模式识别与机器学习

Pattern Recognition and Machine Learning

课程内容

- **模式识别与机器学习概述**
- **模式识别与机器学习的基本方法**
  - 回归分析、线性判别函数、线性神经网络、核方法和支持向量机、决策树分类
  - 贝叶斯统计决策理论、概率密度函数估计
  - 无监督学习和聚类
  - 特征选择与提取

监督式学习方法

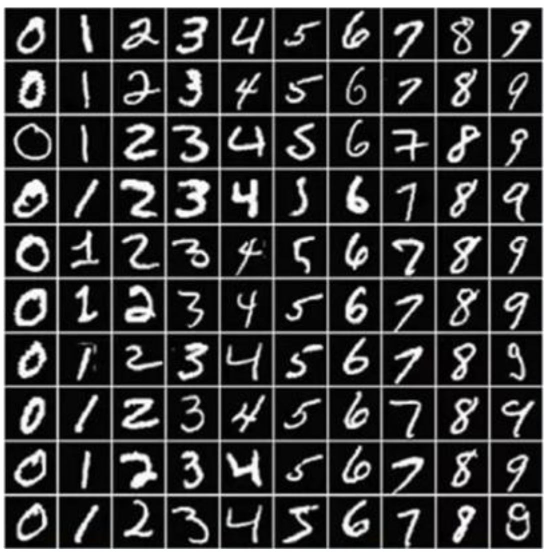
- 监督学习从给定的训练数据集中学习出数据及其标记的映射函数，当新的数据到来时，可以根据这个函数预测新数据的结果。训练集中的数据标记是由人标注的。
- 常见的监督学习算法包括回归分析和分类。
  - 回归：函数的输出是一个连续的值
  - 分类：函数的输出是一个离散的值（二分类、多分类）

课程内容

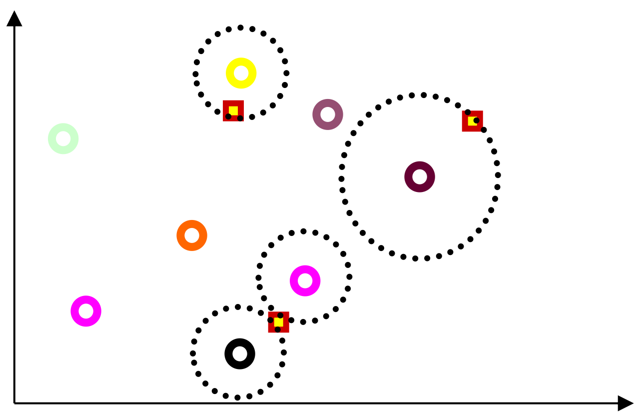
- **线性分类**
  - 分类任务简述
  - 最小距离分类器
  - 线性分类器 --- 决策
  - 线性分类器 --- 训练

最小距离分类器

- 一种简单的分类方法--最小距离分类法
  - 对每类训练数据（已知类别的数据）确定代表点。
  - 测试数据（未知类别的数据）的类别通过计算其与这些代表点的距离最近作决策。



mnist手写数据集



为减少代表点选择影响，
为每一种字符建立模板

$$\delta(\mathbf{x}_k, \mathbf{x}_l) = \left\| \mathbf{x}_k - \mathbf{x}_l \right\|$$

最小距离分类器

- 距离度量的性质：
  - 非负性 $dist(x_i, x_j) \geq 0$
  - 同一性 $dist(x_i, x_j) = 0$ 当且仅当 $x_i = x_j$
  - 对称性 $dist(x_i, x_j) = dist(x_j, x_i)$
  - 直递性 $dist(x_i, x_j) \leq dist(x_i, x_k) + dist(x_k, x_j)$
- 闵可夫斯基距离（Minkowski distance）

$$dist(x_i, x_j) = \left(\sum_{u=1}^{n} |x_{iu} - x_{ju}|^p \right)^{\frac{1}{p}}$$

  - p=1：曼哈顿距离（Manhattan distance）
  - p=2：欧氏距离（Euclidean distance）

最小距离分类器

- 距离度量的性质：
  - 非负性 $dist(x_i, x_j) \geq 0$
  - 同一性 $dist(x_i, x_j) = 0$ 当且仅当 $x_i = x_j$
  - 对称性 $dist(x_i, x_j) = dist(x_j, x_i)$
  - 直递性 $dist(x_i, x_j) \leq dist(x_i, x_k) + dist(x_k, x_j)$
- 马氏距离：

$$D_M(x) = \sqrt{(x-\mu)^T S^{-1} (x-\mu)}$$

其中 $\mu = (\mu_1, \mu_2, \cdots \mu_p)^T, x = (x_1, x_2, \cdots x_p)^T, S$ 是协方差矩阵

最小距离分类器--如何选定距离?

- 一种简单的分类方法--最小距离分类法

已知二维空间的 3 个点 $x_1=(1,1)^T, x2=(5,1)^T, x_3=(4,4)^T$，p 取不同值时，$L_p$ 距离下 x_1 的最近邻点。

解：$L_1(x_1,x_3)=6, L_2(x_1,x_3)=4.24, L_3(x_1,x_3)=3.78, L_4(x_1,x_3)=3.57$，无论 p 为何值，$L_p(x_1,x_2)=4$。

于是，p 等于 1 或 2 时，x_2 是 x_1 的最近邻，当 p 为 3 或 4 时，x_3 是 x_1 的最近邻。

$$d(X,Y)=[\sum_{i=1}^{n}(x_i-y_i)^p]^{1/p}$$

非负、对称性和三角不等式

最小距离分类器--如何选定距离?

■ 一种简单的分类方法--最小距离分类法



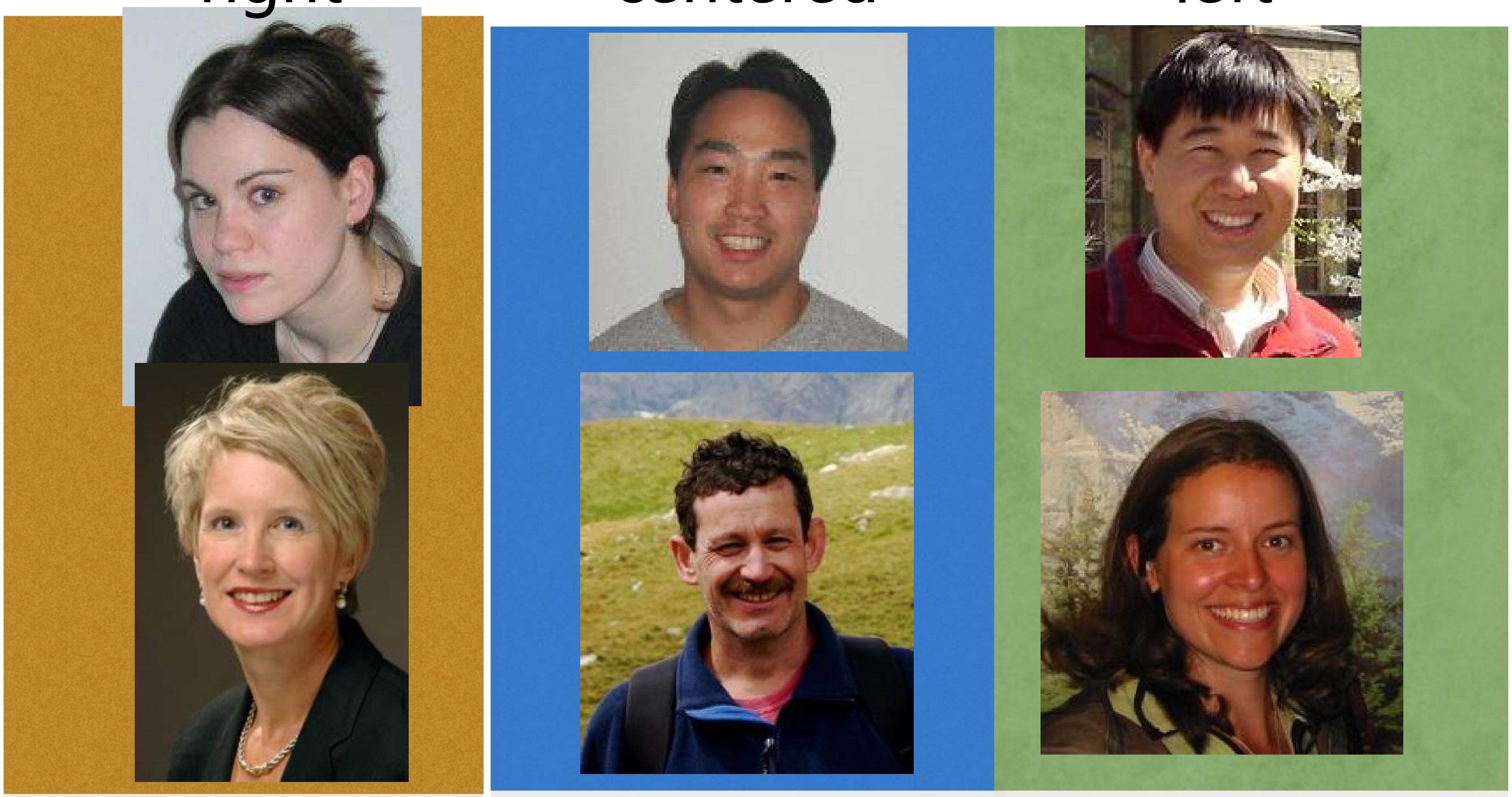



Which images are most similar?

最小距离分类器--如何选定距离?

- 一种简单的分类方法--最小距离分类法

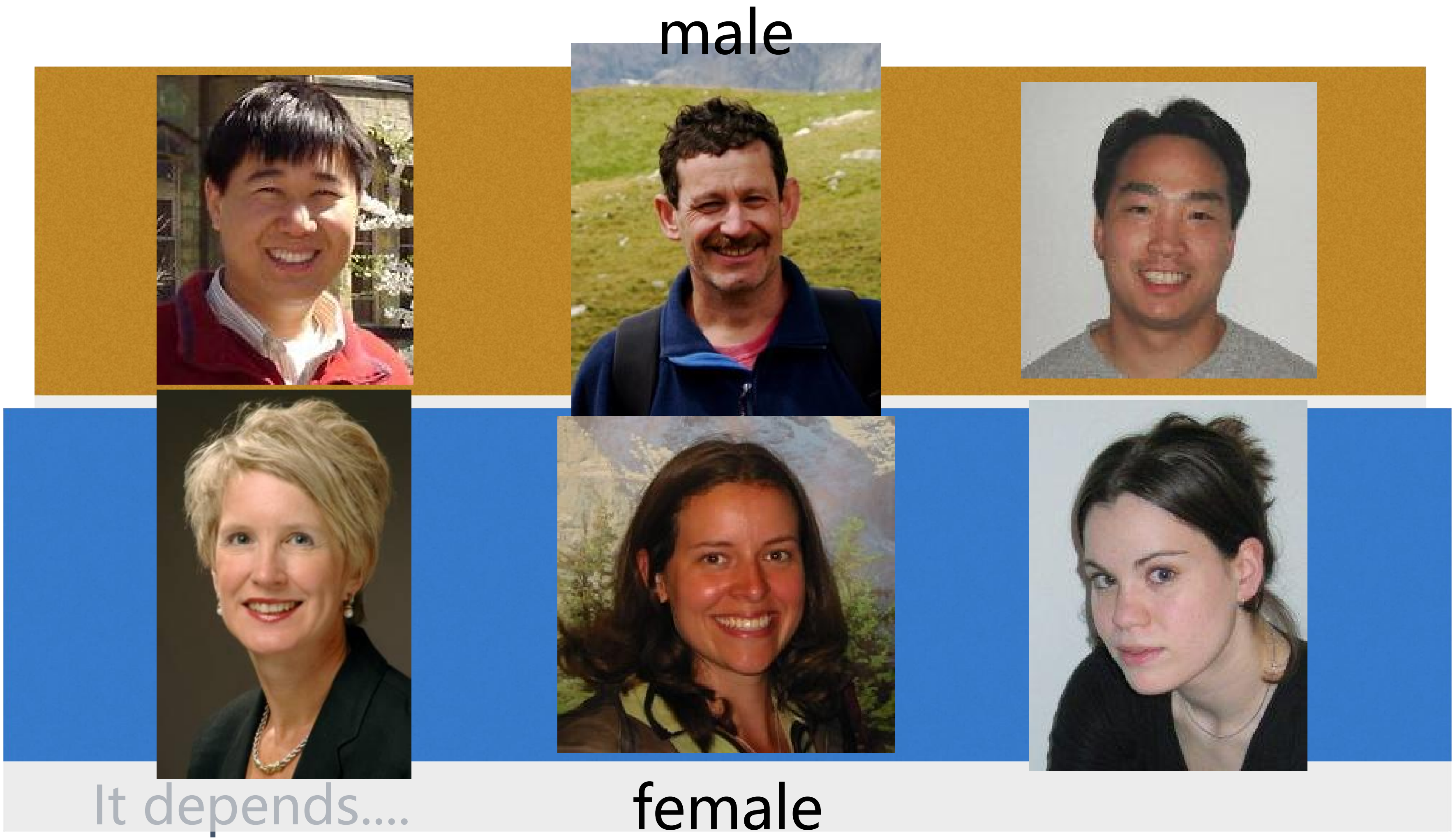
right centered left



It depends....

最小距离分类器--如何选定距离?

- 一种简单的分类方法--最小距离分类法



最小距离分类器--如何选定距离?

- 一种简单的分类方法--最小距离分类法

student

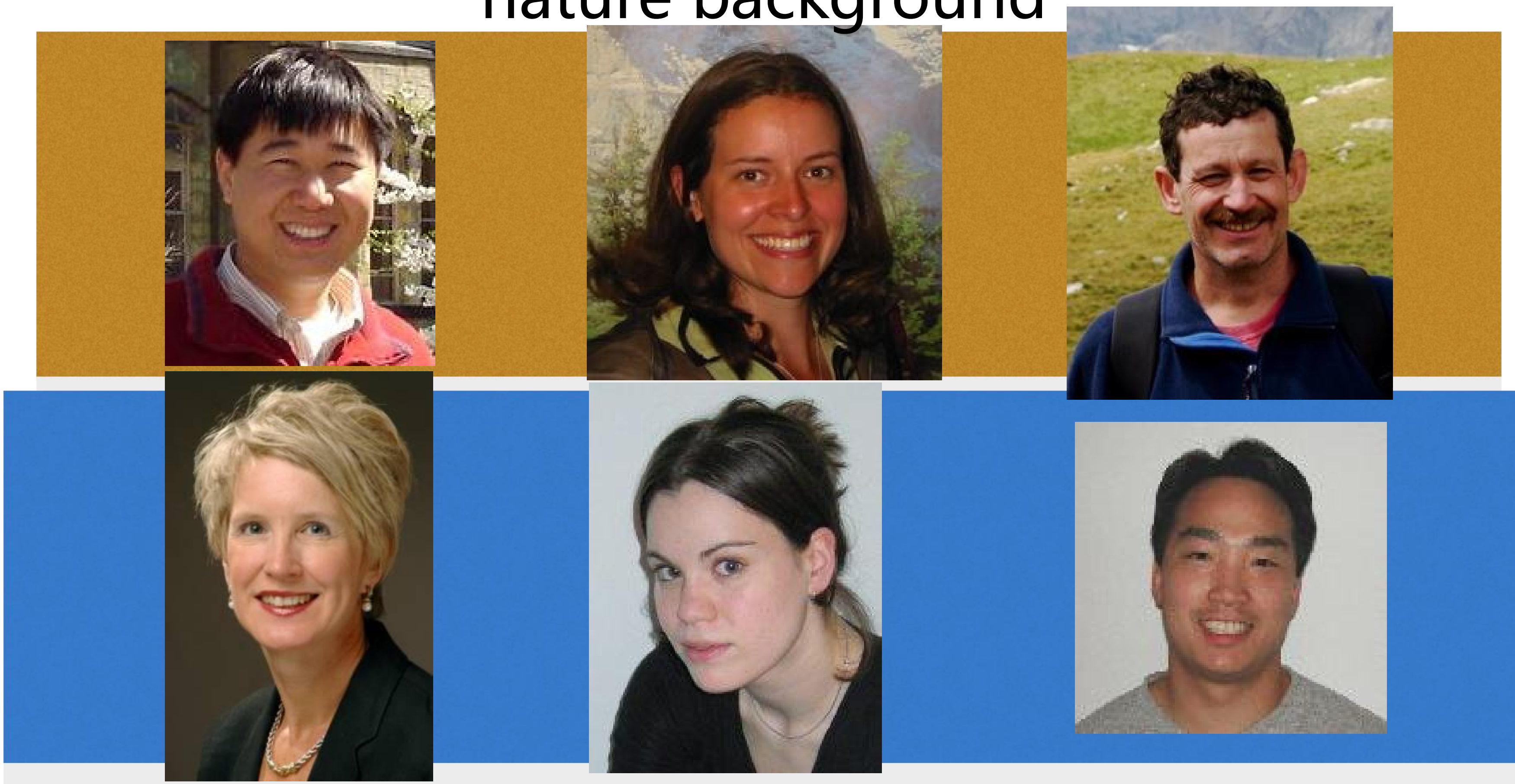
It depends....

professor

最小距离分类器--如何选定距离?

- 一种简单的分类方法--最小距离分类法

nature background

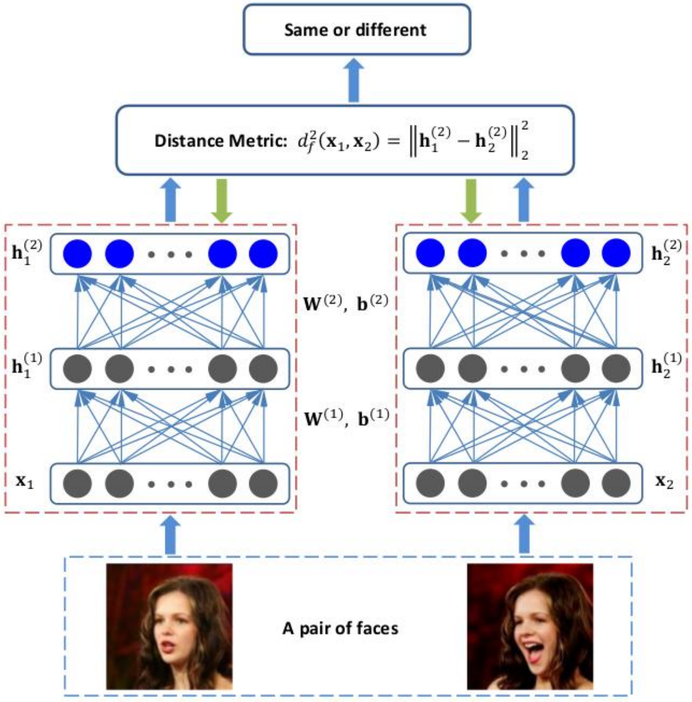


It depends....

plain background

最小距离分类器--度量学习(Metric Learning)

- Many problems may lack a well-defined, relevant distance metric
- A sensible similarity/distance metric may be highly task-dependent or semantic-dependent
- The simplest mapping is a linear transformation
- 学习出适用的距离



监督式学习--三段式解法

- **有监督学习： 找一個函数的能力**

Step 0: What kind of function do you want to find?

Regression, Classification, ……

Step 1: define a set of function

Linear Classifier
Probabilistic
SVM
Decision Tree
Deep Learning

Step 2: goodness of function

Regression(MSE)
Classification
……

Step 3: pick the best function

Gradient Descent
……

参考自李宏毅老师课件

线性分类器

- **有监督学习： 找一個函數的能力**

$y = b + w \cdot x$

A set of function

Model

$f_1, f_2 \cdots$

Goodness of function f

Training Data

$f_1($  $) =$ "cat"

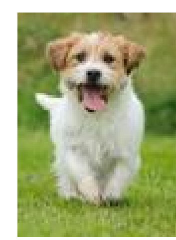
$f_1($  $) =$ "dog"

Better!

$f_2($  $) =$ "monkey"

$f_2($  $) =$ "snake"

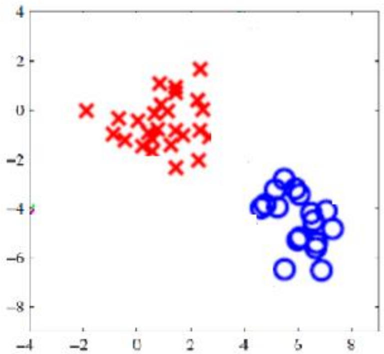
Supervised Learning (监督式学习)

function input:  

function output: "cat" "dog"

线性分类器

- 线性模型一般形式 （分类问题）



  - Function (Model):

$$x \Rightarrow f(x) = b + w \cdot x$$

  - Loss function:

$$L(f) = \sum_n \delta(f(x^n) \neq \hat{y}^n)$$

The number of times f get incorrect results on training data.

  - Find the best function:
    - Example: Perceptron, SVM

Not Today

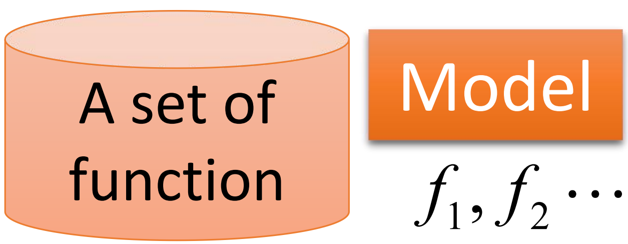
线性分类器

■ 线性模型一般形式 （分类问题）

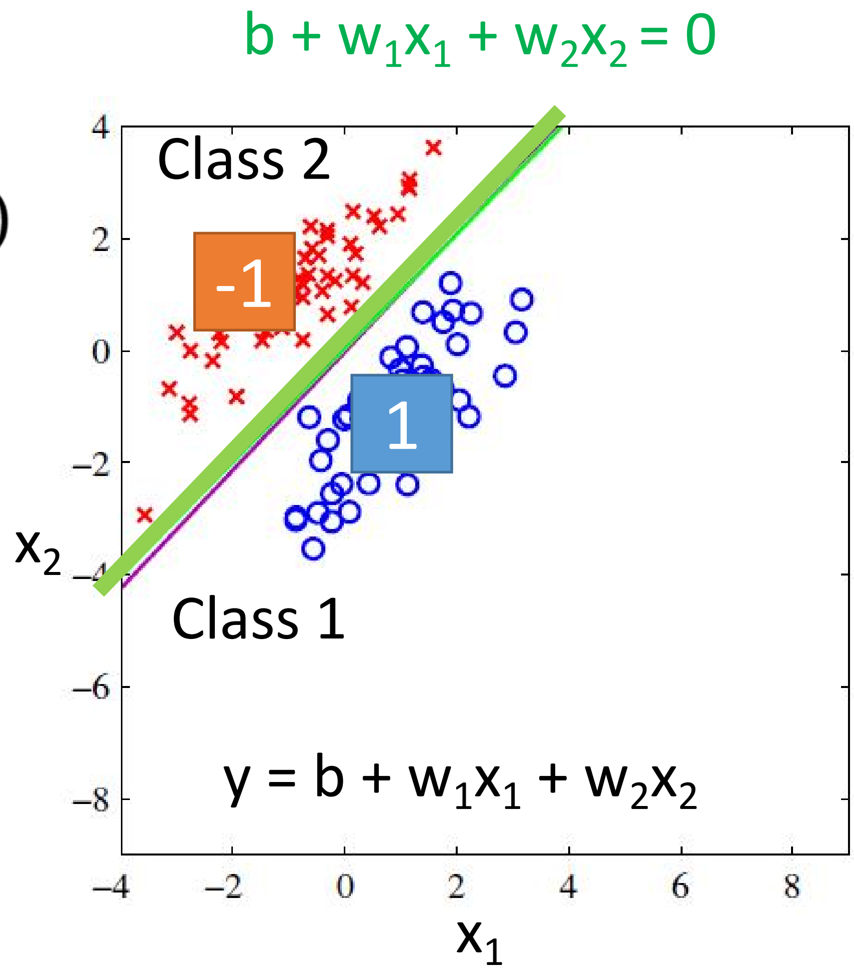
- Loss function:

$$L(f) = \sum_n \delta(f(x^n) \neq \hat{y}^n)$$

$y = b + w \cdot x$



$f_1, f_2 \cdots$



参考自李宏毅老师课件

线性分类器 --- 决策

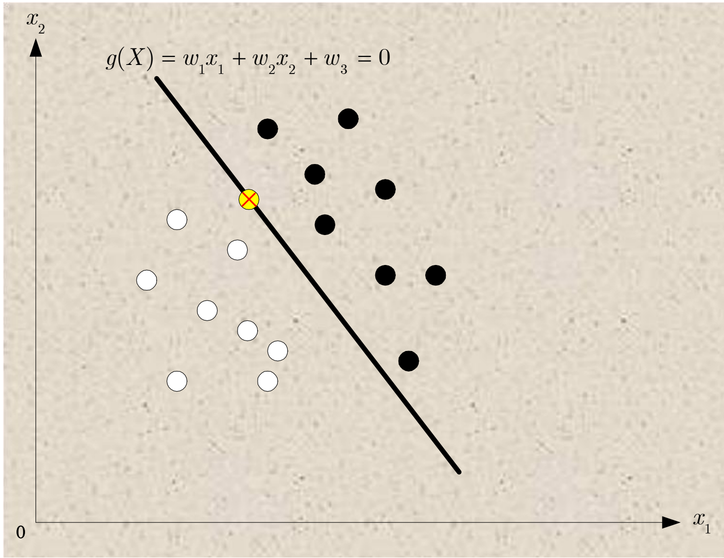
- 线性模型一般形式 （分类问题） y = b + w · x
  - For a d-dimensional feature $x = (x_1, x_2, ..., x_d)^T$

Linear Classifier can be represented as

$$x = (x_1, x_2, ..., x_d, 1)^T \qquad w = (w_1, w_2, ..., w_d, w_{d+1})^T$$

enhanced feature　　　　　enhanced weight

决策：W参数已知　$g(x) = w^T x$



Decision Rule

$$\begin{cases} \text{if } g(x) > 0, & x \in w_1 \\ \text{if } g(x) < 0, & x \in w_2 \end{cases}$$

线性分类器 --- 决策

- 线性模型一般形式 （分类问题） y = b + w · x
  - For a d-dimensional feature $x = (x_1, x_2, ..., x_d)^T$

Linear Classifier can be represented as

$$x = (x_1, x_2, ..., x_d, 1)^T \qquad w = (w_1, w_2, ..., w_d, w_{d+1})^T$$

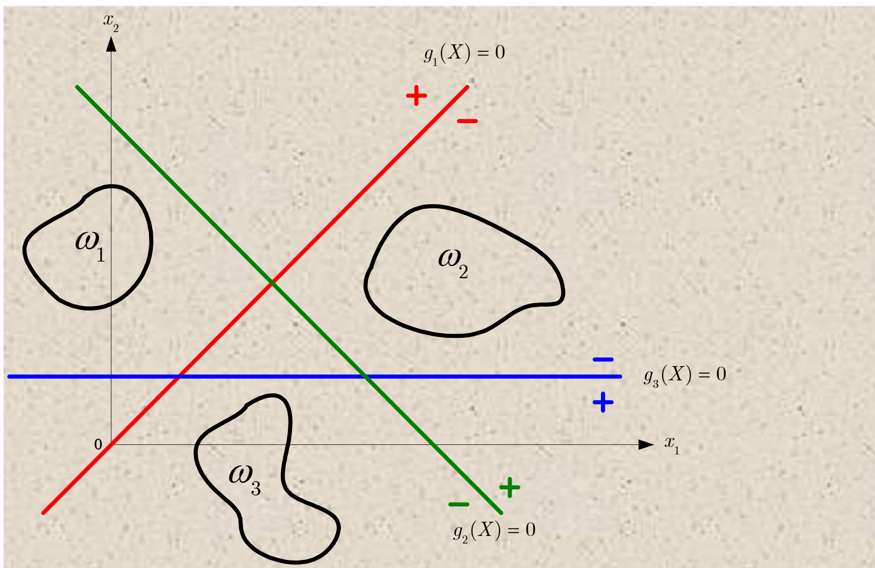
enhanced feature　　　enhanced weight

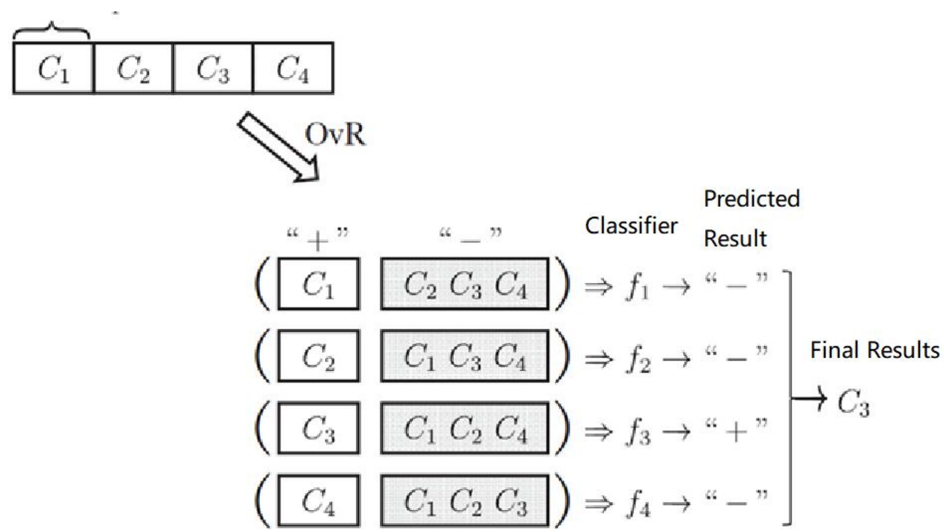
决策：W参数已知 $g(x) = w^T x$

- 扩展到多类(Using multi binary classification)
  - One vs. Rest, OvR
  - One vs. One, OvO
  - Many vs. Many, MvM

线性分类器 --- 决策

- 扩展到多类(Using multi binary classification)
  - One vs. Rest, OvR





$$g_i(X) = W_i^T X = \begin{cases} > 0, \text{当} X \in \omega_i \\ < 0, \textbf{Others} \end{cases}$$

线性分类器 --- 决策

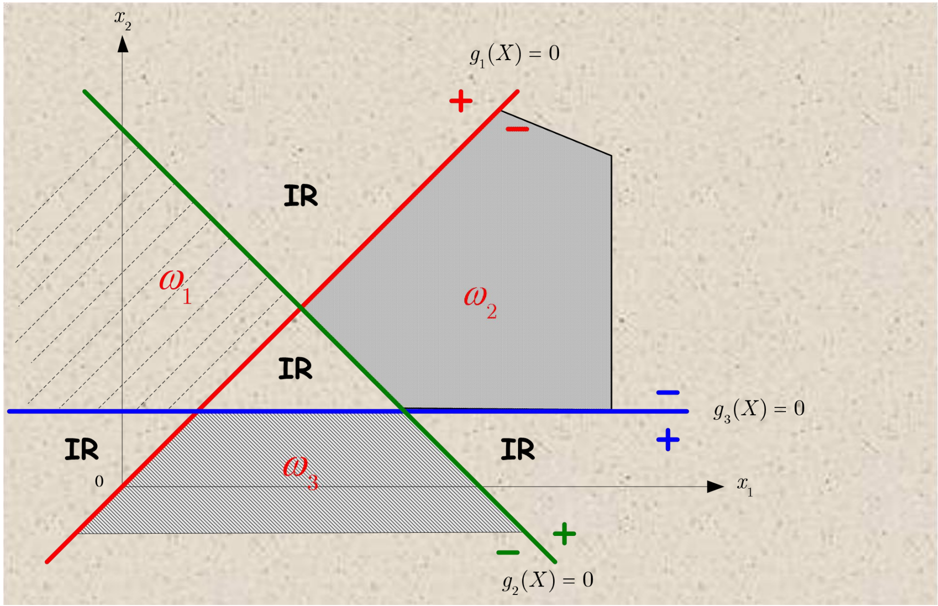
- 扩展到多类(Using multi binary classification)
  - One vs. Rest, OvR

$$g_i(X) = W_i^T X = \begin{cases} > 0, \text{当} X \in \omega_i \\ < 0, \textbf{Others} \end{cases}$$

$$g_1(x) = -10x_1 + 19x_2 + 19$$

$$g_2(x) = x_1 + x_2 - 5$$

$$g_3(x) = -2x_2 + 1$$



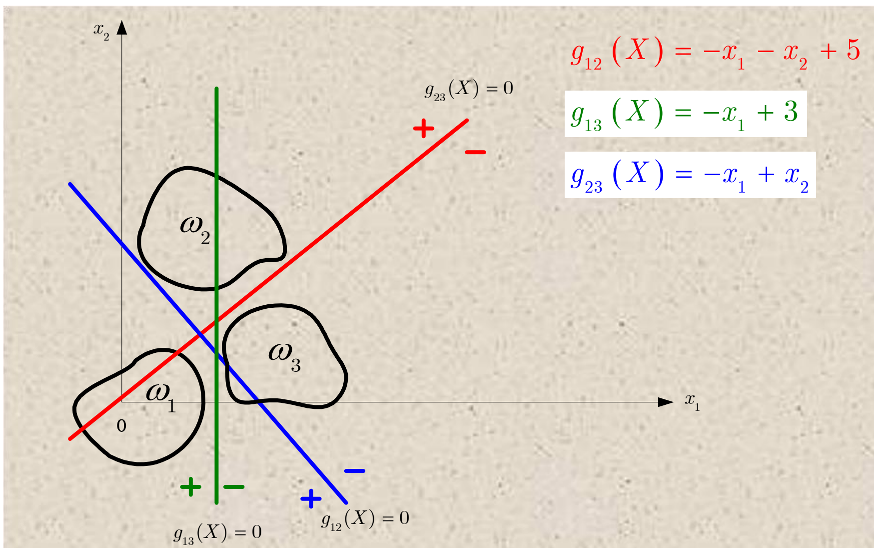
? $x = (6, 2)^T$

$$g_1(x) < 0 \quad g_2(x) > 0 \quad g_3(x) < 0$$

$$x \in w_2$$

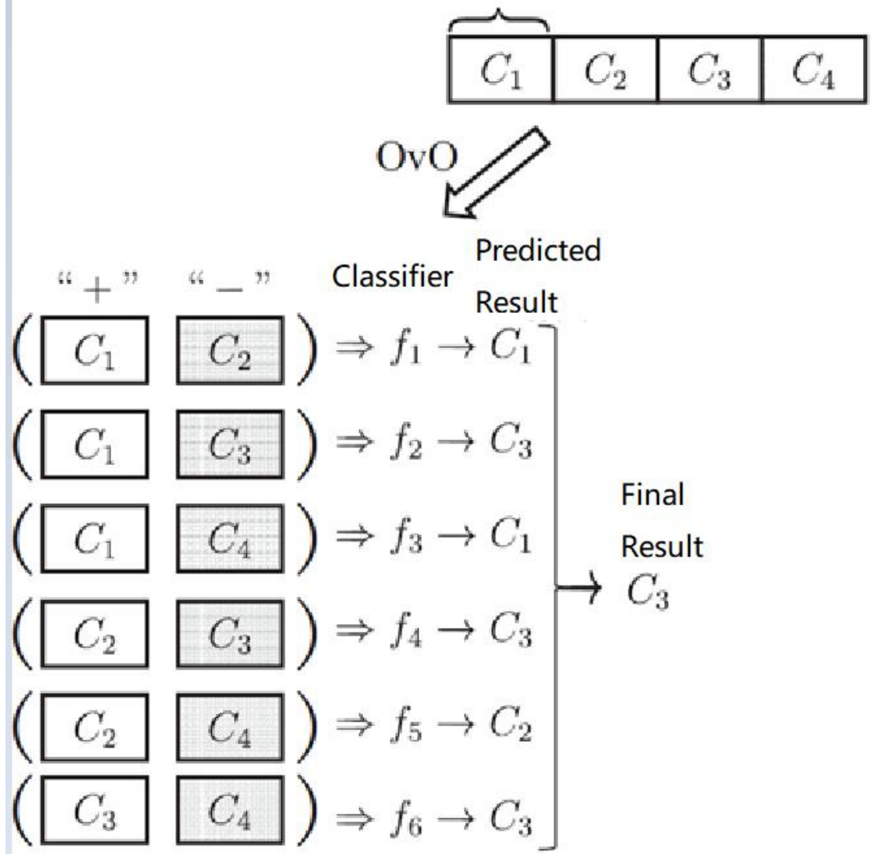
线性分类器 --- 决策

- 扩展到多类(Using multi binary classification)
  - One vs. One, OvO



$$g_{ij}(X) > 0, \forall j \neq i$$

$$X \in \omega_i$$



线性分类器 --- 决策

- 扩展到多类(Using multi binary classification)
  - One vs. Rest, OvR

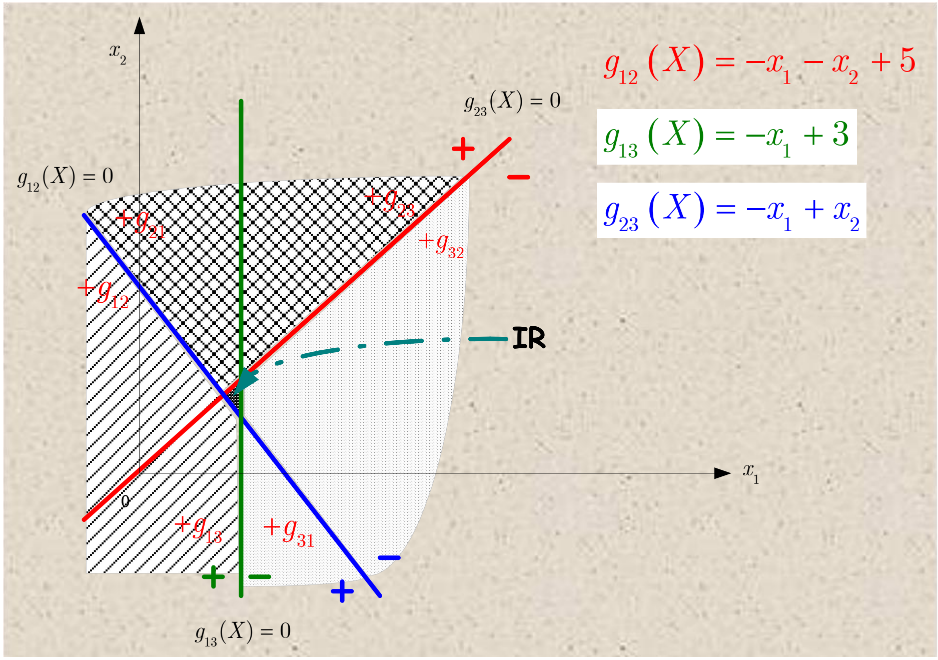
$$g_{ij}(X) > 0, \forall j \neq i \quad X \in \omega_i$$

$$g_{12}(X) = -x_1 - x_2 + 8.2$$

$$g_{13}(X) = -x_1 + 5.5$$

$$g_{23}(X) = -x_1 + x_2 + 0.2$$

? $\mathrm{x} = (8,3)^{\mathrm{T}}$



$$g_{12}(X) = -2.8 \quad g_{21}(X) = 2.8$$

$$g_{13}(X) = -2.5 \quad g_{31}(X) = 2.5$$

$$g_{23}(X) = -4.8 \quad g_{32}(X) = 4.8$$

$\Rightarrow$ $X \in \omega_3$

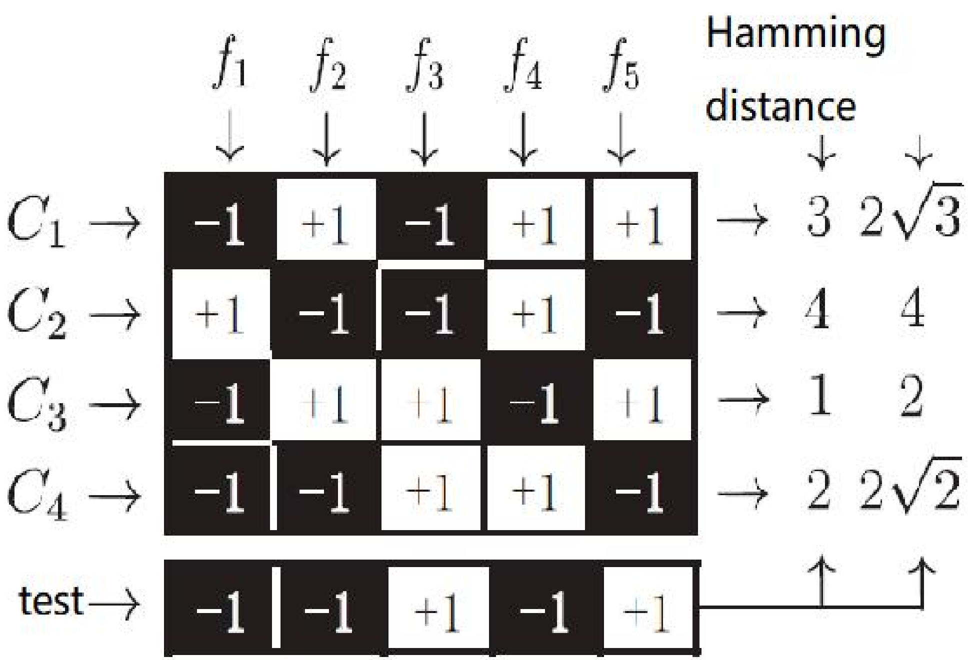
线性分类器 --- 决策

- 扩展到多类(Using multi binary classification)
  - Comparison between OVR and OVO

OVR	OVO
1) R classifiers. Need less storage cost and test time	1) R(R-1)/2 classifiers. Need more storage cost and test time
2) More training time	2) Less training time

线性分类器 --- 决策

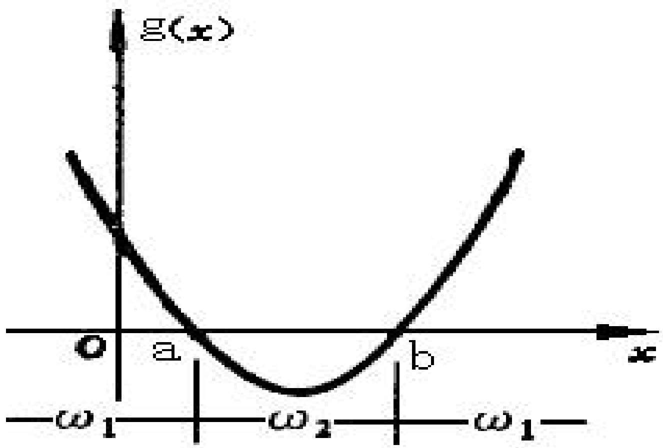
- 扩展到多类(Using multi binary classification)
  - ➢Using multi binary classification
    (Divide M two-classes Assembles)



线性分类器 --- 决策

- 扩展到非线性

For a nonlinear classifier



$$g(x) = (x-a)(x-b)$$

if $x < a$ or $x > b$, $g(x) > 0$ then $x \in w_1$

$a < x < b$, $g(x) < 0$ then $x \in w_2$

线性分类器 --- 决策

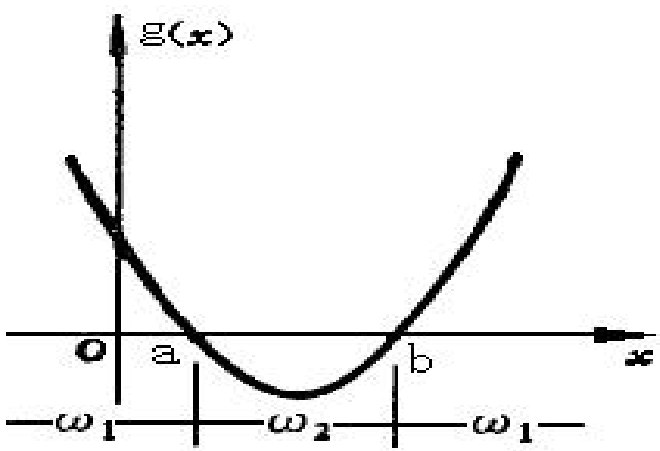
- 扩展到非线性

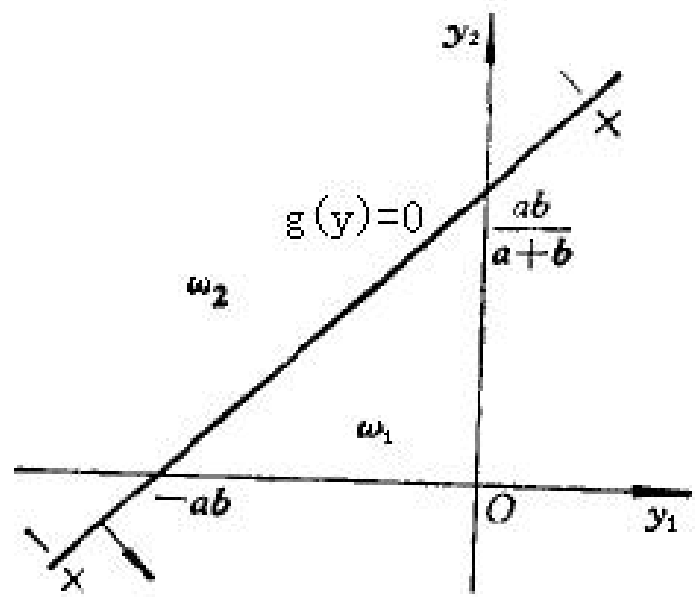
For a nonlinear classifier

$$y_1 = x^2 \quad y_2 = x \qquad g(x) = (x-a)(x-b) = \mathrm{x}^2 - (a+b)\mathrm{x} + ab$$

$$g(y) = w_1 y_1 + w_2 y_2 + w_3$$

$$w_1 = 1, w_2 = -(a+b), w_3 = ab$$





线性分类器 --- 决策

■ 扩展到非线性

For a nonlinear classifier

$$X^* = \left(f_1(X), f_2(X), \cdots, f_k(X)\right)^T, k > n$$

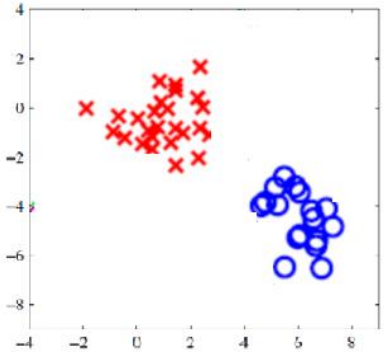
$$\begin{aligned} g(X) &= w_1 f_1(X) + w_2 f_2(X) + \cdots + w_k f_k(X) + w_{k+1} \\ &= W^T X^* \end{aligned}$$

$$W = (w_1, w_2, \cdots, w_{k+1})^T$$

$$X^* = \left(f_1(X), f_2(X), \cdots, f_k(X), 1\right)^T$$

线性分类器 --- 学习

■ 线性模型一般形式 （分类问题）



- Function (Model):

$$x \rightarrow f(x) = \mathrm{b} + \mathrm{w} \cdot \mathrm{x}$$

- Loss function:

$$L(f) = \sum_n \delta(f(x^n) \neq \hat{y}^n)$$

The number of times f get incorrect results on training data.

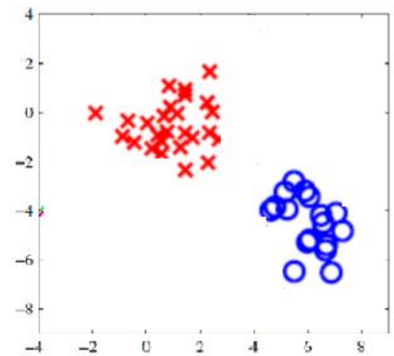
- Find the best function:
  - Example: Perceptron, LMSE,SVM

Not This Chapter

线性分类器 --- 学习

- 线性模型一般形式 （分类问题）
  - Loss function:

$$L(f) = \sum_n \delta(f(x^n) \neq \hat{y}^n)$$



学习：如何确定参数

- Find the best function:
  - Example: Perceptron

$$\{X_A, X_B\} \in \omega_1 \Rightarrow g(X) > 0$$
$$\{X_C, X_D\} \in \omega_2 \Rightarrow g(X) < 0$$

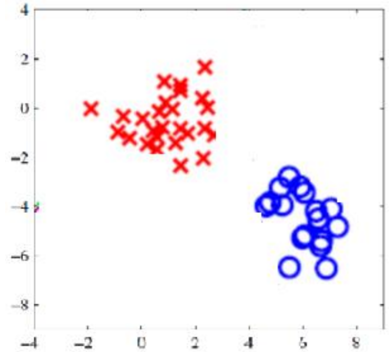
$$\begin{cases} x_{1A}w_1 + x_{2A}w_2 + w_3 > 0 \\ x_{1B}w_1 + x_{2B}w_2 + w_3 > 0 \\ x_{1C}w_1 + x_{2C}w_2 + w_3 < 0 \\ x_{1D}w_1 + x_{2D}w_2 + w_3 < 0 \end{cases}$$

$$\begin{cases} x_{1A}w_1 + x_{2A}w_2 + w_3 > 0 \\ x_{1B}w_1 + x_{2B}w_2 + w_3 > 0 \\ -x_{1C}w_1 - x_{2C}w_2 - w_3 > 0 \\ -x_{1D}w_1 - x_{2D}w_2 - w_3 > 0 \end{cases}$$

线性分类器 --- 学习

- 线性模型一般形式 （分类问题）
  - Loss function:

$$L(f) = \sum_n \delta(f(x^n) \neq \hat{y}^n)$$



学习：如何确定参数

$$W = (w_1, w_2, w_3)^T$$

$$\begin{cases} x_{1A}w_1 + x_{2A}w_2 + w_3 > 0 \\ x_{1B}w_1 + x_{2B}w_2 + w_3 > 0 \\ -x_{1C}w_1 - x_{2C}w_2 - w_3 > 0 \\ -x_{1D}w_1 - x_{2D}w_2 - w_3 > 0 \end{cases}$$

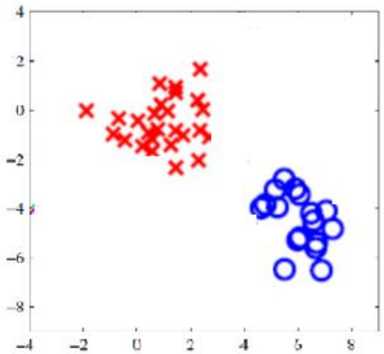
$$X = \begin{bmatrix} x_{1A} & x_{2A} & 1 \\ x_{1B} & x_{2B} & 1 \\ -x_{1C} & -x_{2C} & -1 \\ -x_{1D} & -x_{2D} & -1 \end{bmatrix}$$

$$XW > 0$$

线性分类器 --- 学习

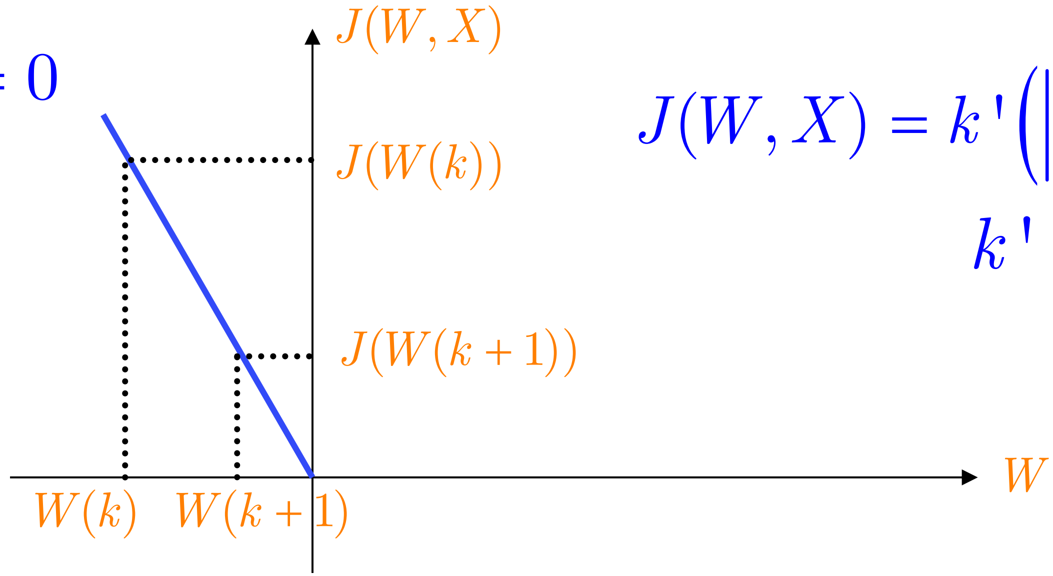
- 线性模型一般形式 （分类问题）
  - Loss function:

$$L(f) = \sum_{n} \delta(f(x^n) \neq \hat{y}^n)$$



学习：如何确定参数

$$J_{\min}(W, X) = 0$$



$$J(W, X) = k'\left(\left|W^T X\right| - W^T X\right)$$

$$k' = const$$

线性分类器 --- 学习

- 线性模型一般形式 （分类问题）
  - ➢Find the best function (Perception): **学习：如何确定参数**

$$J(W,X) = k'\left(\left|W^T X\right| - W^T X\right)$$

$$k' = const$$

Suppose $k' = \dfrac{1}{2}$

$$J(W,X) = \frac{1}{2}\left(\left|W^T X\right| - W^T X\right)$$

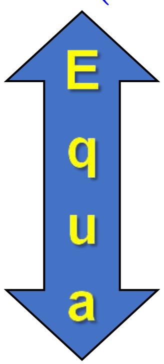
$$\frac{\partial J}{\partial W} = \frac{1}{2}\left[X\,\mathrm{sgn}(W^T X) - X\right]$$

$$\mathrm{sgn}(W^T X) = \begin{cases} 1 & W^T X > 0 \\ -1 & otherwise \end{cases}$$

线性分类器 --- 学习

- 线性模型一般形式 （分类问题）
  - Find the best function (Perception): **学习：如何确定参数**

$$W(k+1)=W(k)+\frac{C}{2}\left\{X(k)-X(k)\,\mathrm{sgn}[W^T(k)X(k)]\right\}$$



$$=\begin{cases} W(k) & W^T(k)X(k)>0 \\ W(k)+CX(k) & otherwise \end{cases}$$

if $X\in\omega_1$ and $g(X)>0$, then $W(k+1)=W(k)$

if $X\in\omega_1$ and $g(X)<0$, then $W(k+1)=W(k)+CX(k)$

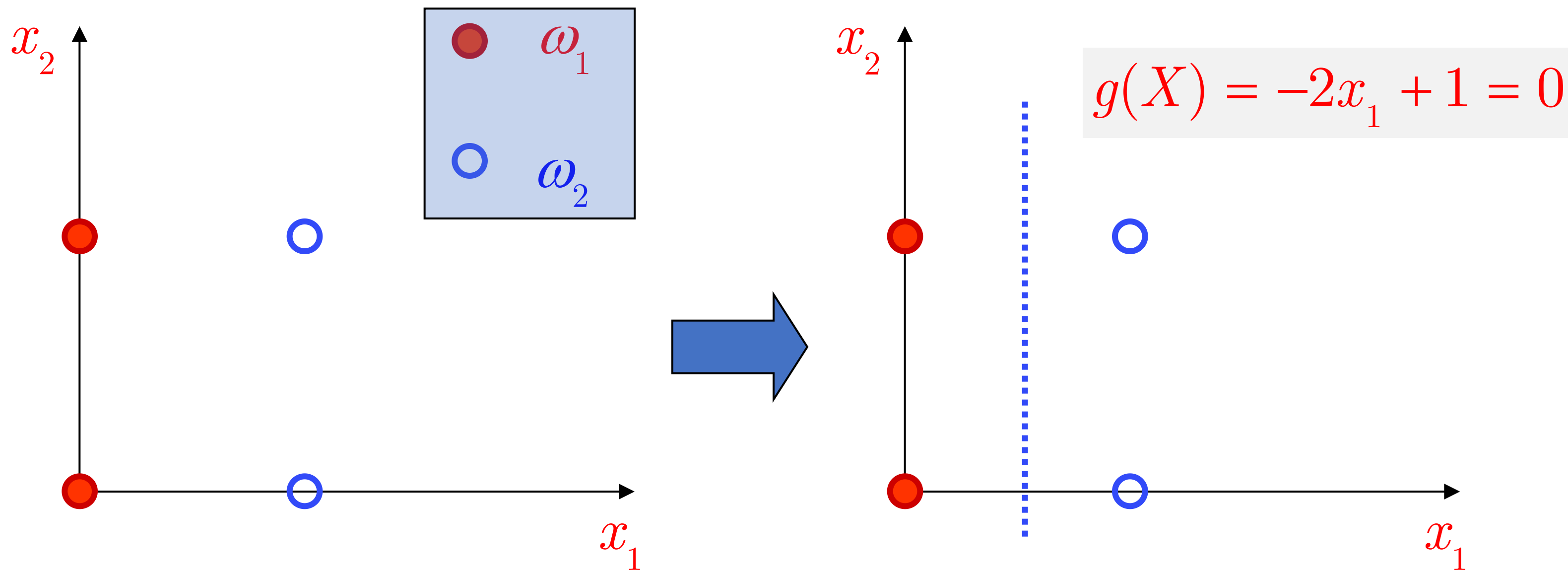
if $X\in\omega_2$ and $g(X)<0$, then $W(k+1)=W(k)$

if $X\in\omega_2$ and $g(X)>0$, then $W(k+1)=W(k)\text{-}CX(k)$

线性分类器 --- 学习

- 线性模型一般形式 （分类问题）
  - 一个例子(Perception)

学习：如何确定参数



Initial

$$X(1) = (0 \quad 0 \quad 1)^T \qquad X(2) = (0 \quad 1 \quad 1)^T$$

$$X(3) = (1 \quad 0 \quad 1)^T \qquad X(4) = (1 \quad 1 \quad 1)^T$$

$$C = 1 \qquad W(1) = (0,0,0)^T$$

线性分类器 --- 学习

- 线性模型一般形式 （分类问题）
  - 一个例子(Perception)

学习：如何确定参数

Penalty, Hold

Step1:

$$W^T(1)X(1) = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 0 \quad \xrightarrow{\text{Pen}} \quad W(2) = W(1) + X(1) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Step2:

$$W^T(2)X(2) = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 1 \quad \xrightarrow{\text{Hol}} \quad W(3) = W(2) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Step3:

$$W^T(3)X(3) = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 1 \quad \xrightarrow{\text{Pen}} \quad W(4) = W(3) - X(3) = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}$$

线性分类器 --- 学习

- 线性模型一般形式 （分类问题）
  - 一个例子(Perception)

学习：如何确定参数

Penalty, Hold

Step4:

$$W^T(4)X(4) = (-1 \quad 0 \quad 0)\begin{pmatrix}1\\1\\1\end{pmatrix} = -1 \xrightarrow{\text{Hol}} W(5) = W(4) = \begin{pmatrix}-1\\0\\0\end{pmatrix}$$

Step5: Recurrent

$$X(5) = X(1) = (0 \quad 0 \quad 1)^T$$

$$X(6) = X(2) = (0 \quad 1 \quad 1)^T$$

$$X(7) = X(3) = (1 \quad 0 \quad 1)^T$$

$$X(8) = X(4) = (1 \quad 1 \quad 1)^T$$

线性分类器 --- 学习

- 线性模型一般形式 （分类问题）
  - 一个例子(Perception)

$W^T(5)X(5) = 0$ **Pen** $W(6) = W(5) + X(5) = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$

$W^T(6)X(6) = 1$ **Hol** $W(7) = W(6) = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$

$W^T(7)X(7) = 0$ **Pen** $W(8) = W(7) - X(7) = \begin{pmatrix} -2 \\ 0 \\ 0 \end{pmatrix}$

$W^T(8)X(8) = -2$ **Hol** $W(9) = W(8) = \begin{pmatrix} -2 \\ 0 \\ 0 \end{pmatrix}$

线性分类器 --- 学习

■ 线性模型一般形式 （分类问题）

- 一个例子(Perception)

$W^T(9)X(9) = 0$ $\Rightarrow$ **Pen** $\Rightarrow$ $W(10) = W(9) + X(9) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$W^T(10)X(10) = 1 > 0$ $\Rightarrow$ **Hol** $\Rightarrow$ $W(11) = W(10) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$W^T(11)X(11) = -1 < 0$ $\Rightarrow$ **Hol** $\Rightarrow$ $W(12) = W(11) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$W^T(12)X(12) = -1 < 0$ $\Rightarrow$ **Hol** $\Rightarrow$ $W(13) = W(12) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

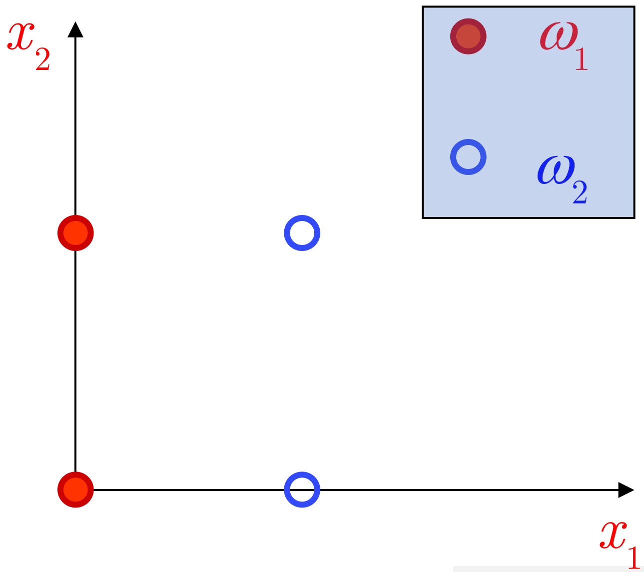
线性分类器 --- 学习

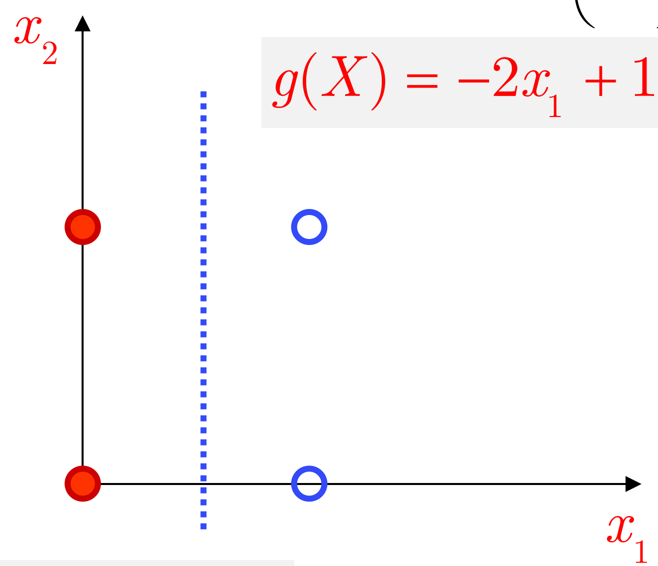
- 线性模型一般形式 （分类问题）
  - 一个例子(Perception)

$W^T(13)X(13) = 1 > 0$

Hol

$$W(14) = W(13) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$$





$g(X) = -2x_1 + 1 = 0$

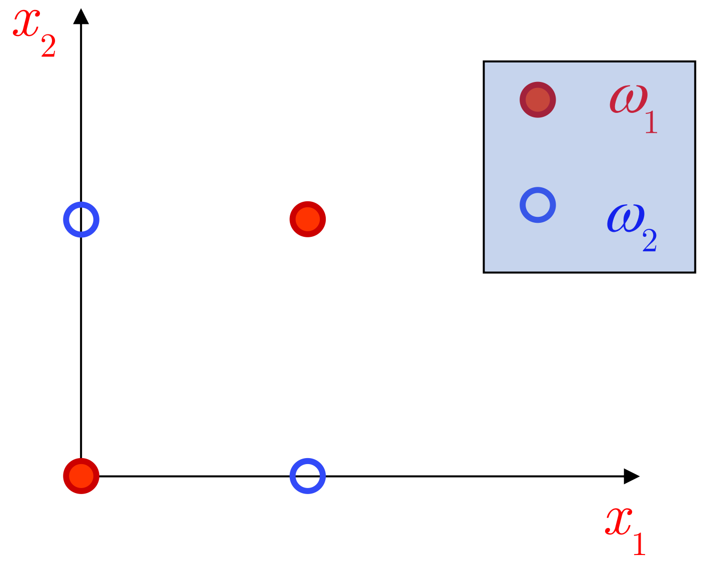
$$g(X) = -2x_1 + 1 = 0$$

The classifier can be found after limited iterations for a linear classification problem.

线性分类器 --- 学习

■ 线性模型一般形式 （分类问题）

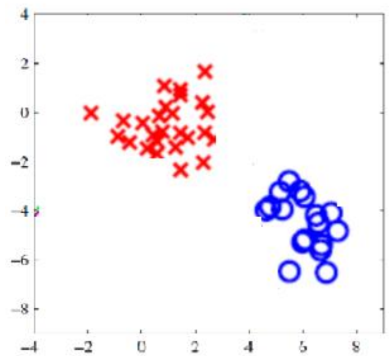
The problem of perception



Can not predict if a dataset is a linear classification problem

线性分类器 --- 学习

■ 线性模型一般形式 （分类问题）



➢Loss function:

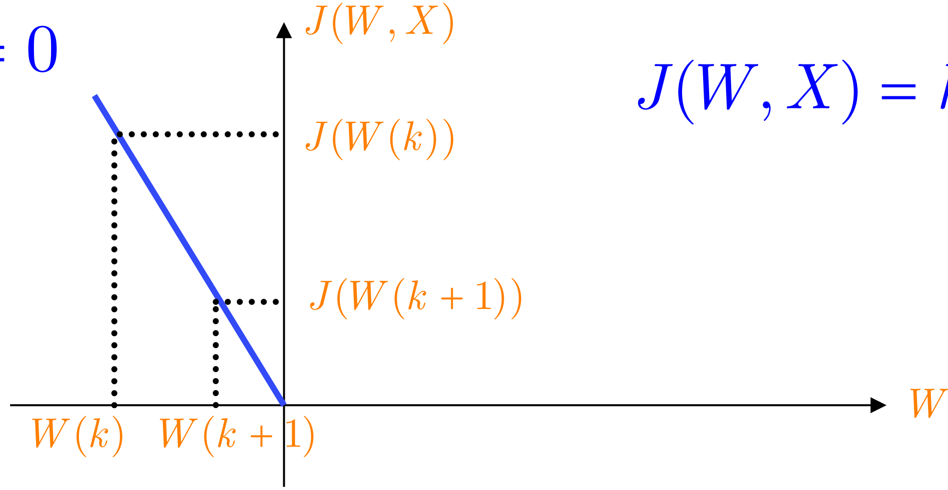
$$L(f) = \sum_n \delta(f(x^n) \neq \hat{y}^n)$$

学习：如何确定参数

- Find the best function:
  - Example: Perceptron

$$XW > 0$$

$$J_{\min}(W, X) = 0$$



$$J(W, X) = k'\left(\left|W^T X\right| - W^T X\right)$$

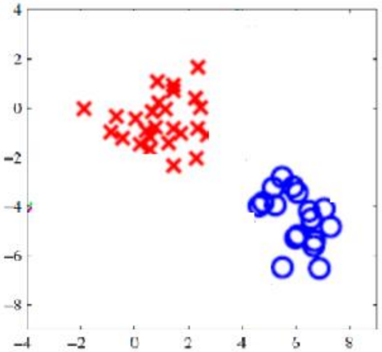
$$k' = const$$

线性分类器 --- 学习

■ 线性模型一般形式 （分类问题）

➢Loss function:

$$L(f) = \sum_{n} \delta(f(x^n) \neq \hat{y}^n)$$



学习：如何确定参数

- Find the best function:
  - Example: LMSE

$$XW > 0 \qquad J(W,X) = k'\left(\left|W^T X\right| - W^T X\right) \qquad k' = const$$

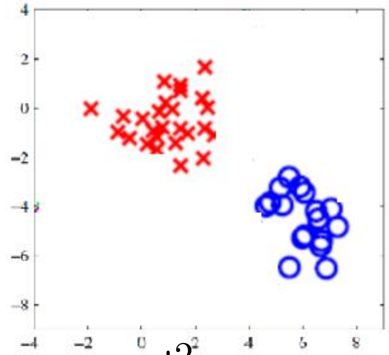
$$XW = b$$

$$J(W,X,b) = \frac{1}{2}\sum_{j=1}^{R}(W^T X_j - b_j) = \frac{1}{2}\left|XW - b\right|^2$$

$$b = (b_1, b_2, \cdots, b_R), b_i > 0 \qquad = \frac{1}{2}(XW - b)^T (XW - b)$$

线性分类器 --- 学习

- 线性模型一般形式 （分类问题）
  - ➢Find the best function (LMSE):



$$J(W,X,b)=\frac{1}{2}\sum_{j=1}^{R}(W^{T}X_{j}-b_{j})=\frac{1}{2}\left|XW-b\right|^{2}$$

$$=\frac{1}{2}(XW-b)^{T}(XW-b)$$

$$\begin{cases}\dfrac{\partial J}{\partial W}=0\\[2ex]\dfrac{\partial J}{\partial b}=0\end{cases}\quad\Rightarrow\quad\begin{cases}X^{T}(XW-b)=0\\XW-b=0\end{cases}$$

线性分类器 --- 学习

- 线性模型一般形式 （分类问题）
  - Find the best function (LMSE):

$$W = (X^T X)^{-1} X^T b \overset{\Delta}{=} X^{\#} b \qquad XW - b = 0$$

$$\Downarrow$$

$$b(k+1) = b(k) + \delta b(k)$$

$$\delta b(k) = \begin{cases} 0 & XW(k) - b(k) \le 0 \\ 2C[XW(k) - b(k)] & \textbf{otherwise} \end{cases}$$

$$= C[XW(k) - b(k) + | XW(k) - b(k) |] \qquad C > 0$$

define $E(k) = XW(k) - b(k)$

$$\delta b(k) = C[E(k) + | E(k) |]$$

线性分类器 --- 学习

- 线性模型一般形式 （分类问题）
  - Find the best function (LMSE):

    Updating equation $\delta b(k) = C[E(k) + | E(k) |]$

$$W(k+1) = X^{\#}b(k+1) = X^{\#}[b(k) + \delta b(k)]$$

$$= X^{\#}b(k) + X^{\#}\delta b(k)$$

$$= W(k) + CX^{\#}[E(k) + | E(k) |]$$

线性分类器 --- 学习

- 线性模型一般形式 （分类问题）
  - Find the best function (LMSE):

Updating equation

$$X^{\#}E(k) = X^{\#}[XW(k) - b(k)]$$

$$= (X^T X)^{-1} X^T [XX^{\#} - 1] b(k)$$

$$= (X^T X)^{-1} (X^T - X^T) b(k) = 0$$

$$W(1) = X^{\#} b(1)$$

$$E(k) = XW(k) - b(k)$$

$$W(k+1) = W(k) + CX^{\#}[E(k) + |E(k)|] = W(k) + CX^{\#}|E(k)|$$

$$b(k+1) = b(k) + C[E(k) + |E(k)|]$$

线性分类器 --- 学习

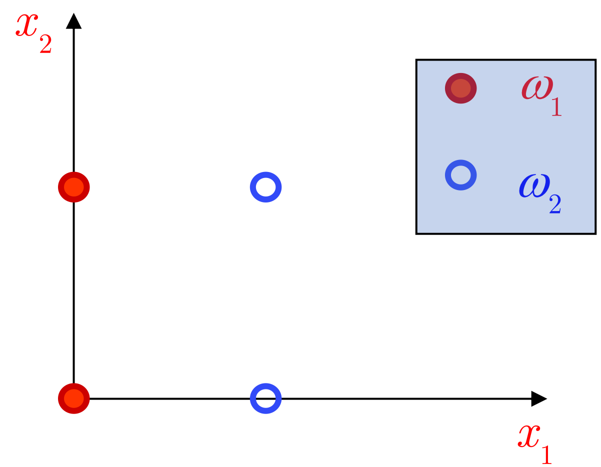
- 线性模型一般形式 （分类问题）
  - 一个例子(LMSE)

An example

学习：如何确定参数

$$\omega_1 : \{(0 \quad 0)^T, (0 \quad 1)^T\}$$

$$\omega_2 : \{(1 \quad 0)^T, (1 \quad 1)^T\}$$



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

$$X^{\#} = (X^T X)^{-1} X^T = \frac{1}{2}\begin{bmatrix} -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & -1 \\ \frac{3}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

线性分类器 --- 学习

■ 线性模型一般形式 （分类问题）

- 一个例子(LMSE)

学习：如何确定参数

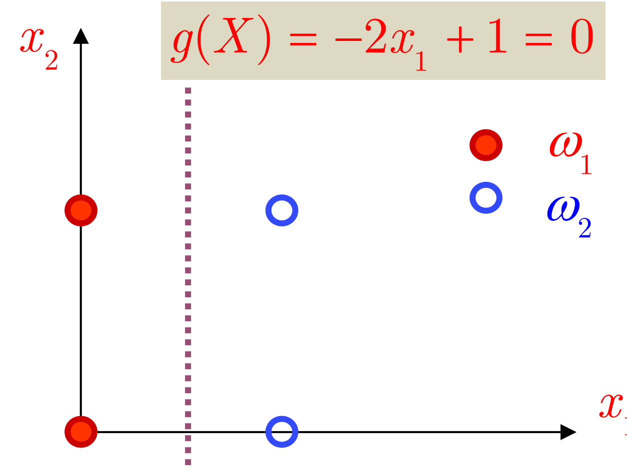
An example

Initial: $b(1) = (1 \quad 1 \quad 1 \quad 1)^T \qquad C = 1$

Iteration: $W(1) = X^{\#} b(1) = (-2 \quad 0 \quad 1)^T$

$$XW(1) = (1 \quad 1 \quad 1 \quad 1)^T$$

$$E(1) = XW(1) - b(1) = (0 \quad 0 \quad 0 \quad 0)^T$$



Ending

$$g(X) = -2x_1 + 1 = 0$$

线性分类器 --- 学习

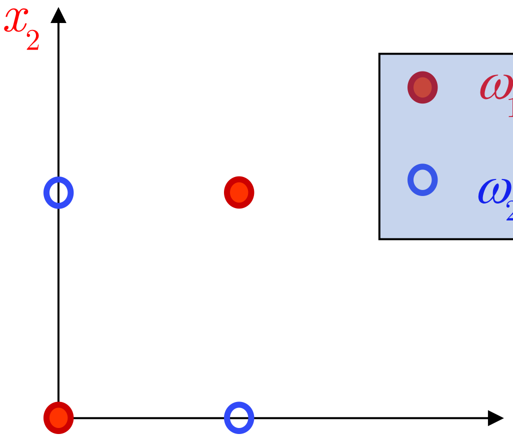
- 线性模型一般形式 （分类问题）
  - 一个例子(LMSE)

An example

学习：如何确定参数

$$\omega_1 : \{(0 \quad 0)^T, (1 \quad 1)^T\}$$

$$\omega_2 : \{(0 \quad 1)^T, (1 \quad 0)^T\}$$



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

$$X^{\#} = (X^T X)^{-1} X^T = \frac{1}{2}\begin{bmatrix} -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ \frac{3}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

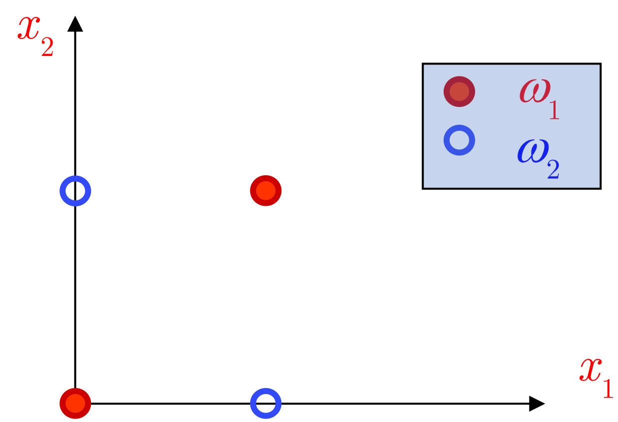
线性分类器 --- 学习

- 线性模型一般形式 （分类问题）
  - 一个例子(LMSE)

学习：如何确定参数

An example

Initial: $b(1) = (1 \quad 1 \quad 1 \quad 1)^T \quad C = 1$



Iteration: $W(1) = X^{\#}b(1) = (0 \quad 0 \quad 0)^T$

$$XW(1) = (0 \quad 0 \quad 0 \quad 0)^T$$

$$E(1) = XW(1) - b(1) = (-1 \quad -1 \quad -1 \quad -1)^T$$

Ending

Nondiscriminative

Thanks