第 5 章：线性判别函数

第一部分：计算与证明

1． 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。假设初始的权向量 $a=(0,1)^T$，且梯度更新步长η_k 固定为 1。试利用批处理感知器算法求解线性判别函数 $g(y)=a^Ty$ 的权向量。

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：

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

此时的决策规则为：对 $j \neq i$, 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。

3． 已知模式样本集：$\omega_1 = \{(0,0)^T, (1,1)^T\}$, $\omega_2 = \{(0,1)^T, (1,0)^T\}$。采用误差平方准则算法（即 Ho-kashyap 算法）验证它是线性不可分的。（提示：迭代时η_k固定取 1,初始 $b=(1,1,1,1)^T$）

4． Consider the hyperplane used in discrimination:

(a) Show that the distance from the hyperplane $g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 = 0$ to the point $\mathbf{x}_a$ is $|g(\mathbf{x}_a)|/||\mathbf{w}||$ by minimizing $||\mathbf{x}-\mathbf{x}_a||^2$ subject to the constraint $g(\mathbf{x}) = 0$.（提示本题需要证明两点：其一，点 $\mathbf{x}_a$ 到超平面 $g(\mathbf{x}) = 0$ 的距离为$|g(\mathbf{x}_a)|/||\mathbf{w}||$；其二，该距离是位于超平面 $g(\mathbf{x}) = 0$ 上使目标函数$||\mathbf{x}-\mathbf{x}_a||^2$ 最小的点 $\mathbf{x}$ 到点 $\mathbf{x}_a$ 的距离。）

(b) Show that the projection of $\mathbf{x}_a$ onto the hyperplane is given by (即证明点 $\mathbf{x}_a$ 到超平面 $g(\mathbf{x}) = 0$ 的投影 $\mathbf{x}_p$ 为如下式子):

$$\mathbf{x}_p = \mathbf{x}_a - \frac{g(\mathbf{x}_a)}{||\mathbf{w}||^2}\mathbf{w}$$

第二部分：计算机编程

本章所使用的数据：

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 (see page 44 or 45 in PPT).

(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.

(c). Explain the difference between the iterations required in the two cases.

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.

3. Consider relaxation methods as described in the PPT. (See the slides for the "Batch Relaxation with Margin" algorithm and page 62 in PPT for the "Single Sample Relaxation with Margin" algorithm):

(a) Implement the batch relaxation with margin, set b = 0.1 and initialize **a** = **0**, and apply it to the data in ω_1 and ω_3. Plot the criterion function as a function of the number of passes through the training set.

(b) Repeat for b = 0.5 and $\mathbf{a}_0$ = **0** (namely, initialize **a** = **0**). Explain qualitatively any differences you find in the convergence rates.

(c) Modify your program to use single sample learning. Again, Plot the criterion function as a function of the number of passes through the training set.