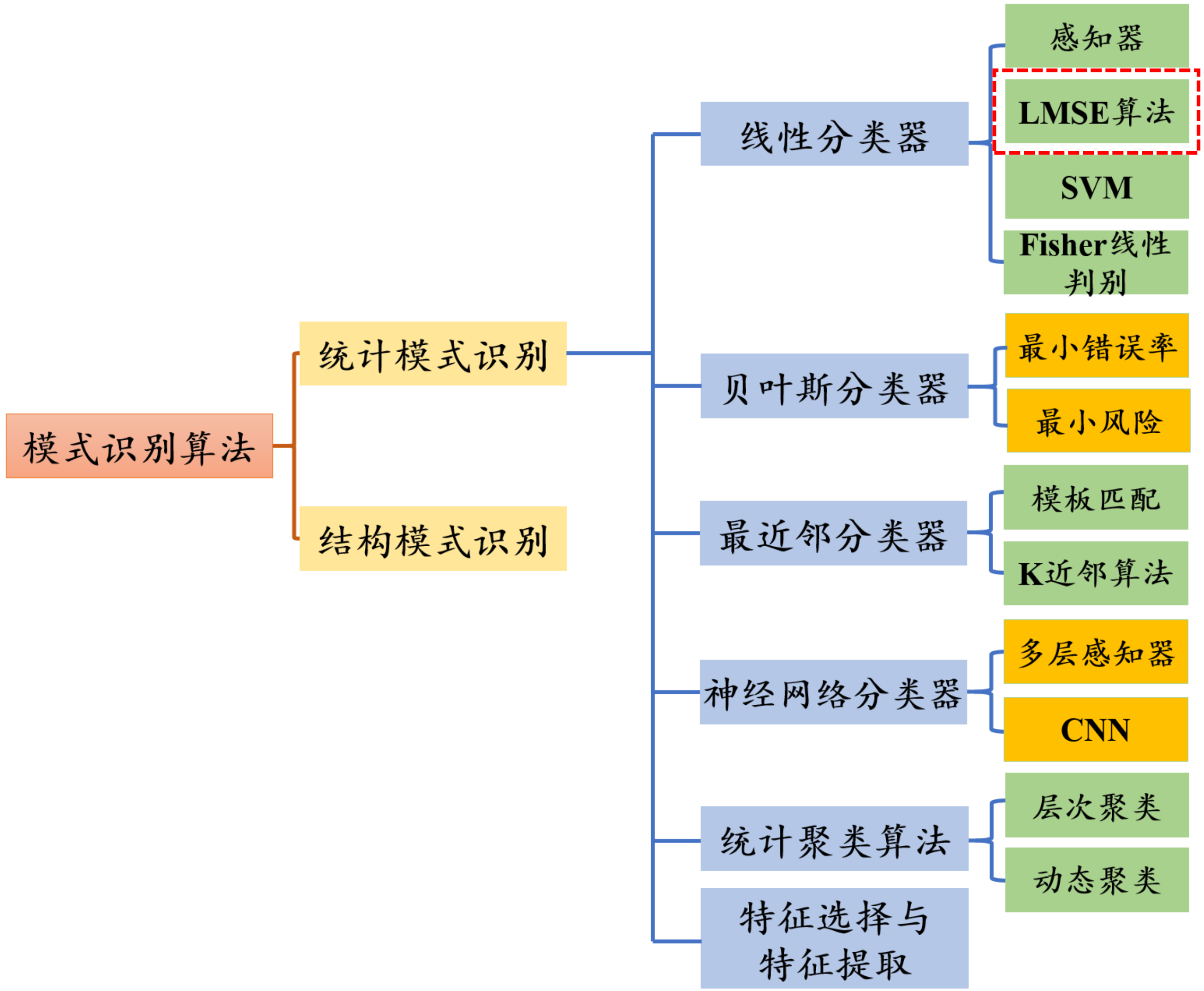
模式识别课程

最小平方误差判别

张俊超

中南大学
航空航天学院

模式识别-线性分类器



模式识别算法
统计模式识别
结构模式识别
线性分类器
感知器
LMSE算法
SVM
Fisher线性判别
贝叶斯分类器
最小错误率
最小风险
最近邻分类器
模板匹配
K近邻算法
神经网络分类器
多层感知器
CNN
统计聚类算法
层次聚类
动态聚类
特征选择与特征提取

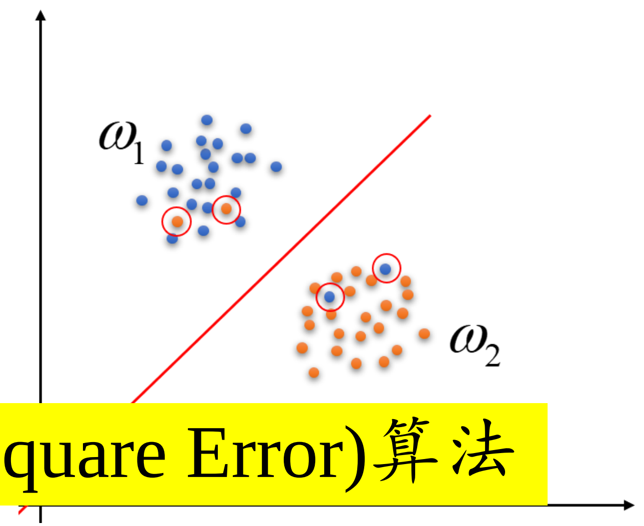
线性分类器–LMSE算法

◆感知器算法：

- 不能处理线性不可分的问题
- 不能判断样本集是否线性可分

希望：

- 能对输入训练样本集是否线性可分作出判断
- 样本线性不可分时，能够给出某种准则下的最优解(比如：错分样本尽可能少)



最小平方误差(Least Mean Square Error)算法

线性分类器-LMSE算法

$$\boldsymbol{\alpha}^T \mathbf{y}_i > 0, (i = 1, 2, \ldots, N)$$

$$\Big\Downarrow \text{转为等式约束}$$

$$\boldsymbol{\alpha}^T \mathbf{y}_i = b_i > 0, (i = 1, 2, \ldots, N)$$

$$\Big\Downarrow \text{矩阵形式，松弛变量}\boldsymbol{b}$$

$$\mathbf{Y}\boldsymbol{\alpha} = \mathbf{b}$$

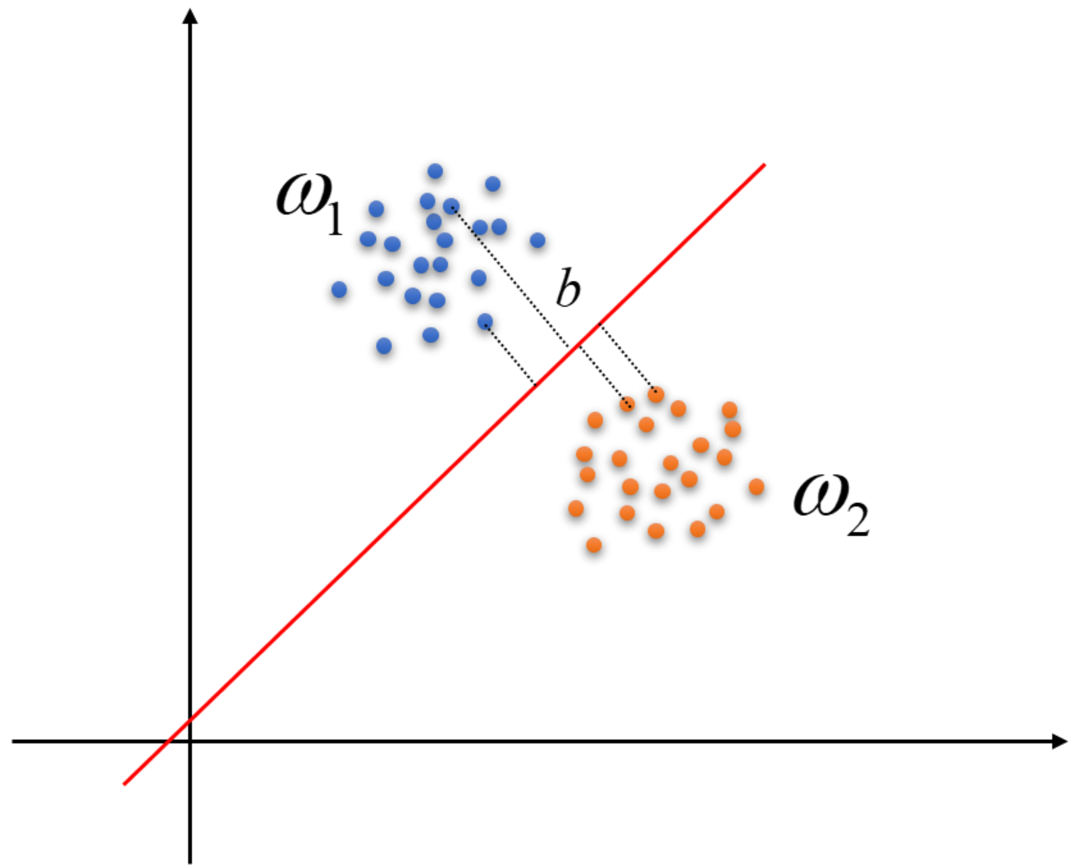
其中

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_N^T \end{bmatrix} \overset{d=n+1}{=} \begin{bmatrix} y_{11} & \cdots & y_{1d} \\ \vdots & \ddots & \vdots \\ y_{N1} & \cdots & y_{Nd} \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix}$$

线性分类器 –LMSE算法

物理含义： $\|\mathbf{w}\| = 1$

代数含义： $\mathbf{Y\alpha} = \mathbf{b}, \mathbf{Y} \in \mathbb{R}^{N \times d}, \boldsymbol{\alpha} \in \mathbb{R}^{d}, \mathbf{b} \in \mathbb{R}^{N}$



线性分类器－LMSE算法

求解超定方程组($N>d$)：

$$\mathbf{Y\alpha}=\mathbf{b},\mathbf{Y}\in\mathbb{R}^{N\times d},\boldsymbol{\alpha}\in\mathbb{R}^{d},\mathbf{b}\in\mathbb{R}^{N}$$

求解方程组的最小平方误差解，即：

$$\begin{aligned}\boldsymbol{\alpha}^{*}&=\arg\min_{\boldsymbol{\alpha}} J_{s}\left(\boldsymbol{\alpha}\right)\\&=\arg\min_{\boldsymbol{\alpha}}\left\|\mathbf{Y\alpha}-\mathbf{b}\right\|^{2}\end{aligned}$$

求解方法：伪逆法和梯度下降法

线性分类器-LMSE算法

- 伪逆法

$$\boldsymbol{\alpha}^* = \arg\min_{\boldsymbol{\alpha}} J_s(\boldsymbol{\alpha})$$

$$= \arg\min_{\boldsymbol{\alpha}} \|\mathbf{Y}\boldsymbol{\alpha} - \mathbf{b}\|^2$$

$$\nabla_{\mathbf{x}}(\mathbf{x}^T \mathbf{A}\mathbf{x}) = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}$$

$$\nabla_{\mathbf{x}}(\mathbf{y}^T \mathbf{x}) = \nabla_{\mathbf{x}}(\mathbf{x}^T \mathbf{y}) = \mathbf{y}$$

$$\nabla J_s(\boldsymbol{\alpha}) = \frac{\partial J_s(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = 0$$

$$\nabla J_s(\boldsymbol{\alpha}) = 2\mathbf{Y}^T(\mathbf{Y}\boldsymbol{\alpha} - \mathbf{b}) = \mathbf{0}$$

$$\boldsymbol{\alpha}^* = (\mathbf{Y}^T\mathbf{Y})^{-1}\boxed{\mathbf{Y}^T}\mathbf{b} = \mathbf{Y}^+\mathbf{b}$$

Q:伪逆法有什么不足吗？

线性分类器–LMSE算法

关于伪逆法的思考：

- 权向量的解依赖于$\boldsymbol{b}$，不同的$\boldsymbol{b}$会得到不同的权向量；
- 对线性可分的情况，不恰当的$\boldsymbol{b}$的选择，会导致伪逆解不一定位于正确的解区域内；
- 矩阵$\mathbf{Y}^T\mathbf{Y}$的逆，不一定存在。

【注意】

当矩阵$\mathbf{Y}^T\mathbf{Y}$维度较大时，矩阵求逆比较耗时，一般会采用梯度下降算法获得最优解：

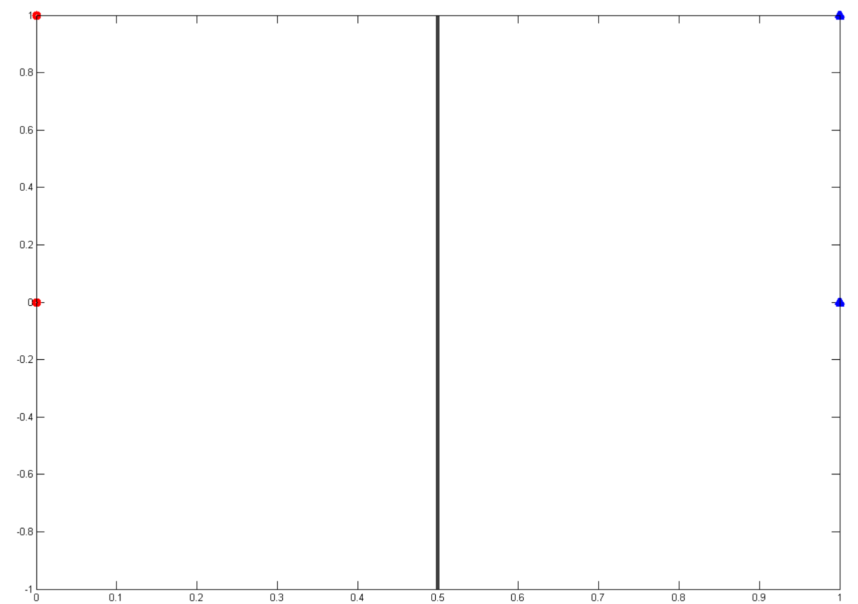
$$\begin{aligned}\boldsymbol{\alpha}(t+1) &= \boldsymbol{\alpha}(t) - \rho_t \nabla J_p(\boldsymbol{\alpha}) \\ &= \boldsymbol{\alpha}(t) - \rho_t \mathbf{Y}^T(\mathbf{Y}\boldsymbol{\alpha}(t) - \mathbf{b}(t))\end{aligned}$$

线性分类器－LMSE算法

例题：已知两类训练样本，(0,0), (0,1)属于w1, (1,0), (1,1)属于w2, 试用LMSE算法求$\mathbf{a}^*$

$$\mathbf{Y}=\begin{bmatrix}0 & 0 & 1\\ 0 & 1 & 1\\ -1 & 0 & -1\\ -1 & -1 & -1\end{bmatrix},\mathbf{b}=\begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

$$\boldsymbol{\alpha}^*=\begin{bmatrix}-2\\0\\1\end{bmatrix}$$

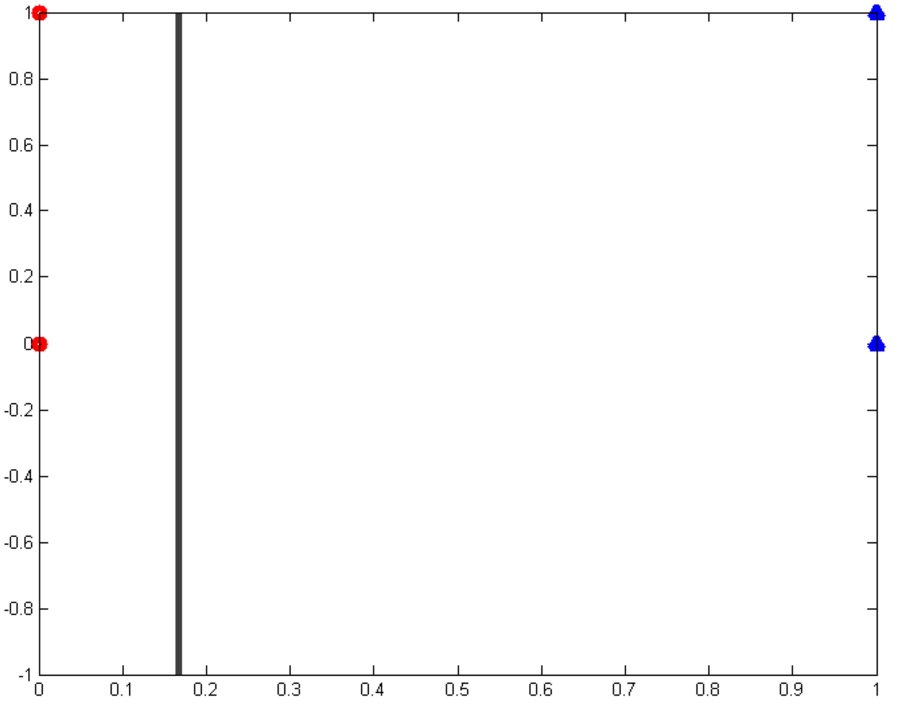


线性分类器－LMSE算法

例题：已知两类训练样本，(0,0), (0,1)属于w1, (1,0), (1,1)属于w2, 试用LMSE算法求$\mathbf{a}^*$

$$\mathbf{Y}=\begin{bmatrix}0 & 0 & 1\\ 0 & 1 & 1\\ -1 & 0 & -1\\ -1 & -1 & -1\end{bmatrix},\mathbf{b}=\begin{bmatrix}1\\1\\5\\5\end{bmatrix}$$

$$\boldsymbol{\alpha}^*=\begin{bmatrix}-6\\0\\1\end{bmatrix}$$

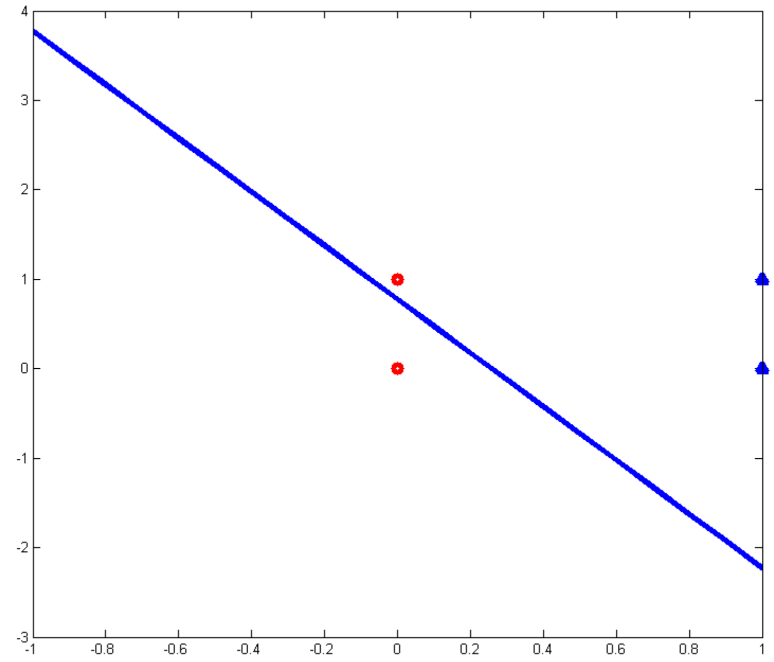


线性分类器 –LMSE算法

例题：已知两类训练样本，(0,0), (0,1)属于w1, (1,0), (1,1)属于w2, 试用LMSE算法求**a***

$$\mathbf{Y}=\begin{bmatrix}0 & 0 & 1\\ 0 & 1 & 1\\ -1 & 0 & -1\\ -1 & -1 & -1\end{bmatrix}, \mathbf{b}=\begin{bmatrix}0.2\\ 0.1\\ 1\\ 2\end{bmatrix}$$

$$\boldsymbol{\alpha}^{*}=\begin{bmatrix}-1.65\\ -0.55\\ 0.425\end{bmatrix}$$



线性分类器 –HK算法

Q: 松弛变量***b***如何给定呢？能不能自动学到理想的松弛变量呢？

优化权向量的同时，对松弛变量也同步进行优化

HK算法

(1965年，何毓琦和R.L.Kashyap共同提出)

线性分类器–HK算法

$$
\begin{aligned}
\boldsymbol{\alpha}^* &= \arg\min_{\boldsymbol{\alpha}} J_s(\boldsymbol{\alpha}) \\
&= \arg\min_{\boldsymbol{\alpha}} \left\| \mathbf{Y}\boldsymbol{\alpha} - \mathbf{b} \right\|^2
\end{aligned}
$$

梯度下降算法更新松弛变量：

$$
\begin{aligned}
\mathbf{b}(t+1) &= \mathbf{b}(t) - \eta_t \frac{\partial J_p(\boldsymbol{\alpha})}{\partial \mathbf{b}} \\
&= \mathbf{b}(t) + 2\eta_t(\mathbf{Y}\boldsymbol{\alpha}(t) - \mathbf{b}(t))
\end{aligned}
$$

因为要始终保证 $b_i > 0$，则松弛变量的更新公式表示为：

$$
\mathbf{b}(t+1) = \mathbf{b}(t) + \eta_t(\mathbf{e}(\mathrm{t}) + |\,\mathbf{e}(\mathrm{t})\,|)
$$

其中 $\mathbf{e}(t) = \mathbf{Y}\boldsymbol{\alpha}(t) - \mathbf{b}(t)$

线性分类器 –HK算法

HK算法流程：

(1) 当t=0时，设定初始松弛向量$\boldsymbol{b}(0)$，并求取初始向量

$$\boldsymbol{\alpha}(0) = (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{b}(0)$$

(2) 计算误差向量：$\mathbf{e}(t) = \mathbf{Y}\boldsymbol{\alpha}(t) - \mathbf{b}(t)$

(3) 计算 $\boldsymbol{\delta}(t) = \eta_t(\mathbf{e}(t) + |\mathbf{e}(t)|)$

(4) 更新权向量：$\boldsymbol{\alpha}(t+1) = (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{b}(t+1)$

$$= (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T(\mathbf{b}(t) + \boldsymbol{\delta}(t))$$

算法的收敛性在线性可分和$0<\eta_t<1$条件下已得到证明。

$$= \boldsymbol{\alpha}(t) + (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\boldsymbol{\delta}(t)$$

(5) 更新松弛变量：

$$\mathbf{b}(t+1) = \mathbf{b}(t) + \boldsymbol{\delta}(t)$$

(6) 重复(2)-(5)，直至所有的 $e=0$ (极小值)，中止迭代。

线性分类器 -HK算法

步骤(2)进一步讨论：

$$\mathbf{e}(t)=\begin{cases}\text{各分量取值}\ge 0 \\ \text{各分量取值有正有负} \\ \text{各分量取值}\le 0\text{，但不都是}0\end{cases}$$

各分量取值有正有负 → **情况不定**

各分量取值≤0，但不都是0 → **线性不可分**

参数不再更新

$$\boldsymbol{\delta}(t)=\eta_t(\mathbf{e}(t)+|\mathbf{e}(t)|)$$

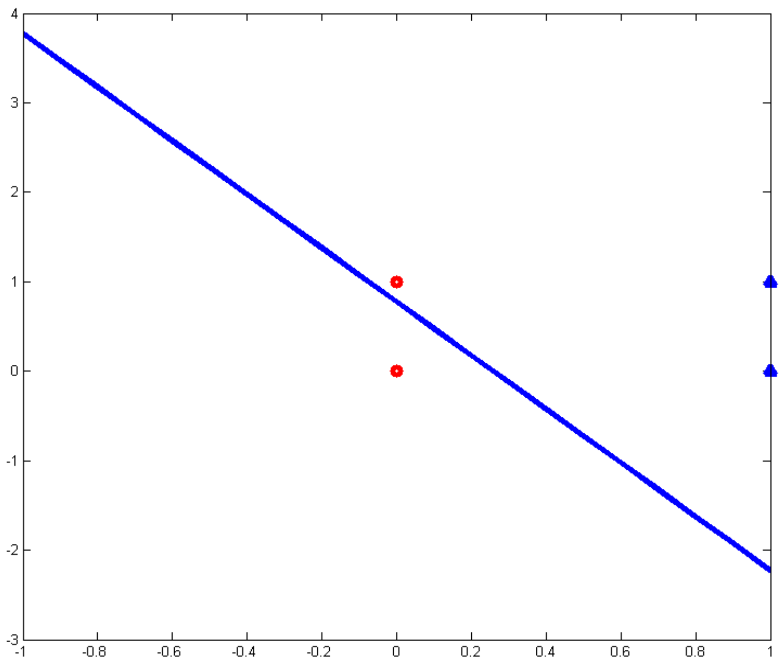
$$\mathbf{b}(t+1)=\mathbf{b}(t)+\boldsymbol{\delta}(t)$$

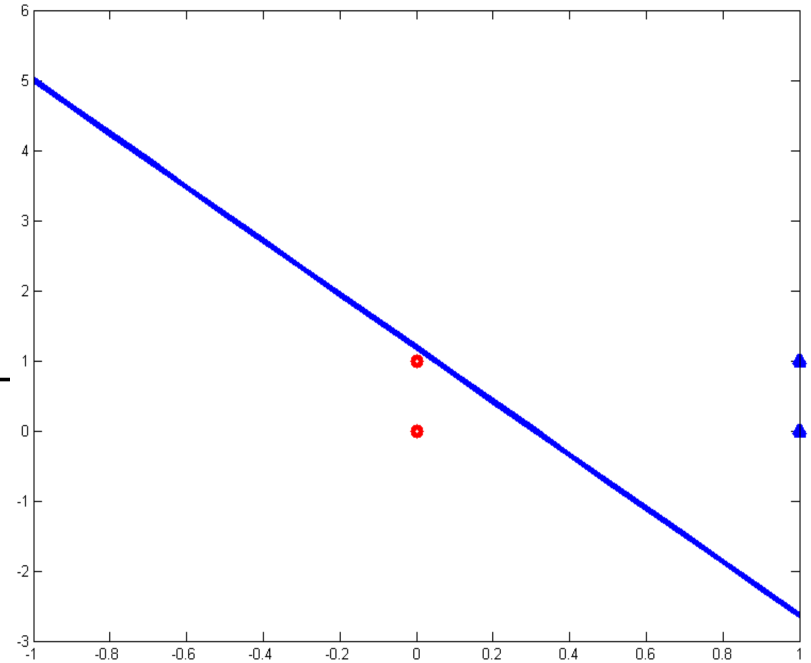
$$\boldsymbol{\alpha}(t+1)=\boldsymbol{\alpha}(t)+(\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\boldsymbol{\delta}(t)$$

【作业】证明***e***各分量小于等于0时，样本集线性不可分。

线性分类器 –HK算法

例题：已知两类训练样本，(0,0), (0,1)属于w1, (1,0), (1,1)属于w2, 试用HK算法求 $\mathbf{a}^*$





0.65

1.45

2

线性分类器 -HK算法

Matlab code:

```
clear all;
clc;
close all;
%%
Y=[0,0,1;0 1 1; -1,0,-1;-1,-1,-1];
b=[0.2;0.1;1;2];
Y_=inv(Y'*Y)*Y';
a=Y_*b;
e = Y*a-b;
r = 0.9;
delta = r*(e+abs(e));
flag = 1;
N = length(b);
C = 0;
```

```
while(flag)
  zenum = 0; nenum = 0;ponum = 0;
  for i = 1:N
    if (abs(e(i))<1e-16)
      zenum = zenum + 1;
    else if (e(i)<0)
        nenum = nenum + 1;
      else
        ponum = ponum + 1;
      end
    end
  end
  if((zenum-N)==0)%%all is 0
    flag=0;
  else if((zenum+ponum-N)==0)%%all is 0 or larger than 0
      flag=0;
    else if((zenum+nenum-N)==0)%%all is 0 or less than 0
        flag=0;
        ['Non-Linear!']
        break;
      else
        C = C + 1
        a = a + Y_*delta;
        b = b + delta;
        e = Y*a-b;
        delta = r*(e+abs(e));
      end
    end
  end
end
```

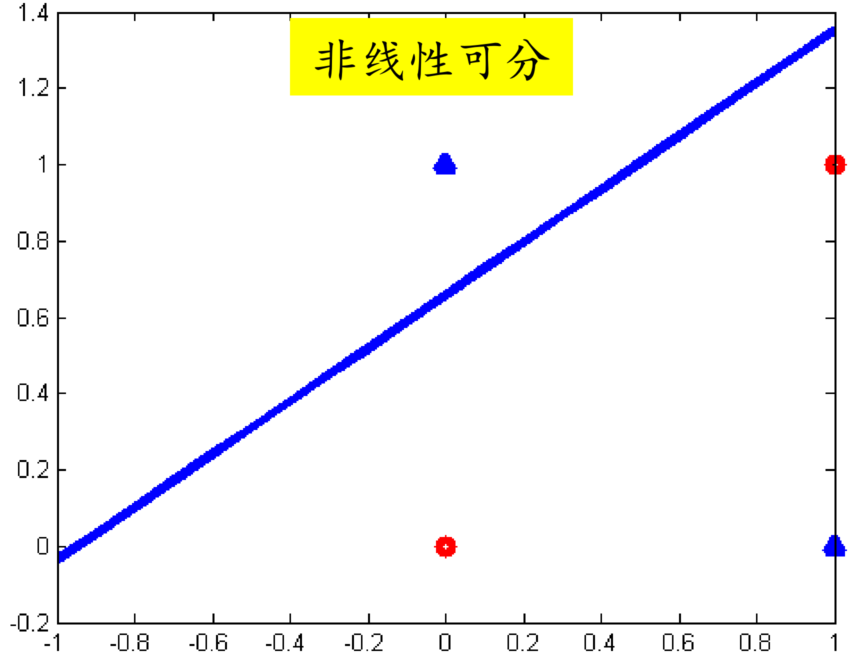
```
figure(1);
x1=(-1:0.1:1);
if abs(a(2))<0.0000001
    x1 = -a(3)/a(1);
    x2 = (-1:0.1:1);
    x1 = ones(length(x2))*x1;
else
    x2=(a(1)*x1+a(3))/(-a(2));
end
plot(x1,x2,'LineWidth',4);
hold on;plot(0,0,'ro','LineWidth',4);
hold on;plot(0,1,'ro','LineWidth',4);
hold on;plot(1,0,'b^','LineWidth',4);
hold on;plot(1,1,'b^','LineWidth',4);
```

线性分类器－HK算法

例题：已知两类训练样本，(0,0), (1,1)属于w1, (1,0), (0,1)属于w2, 试用HK算法求 $\mathbf{a}^*$

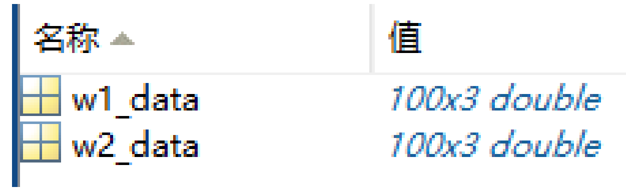
$$\mathbf{Y}=\begin{bmatrix}0 & 0 & 1\\ 0 & -1 & -1\\ -1 & 0 & -1\\ 1 & 1 & 1\end{bmatrix}, \mathbf{b}(0) = {}^{Rand}\begin{bmatrix}0.1493\\ 0.2575\end{bmatrix}$$

$$\mathbf{e}=\begin{bmatrix}-0.3755\\ -0.3755\\ -0.3755\\ -0.3755\end{bmatrix}$$



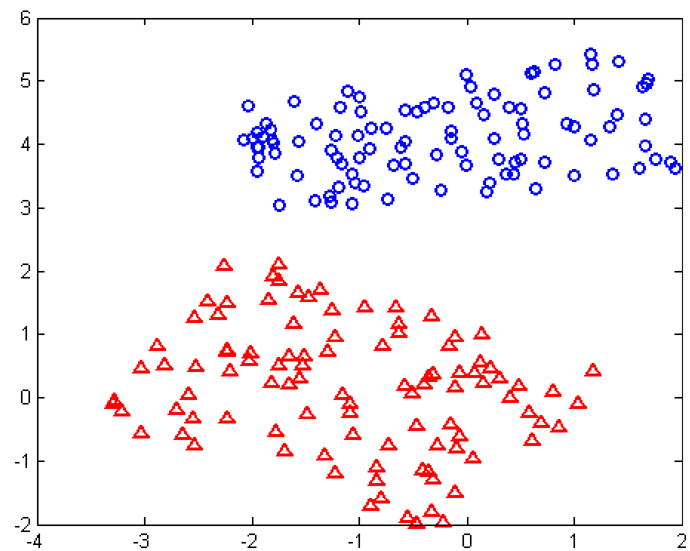
线性分类器 –HK算法

【作业】用HK算法求解权向量，并给出决策面方程。(Matlab编程实现，数据找学习委员copy)



要求：

1. 编程实现HK算法
2. 能实现线性不可分样本的判别
3. (加分项)初始化不同的***b***和学习率，研究分类结果




```
list = randperm(100);
W1 = [w2_data(list(1:10),:,:); w1_data(list(11:end),:,:)];
W2 = [w1_data(list(1:10),:,:); w2_data(list(11:end),:,:)];
```

线性分类器–HK算法

1.编程语言：Matlab/Python

2. 提交实验报告和源代码(**命名规则：HK_学号_姓名，先发给学习委员，学习委员再发到邮箱：** zhangjunchao_work@163.com)

3. 作业迟交n天，本次作业分数乘以0.98^n。

线性分类器－HK算法

若干素材取自网络，特此致谢！

谢谢聆听！