Pattern Recognition

University of Chinese Academy of Sciences

Fall 2023

Shiming Xiang, Gaofeng Meng

Homework 3

Chenkai GUO

2023.11.23

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 $\mathbf{a}=(0,1,0)^T$，其中向量 $\mathbf{a}$ 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 $g(\mathbf{y})=\mathbf{a}^T\mathbf{y}$ 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）。

   由题可得，两类样本的规范化齐次增广表示为：第一类 $(1,4,1)^T,(2,3,1)^T$，第二类 $(-4,-1,-1)^T,(-3,-2,-1)^T$，采用 *batch perception* 算法的损失函数为 $J_p(\mathbf{a})=\sum\limits_{y\in Y}(-\mathbf{a}^T\mathbf{y})$，其中 Y 为错分样本集合，梯度更新准则为 $\mathbf{a}_{k+1}=\mathbf{a}_k+\eta_k\sum\limits_{y\in Y}\mathbf{y}$，初始化权向量 $a_0=(0,1,0)^T$，学习率 $\eta_k=1$

   (*1*) 第一次更新：

   $\mathbf{a}_0^Ty_1=4>0,\mathbf{a}_0^Ty_2=3>0,\mathbf{a}_0^Ty_3=-1<0,\mathbf{a}_0^Ty_4=-2<0$

   更新权向量 $\mathbf{a}_1=\mathbf{a}_0+y_3+y_4=(-7,-2,-2)^T$

   (*2*) 第二次更新：

   $\mathbf{a}_0^Ty_1=-16<0,\mathbf{a}_0^Ty_2=-22<0,\mathbf{a}_0^Ty_3=32>0,\mathbf{a}_0^Ty_4=27>0$

   更新权向量 $\mathbf{a}_2=\mathbf{a}_1+y_1+y_2=(-4,5,0)^T$

   (*3*) 第三次更新：

   $\mathbf{a}_0^Ty_1=16>0,\mathbf{a}_0^Ty_2=7>0,\mathbf{a}_0^Ty_3=11>0,\mathbf{a}_0^Ty_4=2>0$

   此时所有样本分类正确，停止更新权向量，故最终权向量 $\mathbf{a}=(-4,5,0)^T$

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：

$$g_i(\mathbf{x})=\mathbf{w}_i^T\mathbf{x}+w_{i0},\quad i=1,2,\dots,c$$

此时的决策规则为：对 $j\leq i$，如果 $g_i(x)>g_j(x)$，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：

$g_1(x) = -x_1 + x_2$

$g_2(x) = x_1 + x_2 - 1$

$g_3(x) = -x_2$

试画出决策面，指出为何此时不存在分类不确定性区域。

考虑 *one-vs-all* 技巧：

(*1*) 对 ω_1 类，需满足 $g_1(x) > g_2(x), g_1(x) > g_3(x)$

$$\begin{cases} -x_1 + x_2 > x_1 + x_2 - 1 & \Rightarrow x_1 < \frac{1}{2} \\ -x_1 + x_2 > -x_2 & \Rightarrow x_2 > \frac{1}{2}x_1 \end{cases}$$

(*2*) 对 ω_2 类，需满足 $g_2(x) > g_1(x), g_2(x) > g_3(x)$

$$\begin{cases} x_1 + x_2 - 1 > -x_1 + x_2 & \Rightarrow x_1 > \frac{1}{2} \\ x_1 + x_2 - 1 > -x_2 & \Rightarrow x_2 \ -\frac{1}{2}x_1 + \frac{1}{2} \end{cases}$$

(*3*) 对 ω_3 类，需满足 $g_3(x) > g_1(x), g_3(x) > g_2(x)$

$$\begin{cases} -x_2 > -x_1 + x_2 & \Rightarrow x_2 < \frac{1}{2}x_1 \\ -x_2 > x_1 + x_2 - 1 & \Rightarrow x_2 < -\frac{1}{2}x_1 + \frac{1}{2} \end{cases}$$

根据以上决策面可绘制一下决策面示意图

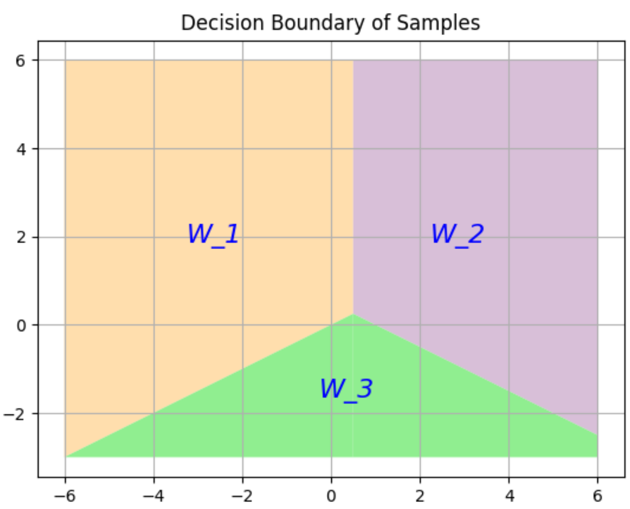

图 1: 样本决策面示意图

显然，所有决策面交于一点，故不存在分类不确定性区域

计算编程题

sample	ω_1 x_1	ω_1 x_2	ω_2 x_1	ω_2 x_2	ω_3 x_1	ω_3 x_2	ω_4 x_1	ω_4 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

图 2: 数据集

3. *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
   (b). Apply your program to the training data from ω_3 and ω_3. Again, note that the number of iterations required for convergence

4. *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*

5. 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。

   3(a) the training results is as follows:

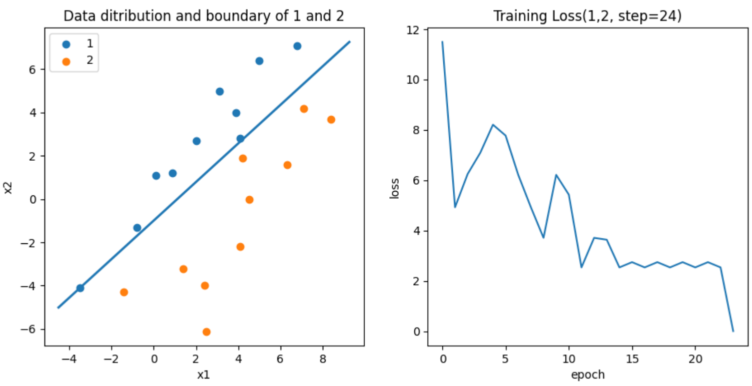

图 3: batch perception between w1 and w2

3(b) the training results is as follows:

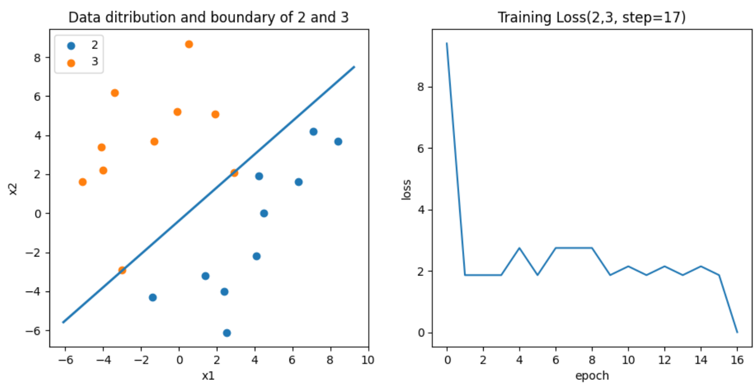

图 4: batch perception between w2 and w2

4.the training results is as follows:

ω_1 *and* ω_3 分类中有错误样本，因此不是线性可分的，故训练次数溢出 k_{max}

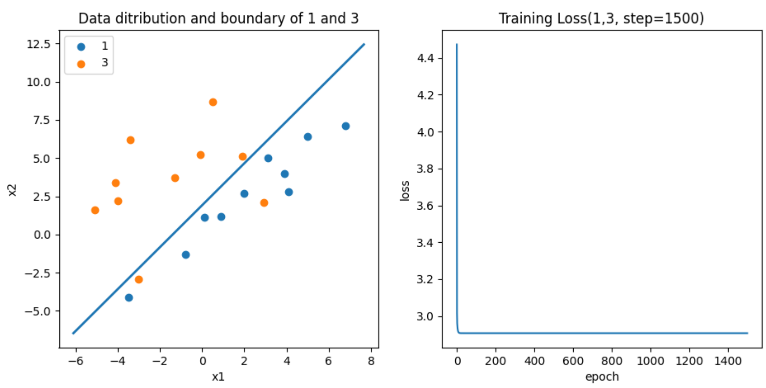

图 5: Ho-Kashyap classification between w2 and w3

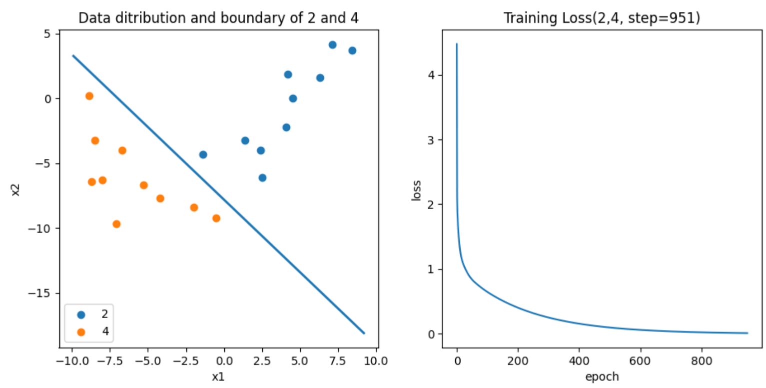

图 6: Ho-Kashyap classification between w2 and w4

5.the MSE-expand model acc is: 100.00%

代码见附件(*homework3.py*)