厦门大学信息学院 本科选修课
2021-2022 第二学期

模 式 识 别
Pattern Recognition

主 讲：王 程

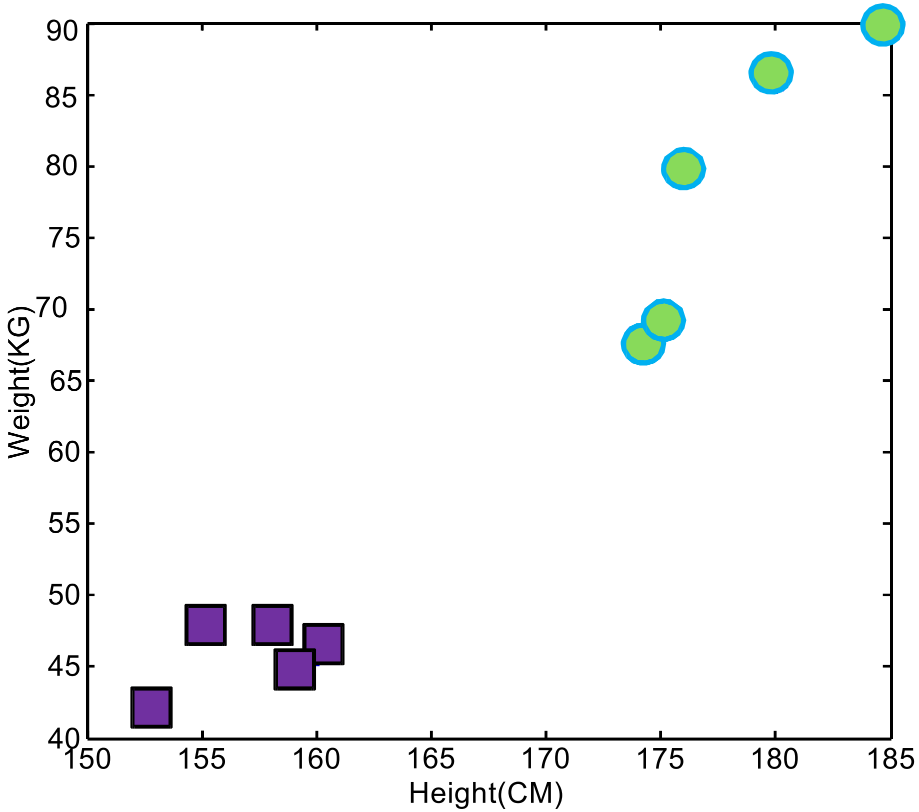
3 判别函数

3.6 可训练的确定性分类器的迭代算法

3.7 Fisher线性判别

3.5 感知器算法

已知一个班a级所有同学的身高和体重数据，学号为{0,2,3,4,7}的同学为男生，学号为{1,5,6,8,9}的同学为女生，如何分开两类数据



学号	0	1	2	3	4	5	6	7	8	9
身高（cm）	174	155	175	180	185	153	158	176	160	159
体重（kg）	68	48	69	86	90	42	48	80	46	45

3.6 可训练的确定性分类器迭代算法

对于每一样本引入类别标签$t_n = \{+1, -1\}, N = 1,2, \dots, N$

如果 $x_n \in \omega_1$，**则**$t_n = +1;$

如果 $x_n \in \omega_2$，**则** $t_n = -1$

学号	0	1	2	3	4	5	6	7	8	9
身高（cm）	174	155	175	180	185	153	158	176	160	159
体重（kg）	68	48	69	86	90	42	48	80	46	45
标签	+1	-1	+1	+1	+1	-1	-1	+1	-1	-1

3.6 可训练的确定性分类器迭代算法

感知器算法和梯度法的关系：

$t_n \boldsymbol{w}^\mathsf{T} \boldsymbol{x}_n > 0$ **样本分类正确**

$t_n \boldsymbol{w}^\mathsf{T} \boldsymbol{x}_n \leq 0$ **样本分类错误**

$$\boldsymbol{w} = (0.0875, 1.0345, -74.9995)^\mathsf{T}$$

$$g(\boldsymbol{x}_n) = \boldsymbol{w}^\mathsf{T} \boldsymbol{x}_n \gtrless 0 \qquad \boldsymbol{x}_n \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

学号	0	1	2	3	4	5	6	7	8	9
身高（cm）	174	155	175	180	185	153	158	176	160	159
体重（kg）	68	48	69	86	90	42	48	80	46	45
标签	+1	-1	+1	+1	+1	-1	-1	+1	-1	-1

3.6 可训练的确定性分类器迭代算法

感知器算法是梯度法的特殊情况，设准则函数为：

$$J(\boldsymbol{w}) = \sum_{n=1}^{N} J_n(\boldsymbol{w}) = \sum_{n=1}^{N} \frac{|t_n \boldsymbol{w}^T \boldsymbol{x}_n| - t_n \boldsymbol{w}^T \boldsymbol{x}_n}{2}$$

如何求解目标函数 $J(\boldsymbol{w})$取得极小值时候的权向量 $\boldsymbol{w}$？

$t_n \boldsymbol{w}^T \boldsymbol{x}_n > 0$ **样本分类正确**

$t_n \boldsymbol{w}^T \boldsymbol{x}_n \leq 0$ **样本分类错误**

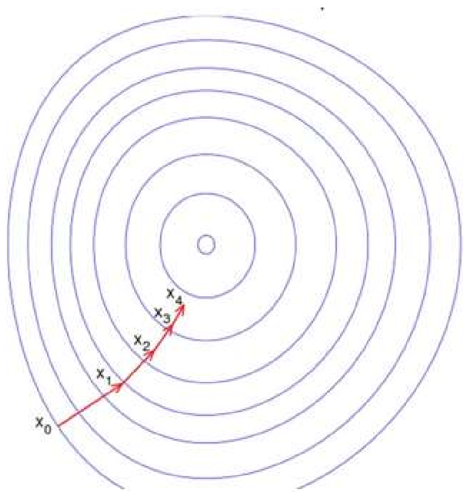
3.6 可训练的确定性分类器迭代算法

梯度法：

$$\nabla J(\boldsymbol{w}) = \begin{bmatrix} \dfrac{\partial J}{\partial w_1} \\ \vdots \\ \dfrac{\partial J}{\partial w_n} \end{bmatrix}$$

$\nabla J(\boldsymbol{w})$定义为准则函数 $J(\boldsymbol{w})$的梯度方向，它是一个向量。

梯度方向是函数在其自变量增加时最增长率的方向，反之，负梯度方向是函数J的最陡下降方向。



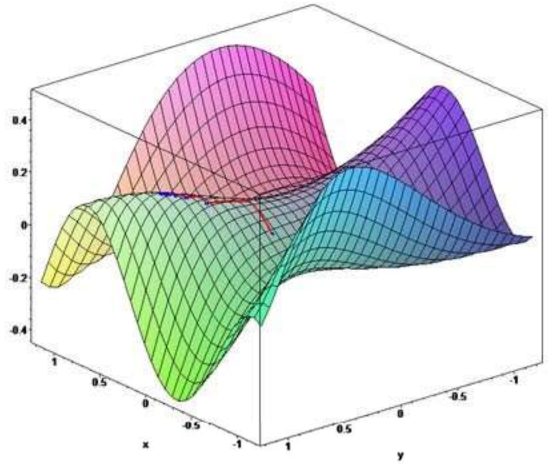
3.6 可训练的确定性分类器迭代算法

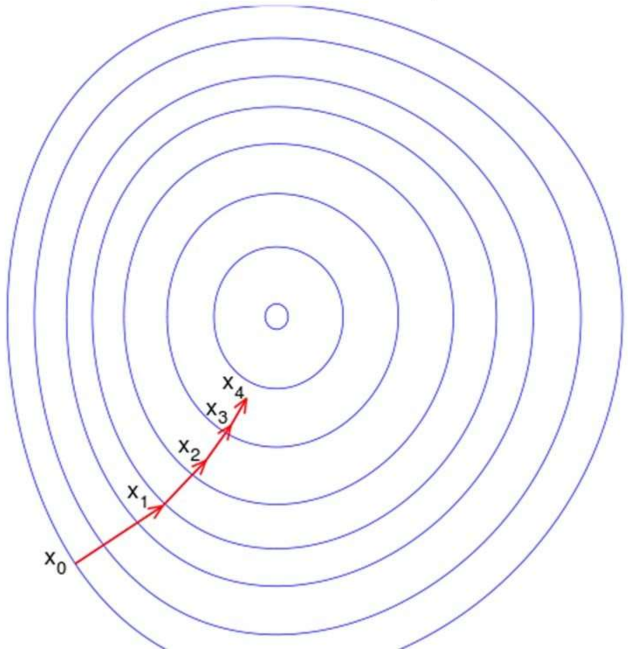
构造关于权向量的准则函数$J(w)$ **,我们通常需要寻找使得**$J(w)$**取得极值时对应的权向量最优解w*。**

每一步迭代中，计算：

$$w^{(k+1)} = w^{(k)} - C\nabla J(w)|_{w=w^{(k)}}$$

其中，C是一个正的比例因子， 作为步长。





3.6 可训练的确定性分类器迭代算法

由$J_n(\boldsymbol{w})=\frac{|t_n\boldsymbol{w}^T x_n|-t_n\boldsymbol{w}^T x_n}{2}$**可得：**

$$J_n(\boldsymbol{w})=\begin{cases}0 & t_n w^T x_n\geq 0\\ -t_n\boldsymbol{w}^T x_n & t_n w^T x_n\leq 0\end{cases}$$

对$J_n(\boldsymbol{w})$关于w**求偏导数可以得到：**

$$\nabla J_n(\boldsymbol{w})=\frac{\partial J_n(\boldsymbol{w})}{\partial w}=\begin{cases}0 & t_n\boldsymbol{w}^T\boldsymbol{x_n}\geq 0\\ -t_n x_n & t_n\boldsymbol{w}^T\boldsymbol{x_n}\leq 0\end{cases}$$

3.6 可训练的确定性分类器迭代算法

$\boldsymbol{w}^T\boldsymbol{x}_{\boldsymbol{n}} = w_1x_{n1} + w_1x_{n2} + \cdots + w_1x_{nD}$

其中 $\boldsymbol{w} = (w_1, w_1, \ldots, w_1)^T$, $\boldsymbol{x}_{\boldsymbol{n}} = (x_{n1}, x_{n2}, \ldots, x_{nD})^T$

求 $w^T x_n$ **关于** w **的偏导数**

$$\frac{\partial \boldsymbol{w}^T\boldsymbol{x}_{\boldsymbol{n}}}{\partial w} = \begin{pmatrix} \frac{\partial \boldsymbol{w}^T\boldsymbol{x}_{\boldsymbol{n}}}{\partial w_1} \\ \vdots \\ \frac{\partial \boldsymbol{w}^T\boldsymbol{x}_{\boldsymbol{n}}}{\partial w_D} \end{pmatrix} = \begin{pmatrix} x_{n1} \\ \vdots \\ x_{nD} \end{pmatrix} = \boldsymbol{x}_{\boldsymbol{n}}$$

3.6 可训练的确定性分类器迭代算法

根据随机梯度法的迭代规则：

$$\boldsymbol{w}^{(k+1)} = \boldsymbol{w}^{(k)} - C\frac{\partial J_n(\boldsymbol{w})}{\partial w}\bigg|_{\boldsymbol{w}=\boldsymbol{w}^{(k)}}$$

$$= \begin{cases} \boldsymbol{w}^{(k)} & t_n \boldsymbol{w}^{(k)^T} x_n \geq 0 \\ \boldsymbol{w}^{(k)} + C t_n \boldsymbol{x_n} & t_n \boldsymbol{w}^{(k)^T} x_n \leq 0 \end{cases}$$

式中，x_n为训练的模式样本，　k为第 k次迭代。

感知器算法是梯度法的特例！

$$\frac{\partial J_n(\boldsymbol{w})}{\partial w} = \begin{cases} 0 & t_n \boldsymbol{w}^{\boldsymbol{T}} \boldsymbol{x_n} \geq 0 \\ -t_n x_n & t_n \boldsymbol{w}^{\boldsymbol{T}} \boldsymbol{x_n} \leq 0 \end{cases}$$

3.6 可训练的确定性分类器迭代算法

感知器方法的优点：

简单容易实现。

缺点：

要求全部样本线性可分，当样本不是线性可分时， 如果仍然使用感知器算法，则不会收敛。

3.6 可训练的确定性分类器迭代算法

最小平方误差（LMSE）：

对于两类分类问题相当于求一组线性不等式的解：

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_{\mathrm{n}} > 0, \forall \boldsymbol{x}_{\mathrm{n}} \in \omega_1$$
$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_{\mathrm{n}} < 0, \forall \boldsymbol{x}_{\mathrm{n}} \in \omega_2$$

若引入样本类别标签$t_n = \{+1, -1\}, N = 1, 2, \dots, N$**可得对于全部模式都需要满足不等式**$t_n w^T x_n > 0$,

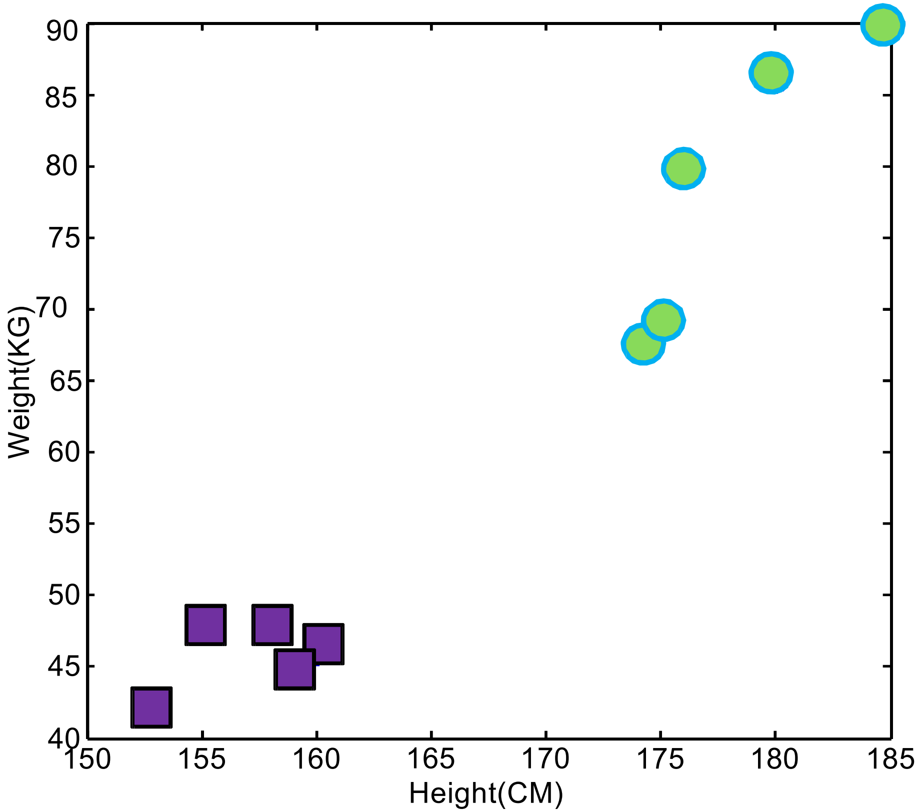
以上用不等式表示为：

$$\boldsymbol{X}\boldsymbol{w} > \mathbf{0}$$

其中X= $\begin{bmatrix} t_1 \boldsymbol{x}_1^{\mathrm{T}} \\ \vdots \\ t_{\mathrm{N}} \boldsymbol{x}_{\mathrm{N}}^{\mathrm{T}} \end{bmatrix}$

3.5 感知器算法

已知一个班a级所有同学的身高和体重数据，学号为{0,2,3,4,7}的同学为男生，学号为{1,5,6,8,9}的同学为女生，如何分开两类数据



学号	0	1	2	3	4	5	6	7	8	9
身高（cm）	174	155	175	180	185	153	158	176	160	159
体重（kg）	68	48	69	86	90	42	48	80	46	45

3.6 可训练的确定性分类器迭代算法

$t_n \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_n > 0$ **样本分类正确**

$t_n \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_n \leq 0$ **样本分类错误**

$$\boldsymbol{w} = (0.0875, 1.0345, -74.9995)^{\mathrm{T}}$$

$$g(\boldsymbol{x}_n) = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_n \underset{<}{\overset{>}{}} 0 \qquad \boldsymbol{x}_n \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

学号	0	1	2	3	4	5	6	7	8	9
身高（cm）	174	155	175	180	185	153	158	176	160	159
体重（kg）	68	48	69	86	90	42	48	80	46	45
标签	+1	-1	+1	+1	+1	-1	-1	+1	-1	-1

$$\mathbf{X} = \begin{bmatrix} t_0 \boldsymbol{x}_0^{\mathrm{T}} \\ \vdots \\ t_9 \boldsymbol{x}_9^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} 174, 68 \\ \vdots \\ -159, -45 \end{bmatrix}$$

3.6 可训练的确定性分类器迭代算法

根据以上分析，如令 分 $\boldsymbol{b} = (b_1, \ldots, b_N)^T$**，且b的所有量均为正值，则** $\boldsymbol{Xw} > \boldsymbol{0}$ **可写为：**

$$\boldsymbol{Xw} = \boldsymbol{b}$$

这里需要同时求解W和b。一般情况下，Xw $= \boldsymbol{b}$ **属 于超定方程，只能求得最小二乘解。**

3.6 可训练的确定性分类器迭代算法

Hoshen–Kopelman algorithm

如上分析，H-K算法定义准则函数

$$J(\boldsymbol{w},\boldsymbol{b})=\frac{1}{2}\|\boldsymbol{X}\boldsymbol{w}-\boldsymbol{b}\|^{2}$$

$$=\frac{1}{2}(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{b})$$

$$=\frac{1}{2}\sum_{n=1}^{N}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_{\mathrm{n}}-\boldsymbol{b})^{2}$$

上式为两个数量的方差之和，故名为最小均方误差 算法。

Hoshen, J.; Kopelman, R. (15 October 1976). "Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm". *Phys. Rev. B*. **14** (8): 3438–3445. doi:10.1103/PhysRevB.14.3438

3.6 可训练的确定性分类器迭代算法

$$J(\boldsymbol{w},\boldsymbol{b}) = \frac{1}{2}(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{b})^\top(\boldsymbol{X}\boldsymbol{w}-\boldsymbol{b})$$

$$= \frac{1}{2}(\boldsymbol{w}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{w}-\boldsymbol{w}^\top\boldsymbol{X}^\top\boldsymbol{b}-\boldsymbol{b}^\top\boldsymbol{X}\boldsymbol{w}+\boldsymbol{b}^\top\boldsymbol{b})$$

$$= \frac{1}{2}\left(\boldsymbol{w}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{w}-2\boldsymbol{w}^\top\boldsymbol{X}^\top\boldsymbol{b}+\boldsymbol{b}^\top\boldsymbol{b}\right)$$

$$\frac{\partial J}{\partial \boldsymbol{w}} = \boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{w}-\boldsymbol{X}^\top\boldsymbol{b}$$

3.6 可训练的确定性分类器迭代算法

$$\boldsymbol{w}^T\boldsymbol{A}\boldsymbol{w} = \sum_{i=1}^{D}\sum_{j=1}^{D} a_{ij}w_iw_j$$

其中$w = (w_1, w_2, \dots, w_D)^T, A = \begin{pmatrix} a_{11} & \cdots & a_{1D} \\ \vdots & \ddots & \vdots \\ a_{D1} & \cdots & a_{DD} \end{pmatrix}$

求$\boldsymbol{w}^T\boldsymbol{A}\boldsymbol{w}$关于$\boldsymbol{w}$的偏导数

$$\frac{\partial \boldsymbol{w}^T\boldsymbol{A}\boldsymbol{w}}{\partial \boldsymbol{w}} = \begin{pmatrix} \frac{\partial \boldsymbol{w}^T\boldsymbol{A}\boldsymbol{w}}{\partial w_1} \\ \vdots \\ \frac{\partial \boldsymbol{w}^T\boldsymbol{A}\boldsymbol{w}}{\partial w_D} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{D}(a_{1j} + a_{j1})w_j \\ \vdots \\ \sum_{j=1}^{D}(a_{Dj} + a_{jD})w_j \end{pmatrix} = (\boldsymbol{A} + \boldsymbol{A}^T)\boldsymbol{w}$$

3.6 可训练的确定性分类器迭代算法

使函数J同时对 w 和b求最小，对于 w 的梯度为：

$$\frac{\partial J}{\partial w}=X^{\mathrm{T}}(Xw-b)$$

令$\frac{\partial J}{\partial w}=0$ **，得**$X^{\mathrm{T}}(Xw-b)=0$

$$w=(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}b=X^{\#}b$$

式中 $X^{\#}=\left(X^{\mathrm{T}}X\right)^{-1}X^{T}$**称为X的伪逆。**

3.6 可训练的确定性分类器迭代算法

利用梯度法求解b的迭代式为：

$$\boldsymbol{b}^{(k+1)} = \boldsymbol{b}^{(k)} - C\frac{\partial J}{\partial \boldsymbol{b}}\Big|_{\boldsymbol{b}=\boldsymbol{b}^{(k)}}$$

为了保证每次迭代中的为$\boldsymbol{b}^{(k)}$正值，应使$\frac{\partial J}{\partial \boldsymbol{b}}\Big|_{\boldsymbol{b}=\boldsymbol{b}^{(k)}}$为负值。

因为：

$$\frac{\partial J}{\partial \boldsymbol{b}} = -(\boldsymbol{Xw} - \boldsymbol{b})$$

为保证为$\boldsymbol{b}^{(k+1)}$正数，故令：

$$\frac{\partial J}{\partial \boldsymbol{b}} = -\frac{1}{2}[(\boldsymbol{Xw} - \boldsymbol{b}) + |\boldsymbol{Xw} - \boldsymbol{b}|]$$

3.6 可训练的确定性分类器迭代算法

H-K算法流程：

已知$\boldsymbol{b}^{(0)}$

$\boldsymbol{w}^{(0)} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{b}^{(0)} = \boldsymbol{X}^{\#}\boldsymbol{b}^{(0)}$;

…

$\boldsymbol{b}^{(k)} = \boldsymbol{b}^{(k-1)} - C\frac{\partial J}{\partial \boldsymbol{b}}\Big|_{\boldsymbol{b}=\boldsymbol{b}^{(k-1)}}$

$\boldsymbol{w}^{(k)} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{b}^{(k)} = \boldsymbol{X}^{\#}\boldsymbol{b}^{(k)}$;

当$\boldsymbol{X}\boldsymbol{w}^{(k)} - \boldsymbol{b}^{(k)} = 0$**,停止迭代，有解**

当$\boldsymbol{X}\boldsymbol{w}^{(k)} - \boldsymbol{b}^{(k)} > 0$**,继续迭代，有解**

当$\boldsymbol{X}\boldsymbol{w}^{(k)} - \boldsymbol{b}^{(k)}$**全部分量停止变为正值，表明该模式类别线性不可分。**

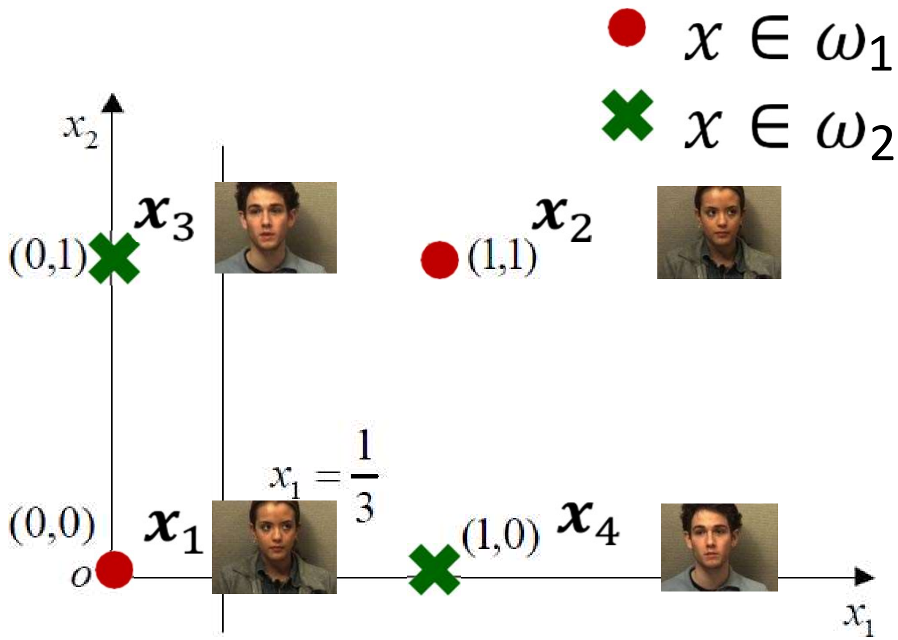
3.6 可训练的确定性分类器迭代算法

例：已知模式样本集：

ω_1: $\{(0,0)^T, \quad (1,1)^T\}$

ω_2: $\{(0,1)^T, \quad (1,0)^T\}$

它是线性不可分的。



3.6 可训练的确定性分类器迭代算法

取 $C=1$**和** $\boldsymbol{b}^{(0)}=(1,1,1,1)^{\mathrm{T}}$，**则**

$$\boldsymbol{X}=\begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & -1 & -1 \\ -1 & 0 & -1 \end{pmatrix}$$

伪逆矩阵为：

$$\boldsymbol{X}^{\#}=\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}=\begin{pmatrix} -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ \frac{3}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

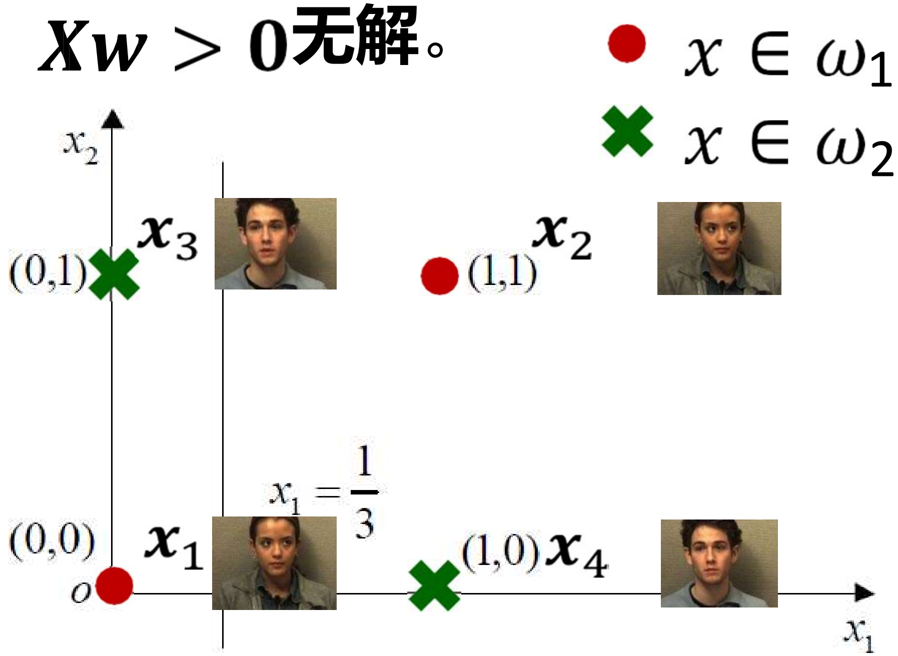
3.6 可训练的确定性分类器迭代算法

$$\boldsymbol{w}^{(0)} = \boldsymbol{X}^{\#}\boldsymbol{b}^{(0)} = (0,0,0)^T$$

$$\boldsymbol{X}\boldsymbol{w}^{(0)} - \boldsymbol{b}^{(0)} = (-1,-1,-1,-1)^T$$

向量$\boldsymbol{X}\boldsymbol{w}^{(0)} - \boldsymbol{b}^{(0)}$的全部分量都为负，

这表示不等式组 $\boldsymbol{X}\boldsymbol{w} > \boldsymbol{0}$**无解。**



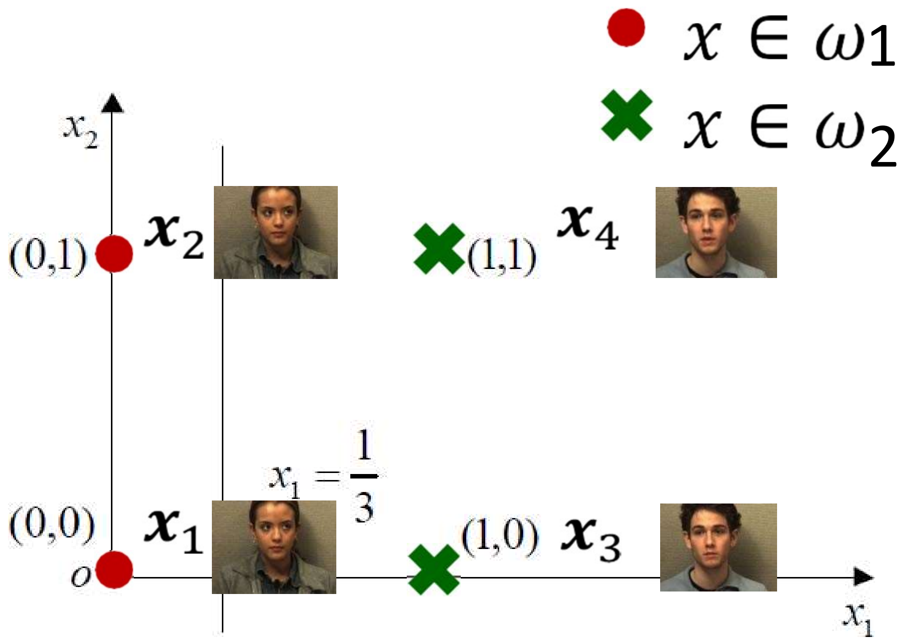
3.6 可训练的确定性分类器迭代算法

例：已知两类样本：

ω_1：$\{(0,0)^{\mathrm{T}}, (0,1)^{\mathrm{T}}\}$；

ω_2：$\{(1,0)^{\mathrm{T}}, (1,1)^{\mathrm{T}}\}$；

试计算能区分两类样本的判别界面方程。



3.6 可训练的确定性分类器迭代算法

写出模式的增广矩阵为：

$$\boldsymbol{X}=\begin{pmatrix}0 & 0 & 1\\ 1 & 1 & 1\\ -1 & 0 & -1\\ -1 & -1 & -1\end{pmatrix}$$

则伪逆为：

$$\boldsymbol{X}^{\#}=\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}=\frac{1}{2}\begin{pmatrix}-1 & -1 & -1 & -1\\ -1 & 1 & 1 & -1\\ \frac{3}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2}\end{pmatrix}$$

3.6 可训练的确定性分类器迭代算法

取C=1和$\boldsymbol{b}^{(0)}=(1,1,1,1)^T$，则

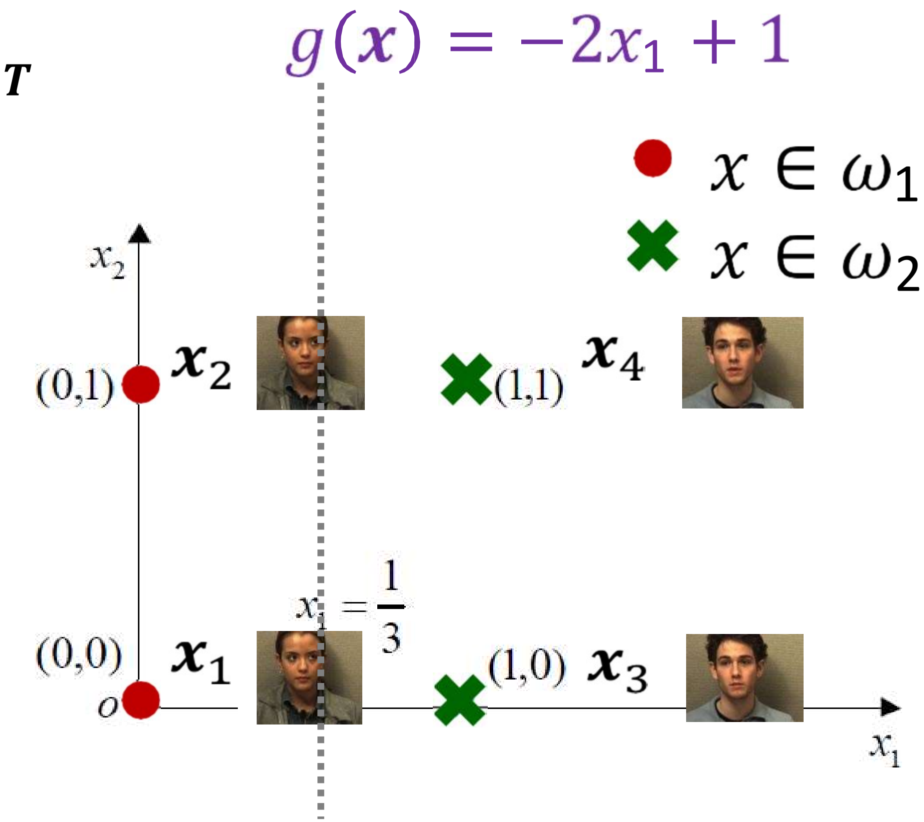
$$\boldsymbol{w}^{(0)}=X^{\#}\boldsymbol{b}^{(0)}=(-2,0,1)^T$$

因$X\boldsymbol{w}^{(0)}=(1,1,1,1)^T$

$$X\boldsymbol{w}^{(0)}-\boldsymbol{b}^{(0)}=(0,0,0,0)^T$$

故

$$\boldsymbol{w}=\boldsymbol{w}^{(0)}=(-2,0,1)^T$$



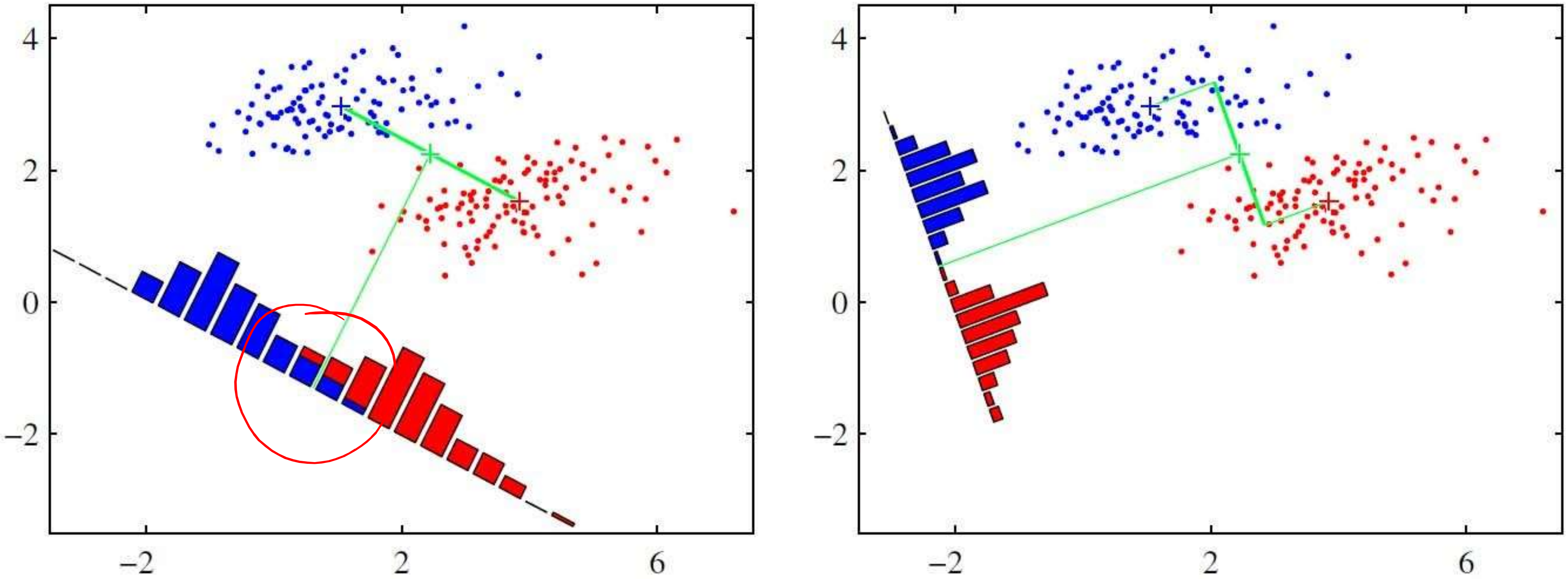
3 判别函数

3.6 可训练的确定性分类器的迭代算法

3.7 Fisher线性判别

3.7 Fisher线性判别

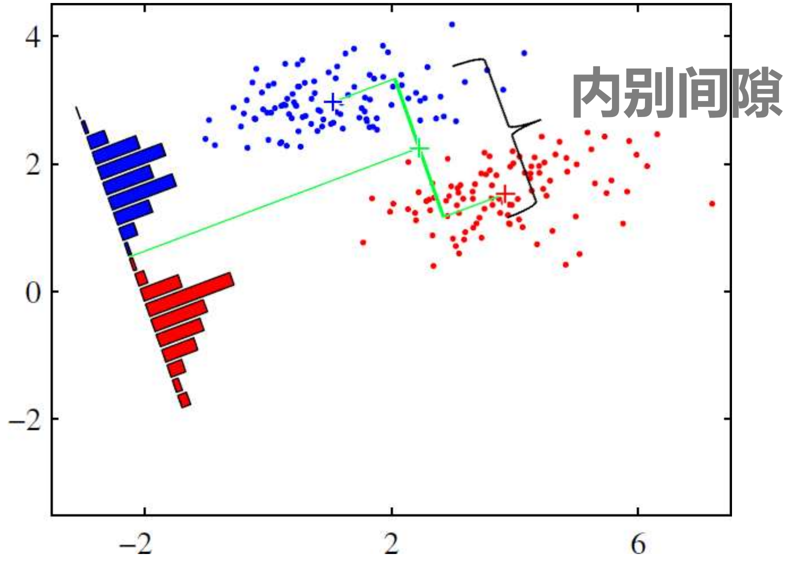
Fisher判别法是1936年提出来的，该方法的主要思 想是通过将多维数据投影到某个方向上，投影的原 则是将总体与总体之间尽可能的分开，然后再选择 合适的判别规则，将新的样本进行分类判别。



3.7 Fisher线性判别

假设有样本集$\{x_1, x_2 \ldots, x_N\}$以及它们所属类别ω_1和ω_2，我们希望寻找一个投影方向w，期望对于属于ω_1的样本x_n，其投影后的值$w^T x_n > w_0$,对于属于ω_2的样本x_n，其投影后的值$w^T x_n < w_0$。

Fisher线性判别准则同时期望两个类别样本在投影 后的类别间隙最大，类内样本尽量集中。



3.7 Fisher线性判别

类别中心：

$$\boldsymbol{\mu}_1 = \frac{1}{N} \sum_{{}^{1}x_n \in \omega_1} \boldsymbol{x}_\text{n}$$

$$\boldsymbol{\mu}_2 = \frac{1}{N} \sum_{{}^{2}x_n \in \omega_2} \boldsymbol{x}_\text{n}$$

变换后各个模式在投影方向w上的均值可以表示为：

$$\mu'_1 = \frac{1}{N_1} \sum_{x_n \in \omega_1} \boldsymbol{w}^\text{T} \boldsymbol{x}_\text{n} = \boldsymbol{w}^\text{T} \boldsymbol{\mu}_1$$

$$\mu'_2 = \frac{1}{N_1} \sum_{x_n \in \omega_2} \boldsymbol{w}^\text{T} \boldsymbol{x}_\text{n} = \boldsymbol{w}^\text{T} \boldsymbol{\mu}_2$$

3.7 Fisher线性判别

类内散度：

$$S'_W = \sum_{x_n \epsilon \omega_1} (w^T x_n - \mu_1)^2 + \sum_{x_n \epsilon \omega_2} (w^T x_n - w^T \mu_2)^2 = w^T S_W w$$

表示同类样本的集中程度。 类内散度矩阵：

$$S_W = \sum_{x_n \in \omega_1} (\boldsymbol{x}_n - \boldsymbol{\mu}_1)(\boldsymbol{x}_n - \boldsymbol{\mu}_1)^T + \sum_{x_n \in \omega_2} (\boldsymbol{x}_n - \boldsymbol{\mu}_2)(\boldsymbol{x}_n - \boldsymbol{\mu}_2)^T$$

3.7 Fisher线性判别

类间散度：

$$s'_{\mathrm{B}} = (\mu'_2 - \mu'_1)^2$$
$$= (\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\mu}_2 - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\mu}_1)^2 = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{B}}\mathbf{w}$$

表示不同类样本之间的分离程度。

类间散度矩阵：

$$\boldsymbol{S}_{\mathrm{B}} = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^{\mathrm{T}}$$

根据Fisher 准则的定义，期望类内散度 s'_{W}越小，类间散度s_{B}越大越好，根据以上准则定义目标函数：

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{B}}\mathbf{w}}{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{W}}\mathbf{w}}$$

3.7 Fisher线性判别

求解 $J(\boldsymbol{w})=\frac{w^T S_B \boldsymbol{w}}{w^T S_W \boldsymbol{w}}$**的最大值转化为优化问题：**

$$\max_{\mathrm{w}} \ \boldsymbol{w}^{\mathrm{T}} \boldsymbol{S}_{\mathrm{B}} \boldsymbol{w}$$
$$s.t. \ \boldsymbol{w}^{\mathrm{T}} \boldsymbol{S}_{\mathrm{W}} \boldsymbol{w}=c \neq 0$$

s.t. 表示优化问题中需要满足约束的条件 （subject to的缩写）

聂曼－皮尔逊判别准则

拉格朗日乘子法：

假设需要求极值的目标函数(objective function)为，

$f(x,y)$，**限制条件为**$\varphi(x,y)=M$

设$g(x,y)=M-\varphi(x,y)$

定义一个新函数

$$F(x,y,\lambda)=f(x,y)+\lambda g(x,y)$$

则用偏导数方法列出方程：

$$\partial F/\partial x=0$$

$$\partial F/\partial y=0$$

$$\partial F/\partial \lambda=0$$

求出x,y,λ**的值，代入即可得到目标函数的极值**

聂曼－皮尔逊判别准则

扩展为多个变量的式子为：

$$F(x_1, x_2, \ldots, \lambda) = f(x_1, x_2, \ldots) + \lambda g(x_1, x_2, \ldots)$$

则求极值点的方程为

$$\partial F / \partial x_i = 0$$

$$\partial F / \partial \lambda = 0$$

3.7 Fisher线性判别

通过引入拉格朗日乘子转化为以下拉格朗日无约束极值问题：

$$L(\boldsymbol{w},\lambda) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{B}}\boldsymbol{w} - \lambda(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{W}}\boldsymbol{w} - c)$$

令导数

$$\frac{\partial L(\boldsymbol{w},\lambda)}{\partial \boldsymbol{w}} = \boldsymbol{0}$$

可得：

$$\boldsymbol{S}_{\mathrm{B}}\boldsymbol{w} - \lambda\boldsymbol{S}_{\mathrm{W}}\boldsymbol{w} = \boldsymbol{0}$$

3.7 Fisher线性判别

可得：

$$\boldsymbol{S}_{\mathrm{B}}\boldsymbol{w} = \lambda \boldsymbol{S}_{\mathrm{W}}\boldsymbol{w}$$

假定 $\boldsymbol{S}_{\mathrm{W}}$ 为非奇异，则：

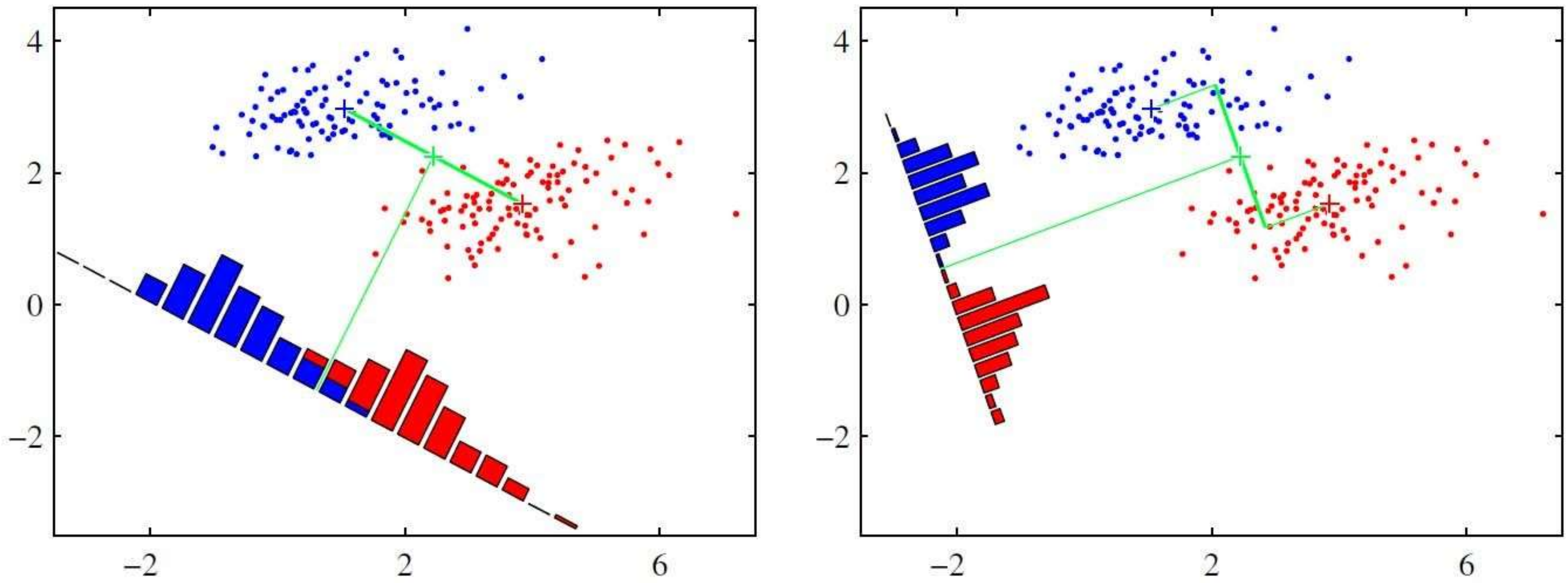
$$\boldsymbol{S}_{\mathrm{W}}^{-1}\boldsymbol{S}_{\mathrm{B}}\boldsymbol{w} = \lambda \boldsymbol{w}$$

即最佳投影方向为 向 $\boldsymbol{S}_{\mathrm{W}}^{-1}\boldsymbol{S}_{\mathrm{B}}$ 的最大特征值对应的特征量。

3.7 Fisher线性判别

取 $w_0 = \frac{1}{2}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\mu}_2 + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\mu}_1)$**，对于任意模式向量**$\boldsymbol{x}_{\mathrm{n}}$**，计算它在** w **上的投影，即可判断类别。**

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_{\mathrm{n}} \gtrless w_0，\quad \boldsymbol{x}_{\mathrm{n}} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$



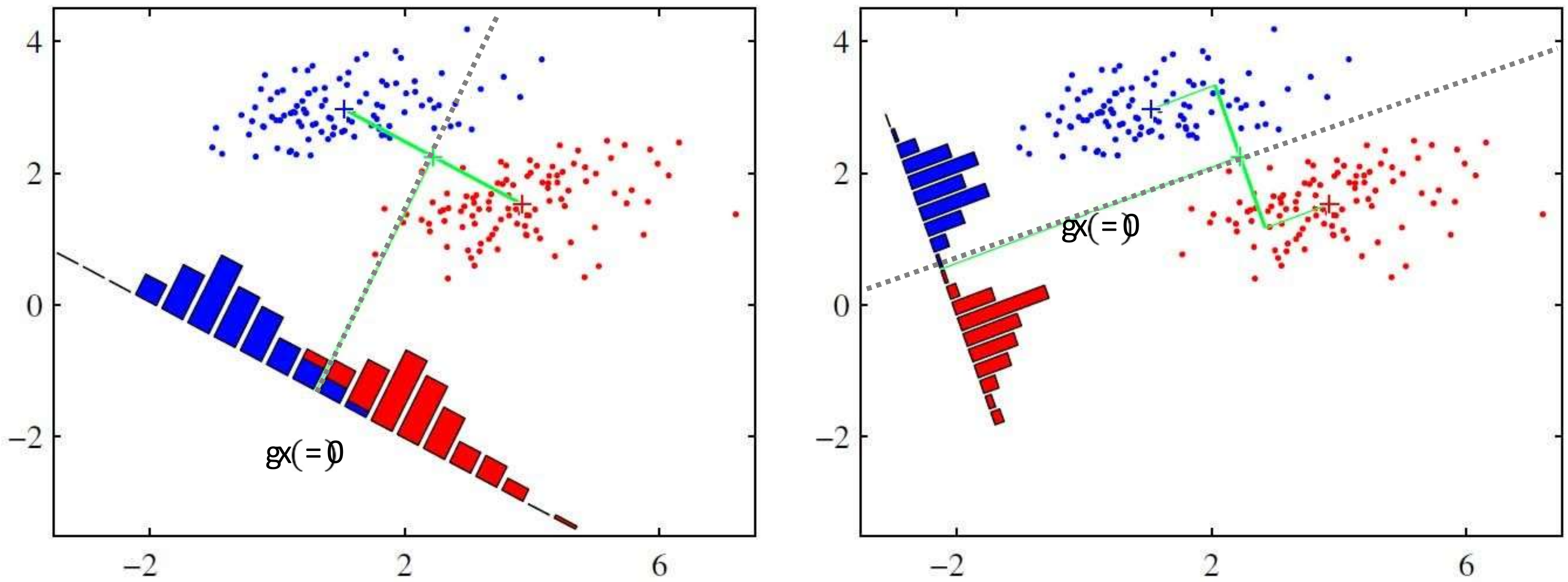
3.7 Fisher线性判别

判别规则为：

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_{\mathrm{n}} \underset{<}{>} w_0 \text{，} \boldsymbol{x}_{\mathrm{n}} \in \begin{cases}\omega_1 \\ \omega_2\end{cases}$$

等价于：

$$g(\boldsymbol{x}_{\mathrm{n}}) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_{\mathrm{n}} - w_0 \underset{<}{>} 0, \ \boldsymbol{x}_{\mathrm{n}} \in \begin{cases}\omega_1 \\ \omega_2\end{cases}$$



3.7 Fisher线性判别

有两类样本：

ω_1： $\{(1,0)^{\mathsf{T}},(-1,0)^{\mathsf{T}},\ (0,2)^{\mathsf{T}},(0,-2)^{\mathsf{T}}\}$和

ω_2： $\{(5,8)^{\mathsf{T}},(3,8)^{\mathsf{T}},(4,10)^{\mathsf{T}},\ (4,6)^{\mathsf{T}}\}$，

用Fisher准则进行分类，得到判别规则。

解：计算两个类别的中心μ_1和μ_2，

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\quad \boldsymbol{\mu}_2 = \begin{pmatrix} 4 \\ 8 \end{pmatrix}$$

3.7 Fisher线性判别

由类内散度矩阵的计算公式

$$S_W = \sum_{x_n \in \omega_1} (x_n - \mu_1)(x_n - \mu_1)^T + \sum_{x_n \in \omega_2} (x_n - \mu_2)(x_n - \mu_2)^T$$

得到

$$\boldsymbol{S}_{\mathrm{W}} = \begin{pmatrix} 4 & 0 \\ 0 & 16 \end{pmatrix}$$

3.7 Fisher线性判别

由类间散度矩阵的计算公式

$$\boldsymbol{S}_{\mathrm{B}} = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^{\mathrm{T}}$$

得到

$$\boldsymbol{S}_{\mathrm{B}} = \begin{pmatrix} 16 & 32 \\ 32 & 64 \end{pmatrix}$$

3.7 Fisher线性判别

根据Fisher准则，最佳投影方向 w 应该为 $S_W^{-1}S_B$ 的最大值对应的特征向量，

$$S_W^{-1}S_B = \begin{pmatrix} 1/4 & 0 \\ 0 & 1/16 \end{pmatrix}\begin{pmatrix} 16 & 32 \\ 32 & 64 \end{pmatrix} = \begin{pmatrix} 4 & 8 \\ 2 & 4 \end{pmatrix}$$

$S_W^{-1}S_B$ 对应的特征值为

$$\lambda_1 = 8, \ \lambda_2 = 0,$$

其中较大特征值 λ_1 对应的特征向量为 $\left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right)^T$，因此最佳投影方向 $w = \left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right)^T$

3.7 Fisher线性判别

对于任意样本x_n，判别规则为：

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_{\mathrm{n}} \gtrless w_0,\quad \boldsymbol{x}_{\mathrm{n}} \in \begin{cases}\omega_1\\ \omega_2\end{cases},$$

$$w_0 = \frac{1}{2}\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\mu}_2 + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\mu}_1\right) = \frac{1}{2}\left(\frac{2}{\sqrt{5}},\frac{1}{\sqrt{5}}\right)^{\mathrm{T}}\begin{pmatrix}4\\8\end{pmatrix} = \frac{8}{\sqrt{5}}\text{。}$$

3.7 Fisher线性判别

3.7 Fisher线性判别

3.7 Fisher线性判别

1.计算类内散度矩阵：

$$\boldsymbol{S_W} = \sum_{x_n \in \omega_1} (x_n - \mu_1)(x_n - \mu_1)^T + \sum_{x_n \in \omega_2} (x_n - \mu_2)(x_n - \mu_2)^T$$

2.计算类间散度矩阵：

$$\boldsymbol{S}_{\mathrm{B}} = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^{\mathrm{T}}$$

3.7 Fisher线性判别

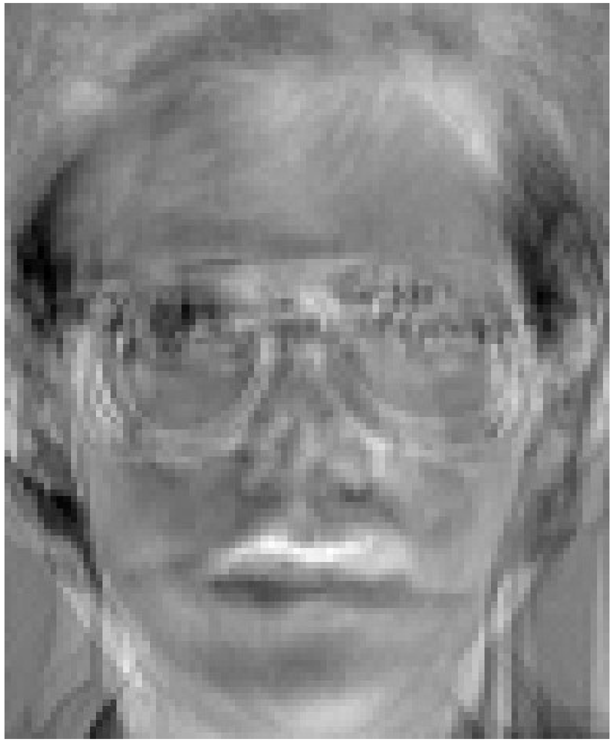
平均脸

平均脸

3.7 Fisher线性判别

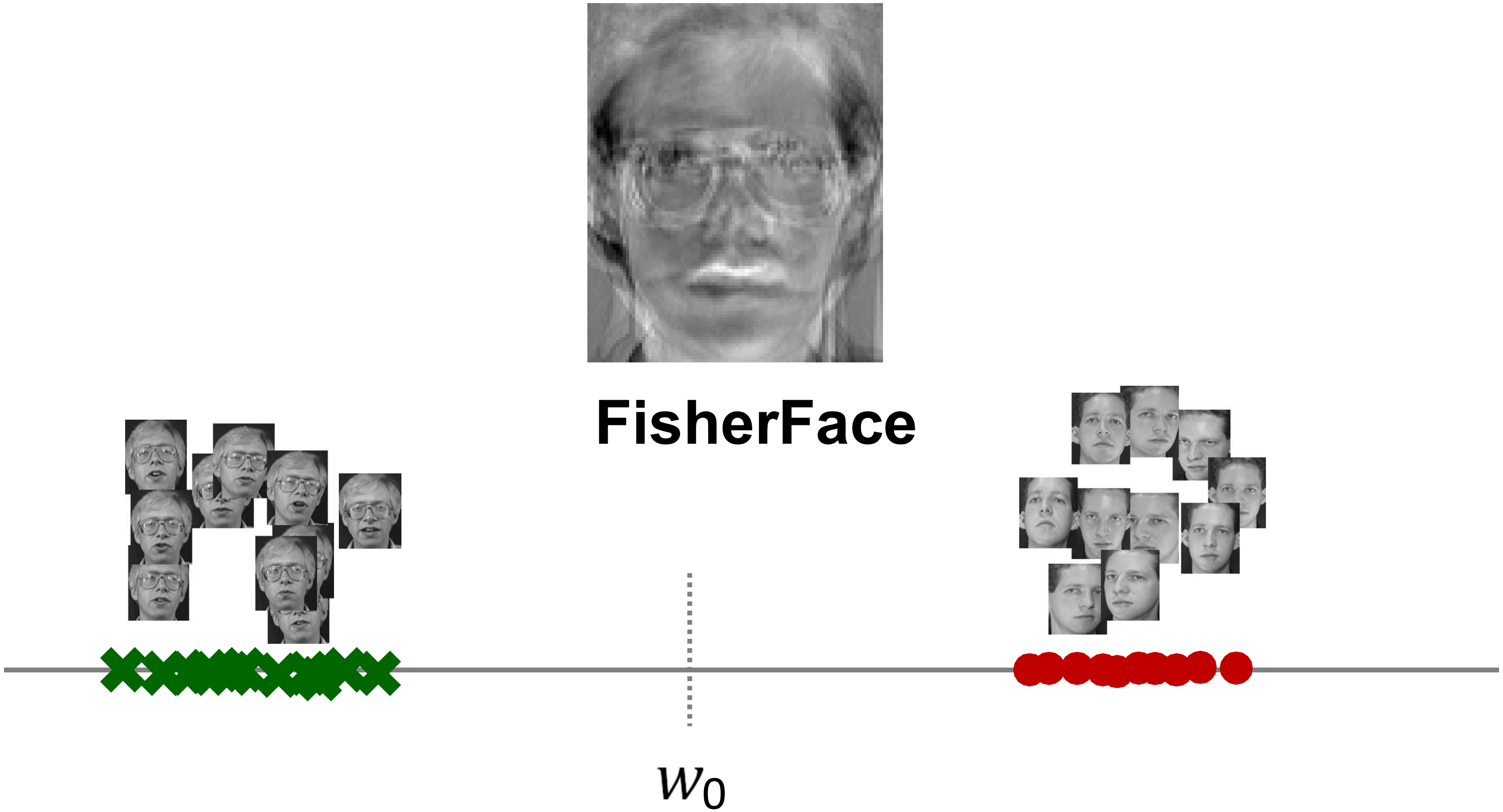
3.**求解特征值：**

$$\boldsymbol{S}_{\mathrm{W}}^{-1}\boldsymbol{S}_{\mathrm{B}}\boldsymbol{w} = \lambda \boldsymbol{w}$$



FisherFace w

3.7 Fisher线性判别



4.**进行判别** $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_n - w_0 \underset{<}{\overset{>}{}} 0 \qquad \boldsymbol{x}_n \in \begin{cases}\omega_1\\ \omega_2\end{cases}$

Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection， IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 19, NO. 7, JULY 1997

原始脸 FisherFace

End