202418020428007

1 计算与证明

1. 现有四个来自于两个类别的二维空间中的样本，其中第一类的两个样本为 $(1,4)^T$ 和 $(2,3)^T$，第二类的两个样本为 $(4,1)^T$ 和 $(3,2)^T$。这里，上标 T 表示向量转置。若采用规范化增广样本表示形式，并假设初始的权向量 $a=(0,1,0)^T$，**其中向量 a 的第三维对应于样本的齐次坐标**。同时，假定梯度更新步长 η_k 固定为 1。试利用批处理感知准则函数方法求解线性判别函数 $g(\mathbf{x})=a^T\mathbf{x}$ 的权向量 $\mathbf{a}$，（注："规范化增广样本表示"是指对齐次坐标表示的样本进行规范化处理）。

答：第一类样本规范化后为 $x_1=(1,4,1)^T$，$x_2=(2,3,1)^T$，第二类样本规范化后为 $x_3=(-4,-1,-1)^T$，$x_4=(-3,-2,-1)^T$。初始化权向量为 $a=(0,1,0)^T$，根据批处理感知准则函数方法，有：

$$x_1: \quad (1,4,1)^T\cdot(0,1,0)^T=4>0$$

$$x_2: \quad (2,3,1)^T\cdot(0,1,0)^T=3>0$$

$$x_3: \quad (-4,-1,-1)^T\cdot(0,1,0)^T=-1<0$$

$$x_4: \quad (-3,-2,-1)^T\cdot(0,1,0)^T=-2<0$$

希望判别结果全为正数，x_3 和 x_4 被错误分类，更新步长 $\eta_k=1$，则权向量更新为：

$$a=(0,1,0)^T+(-4,-1,-1)^T+(-3,-2,-1)^T=(-7,-2,-2)^T$$

再次计算分类结果：

$$x_1: \quad (1,4,1)^T\cdot(-7,-2,-2)^T=-7-8-2=-17<0$$

$$x_2: \quad (2,3,1)^T\cdot(-7,-2,-2)^T=-14-6-2=-22<0$$

$$x_3: \quad (-4,-1,-1)^T\cdot(-7,-2,-2)^T=28+2+2=32>0$$

$$x_4: \quad (-3,-2,-1)^T\cdot(-7,-2,-2)^T=21+4+2=27>0$$

希望判别结果全为正数，x_1 和 x_2 被错误分类，更新步长 $\eta_k=1$，则权向量更新为：

$$a=(-7,-2,-2)^T+(1,4,1)^T+(2,3,1)^T=(-4,5,0)^T$$

再次计算分类结果：

$$x_1: \quad (1,4,1)^T \cdot (-4,5,0)^T = -4+20+0 = 16 > 0$$

$$x_2: \quad (2,3,1)^T \cdot (-4,5,0)^T = -8+15+0 = 7 > 0$$

$$x_3: \quad (-4,-1,-1)^T \cdot (-4,5,0)^T = 16-5+0 = 11 > 0$$

$$x_4: \quad (-3,-2,-1)^T \cdot (-4,5,0)^T = 12-10+0 = 2 > 0$$

分类结果均为正，无错误分类，则权向量为：

$$a = (-4,5,0)^T$$

2. 对于多类分类情形，考虑 one-vs-all 技巧，即构建 c 个线性判别函数：

$$g_i(\mathbf{x}) = \mathbf{w_i}^T\mathbf{x} + w_{i0}, i = 1,2,\cdots,c$$

此时的决策规则为：对于 $j \neq i$，如果 $g_i(\mathbf{x}) > g_j(\mathbf{x})$，$\mathbf{x}$ 则被分为 ω_i 类。现有三个二维空间内的模式分类器，其判别函数为：

$$g_1(\mathbf{x}) = -x_1 + x_2$$

$$g_2(\mathbf{x}) = x_1 + x_2 - 1$$

$$g_3(\mathbf{x}) = -x_2$$

试画出决策面，指出为何此时不存在分类不确定性区域。

答：若 $\mathbf{x}$ 属于 ω_i 类，则有：

$$g_1(\mathbf{x}) > g_2(\mathbf{x}) \Rightarrow x_1 < \frac{1}{2}$$

$$g_1(\mathbf{x}) > g_3(\mathbf{x}) \Rightarrow x_1 < 2x_2$$

若 $\mathbf{x}$ 属于 ω_2 类，则有：

$$g_2(\mathbf{x}) > g_1(\mathbf{x}) \Rightarrow x_1 > \frac{1}{2}$$

$$g_2(\mathbf{x}) > g_3(\mathbf{x}) \Rightarrow x_1 + 2x_2 > 1$$

若 $\mathbf{x}$ 属于 ω_3 类，则有：

$$g_3(\mathbf{x}) > g_1(\mathbf{x}) \Rightarrow x_1 > 2x_2$$

$$g_3(\mathbf{x}) > g_2(\mathbf{x}) \Rightarrow x_1 + 2x_2 < 1$$

因此，决策边界交于 $x_1 = \frac{1}{2}$，$x_2 = \frac{1}{4}$ 的位置，且不存在分类不确定性区域，如图1所示：

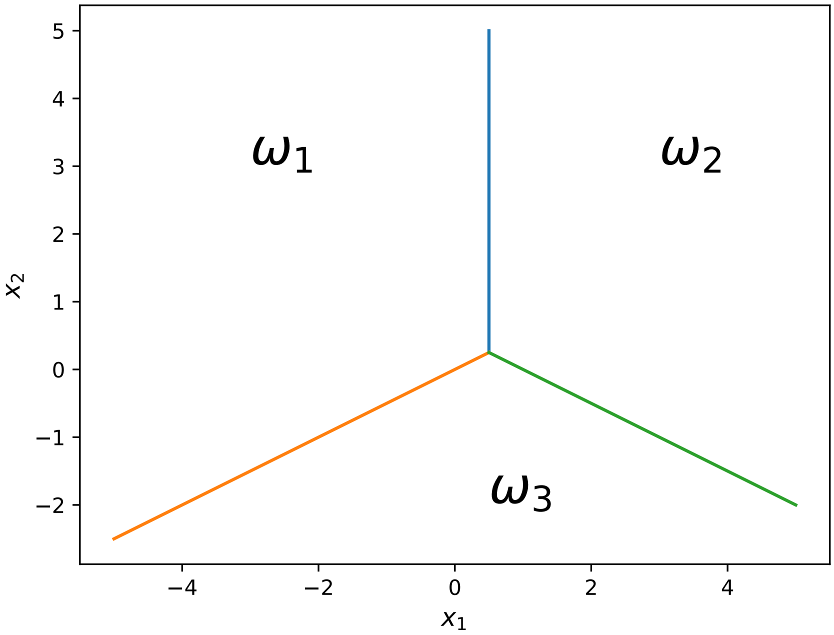

Figure 1: 分类决策面

2 计算机编程

本章所使用的数据如下表：

sample	ω_1		ω_2		ω_3		ω_4	
	x_1	x_2	x_1	x_2	x_1	x_2	x_1	x_2
1	0.1	1.1	7.1	4.2	-3.0	-2.9	-2.0	-8.4
2	6.8	7.1	-1.4	-4.3	0.5	8.7	-8.9	0.2
3	-3.5	-4.1	4.5	0.0	2.9	2.1	-4.2	-7.7
4	2.0	2.7	6.3	1.6	-0.1	5.2	-8.5	-3.2
5	4.1	2.8	4.2	1.9	-4.0	2.2	-6.7	-4.0
6	3.1	5.0	1.4	-3.2	-1.3	3.7	-0.5	-9.2
7	-0.8	-1.3	2.4	-4.0	-3.4	6.2	-5.3	-6.7
8	0.9	1.2	2.5	-6.1	-4.1	3.4	-8.7	-6.4
9	5.0	6.4	8.4	3.7	-5.1	1.6	-7.1	-9.7
10	3.9	4.0	4.1	-2.2	1.9	5.1	-8.0	-6.3

1. Write a program to implement the “batch perception” algorithm.

a. Starting with $a = 0$, apply your program to the training data from ω_1 and ω_2. Note that the number of iterations required for convergence（即记录下收敛的步数）。

b. Apply your program to the training data from ω_3 and ω_2. Again, note that the number of iterations required for convergence.

答：程序见附件中的 batch_perception.py 文件，运行结果如下表所示：

任务	Iteration	权向量	更新步长
a	23	$(-0.304, 0.341, 0.340)$	0.01
b	16	$(-0.041, 0.049, 0.019)$	0.01

2. Implement the Ho-Kashyap algorithm and apply it to the training data from ω_1 and ω_3. Repeat to apply it to the training data from ω_2 and ω_4. Point out the training errors, and give some analyses.

答：程序见附件中的 Ho_Kashyap.py 文件，设置更新因子为 0.01，初始误差为 1，初始 $b_{ini} = 0.01$，最小 $b_{min} = 0.001$，最大迭代次数为 150000 次，运行结果如下表所示：

任务样本	结果	迭代次数	权向量
ω_1、ω_3	不收敛	150000	$(0.003454, -0.002515, 0.004814)$
ω_2、ω_4	收敛	23076	$(0.005339, 0.004746, 0.03688)$

样本 ω_2 和 ω_4 正常收敛，但样本 ω_1 和 ω_3 不收敛，对样本分布进行可视化，结果如图 2。可见 ω_2 和 ω_4 是线性可分的，但 ω_1 和 ω_3 不是线性可分，因此 Ho-Kashyap 算法在分类 ω_1 和 ω_3 上不收敛。

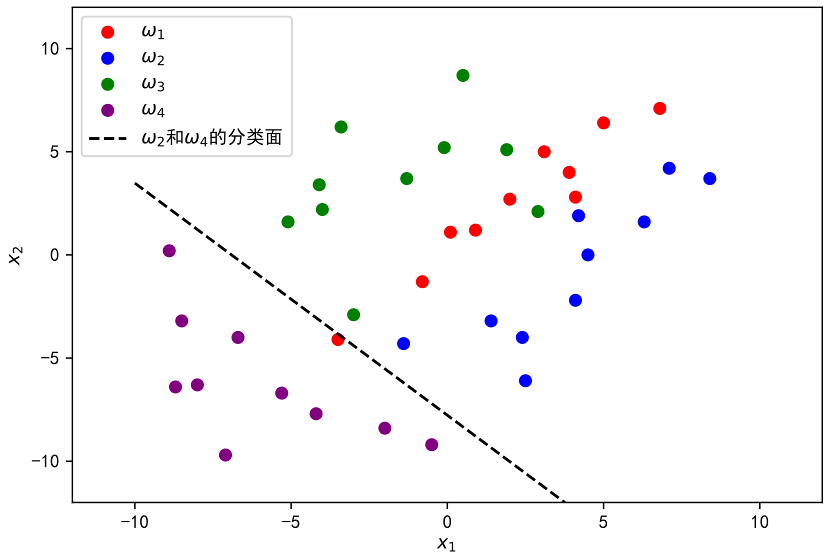

Figure 2: 分类可视化

3. 请写一个程序，实现 MSE 多类扩展方法。每一类用前 8 个样本来构造分类器，用后两个样本作测试。请写出主要计算步骤，并给出你的正确率。

答：代码见文件中 MSE.py 文件，主要步骤如下：

1. 构造分类器，用前 8 个样本来构造分类器，用后两个样本作测试，回归值使用 one-hot 编码。
2. 根据训练样本计算出权值 W。
3. 根据此权值和测试样本计算出预测值 E。
4. 对预测值取 argmax，得到分类输出，判断正确率。

经过测试，分类结果正确率为 100%。