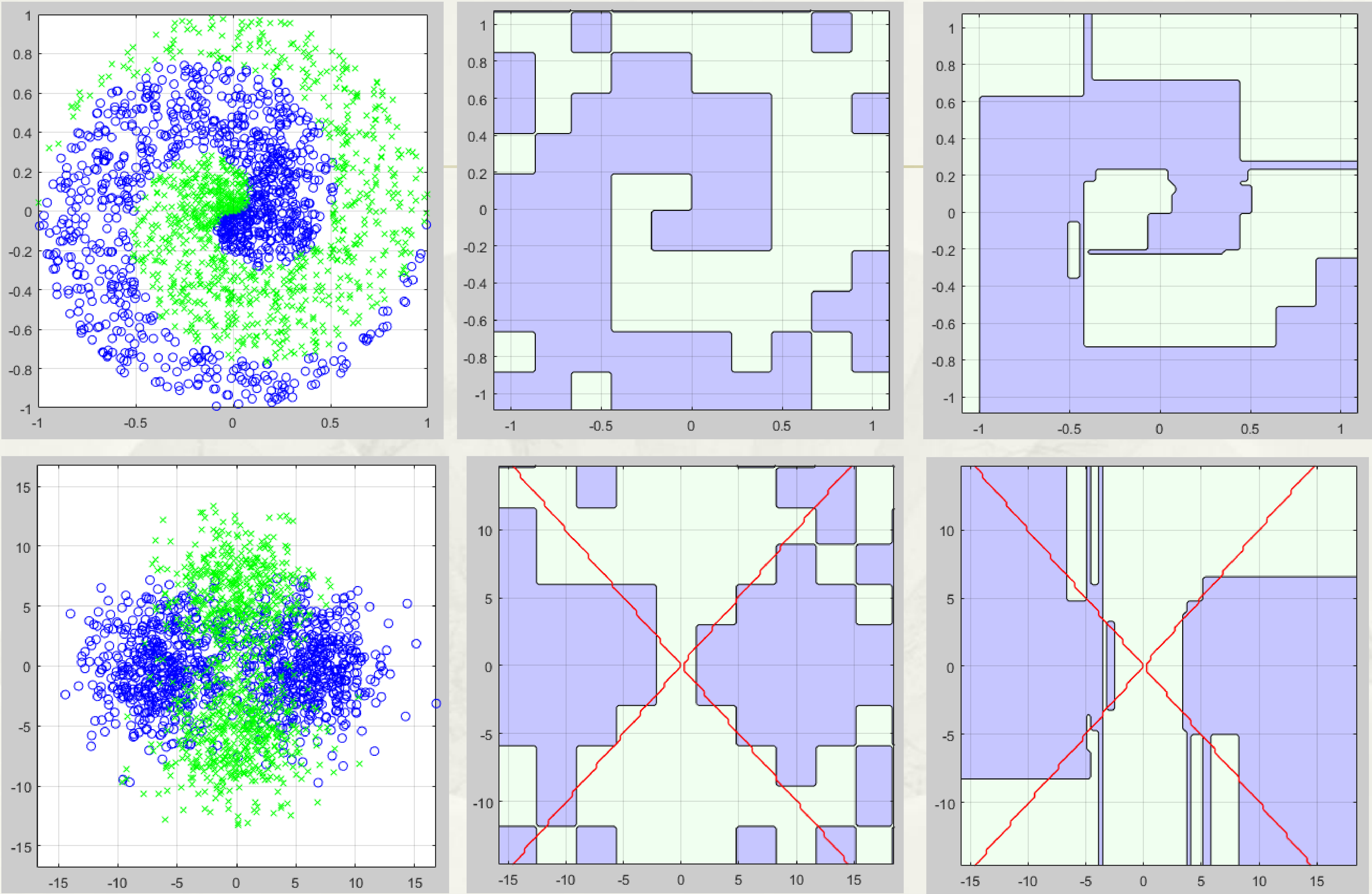
机器学习
Machine Learning

（三）统计学习

何劲松
中国科学技术大学


1 0.8 0.6 0.4 0.2 0 -0.2 -0.4 -0.6 -0.8 -1
-1 -0.5 0 0.5 1
1 0.8 0.6 0.4 0.2 0 -0.2 -0.4 -0.6 -0.8 -1
-1 -0.5 0 0.5 1
1 0.8 0.6 0.4 0.2 0 -0.2 -0.4 -0.6 -0.8 -1
-1 -0.5 0 0.5 1
15 10 5 0 -5 -10 -15
-15 -10 -5 0 5 10 15
10 5 0 -5 -10
-15 -10 -5 0 5 10 15
10 5 0 -5 -10
-15 -10 -5 0 5 10 15

iris数据

Iris数据集是常用的分类实验数据集，由Fisher收集整理。Iris也称鸢尾花卉数据集，是一种多重变量分析的数据集。数据集包含150个数据样本，分为3类，每类50个数据，每个数据包含4个属性。可通过花萼长度，花萼宽度，花瓣长度，花瓣宽度4个属性预测鸢尾花卉属于（Setosa，Versicolour，Virginica）三个种类中的哪一类。

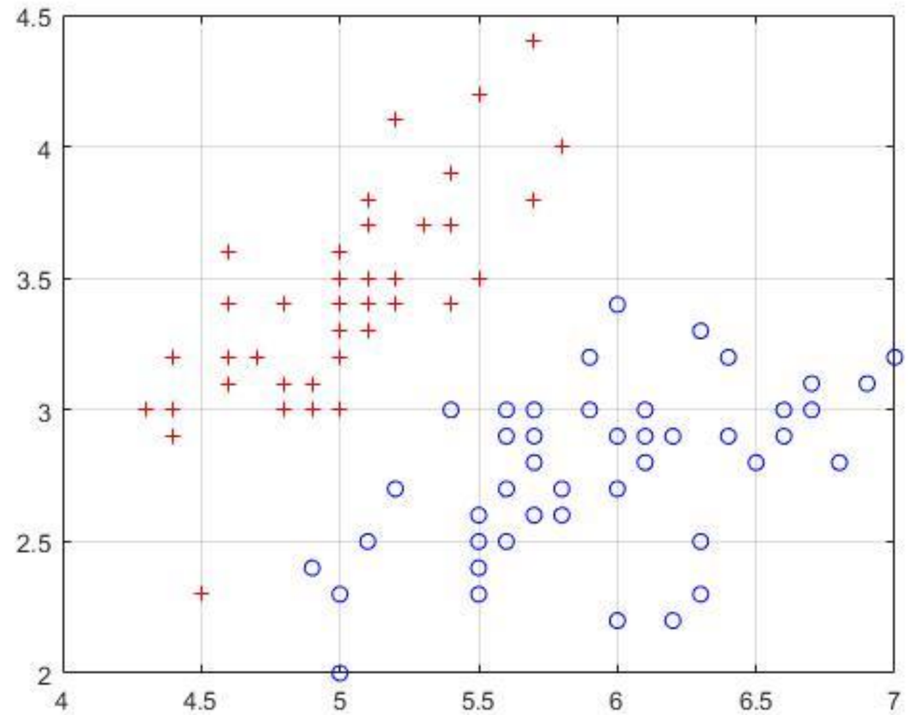
☞Title: Iris Plants Database

☞ Sources:
(a) Creator: R.A. Fisher
(b) Donor: Michael Marshall
(c) Date: July, 1988

☞ Attribute Information:
1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:
   -- Iris Setosa
   -- Iris Versicolour
   -- Iris Virginica

1	5.1,3.5,1.4,0.2,Iris-setosa
2	4.9,3.0,1.4,0.2,Iris-setosa
3	4.7,3.2,1.3,0.2,Iris-setosa
4	4.6,3.1,1.5,0.2,Iris-setosa
5	5.0,3.6,1.4,0.2,Iris-setosa
6	5.4,3.9,1.7,0.4,Iris-setosa
7	4.6,3.4,1.4,0.3,Iris-setosa
8	5.0,3.4,1.5,0.2,Iris-setosa
9	4.4,2.9,1.4,0.2,Iris-setosa
10	4.9,3.1,1.5,0.1,Iris-setosa
11	5.4,3.7,1.5,0.2,Iris-setosa
12	4.8,3.4,1.6,0.2,Iris-setosa
13	4.8,3.0,1.4,0.1,Iris-setosa
14	4.3,3.0,1.1,0.1,Iris-setosa
15	5.8,4.0,1.2,0.2,Iris-setosa
16	5.7,4.4,1.5,0.4,Iris-setosa
17	5.4,3.9,1.3,0.4,Iris-setosa
18	5.1,3.5,1.4,0.3,Iris-setosa
19	5.7,3.8,1.7,0.3,Iris-setosa
20	5.1,3.8,1.5,0.3,Iris-setosa
21	5.4,3.4,1.7,0.2,Iris-setosa
22	5.1,3.7,1.5,0.4,Iris-setosa
23	4.6,3.6,1.0,0.2,Iris-setosa
24	5.1,3.3,1.7,0.5,Iris-setosa
25	4.8,3.4,1.9,0.2,Iris-setosa
26	5.0,3.0,1.6,0.2,Iris-setosa
27	5.0,3.4,1.6,0.4,Iris-setosa
28	5.2,3.5,1.5,0.2,Iris-setosa
29	5.2,3.4,1.4,0.2,Iris-setosa
30	4.7,3.2,1.6,0.2,Iris-setosa
31	4.8,3.1,1.6,0.2,Iris-setosa
32	5.4,3.4,1.5,0.4,Iris-setosa
33	5.2,4.1,1.5,0.1,Iris-setosa
34	5.5,4.2,1.4,0.2,Iris-setosa
35	4.9,3.1,1.5,0.1,Iris-setosa
36	5.0,3.2,1.2,0.2,Iris-setosa
37	5.5,3.5,1.3,0.2,Iris-setosa
38	4.9,3.1,1.5,0.1,Iris-setosa
39	4.4,3.0,1.3,0.2,Iris-setosa
40	5.1,3.4,1.5,0.2,Iris-setosa
41	5.0,3.5,1.3,0.3,Iris-setosa
42	4.5,2.3,1.3,0.3,Iris-setosa
43	4.4,3.2,1.3,0.2,Iris-setosa
44	5.0,3.5,1.6,0.6,Iris-setosa
45	5.1,3.8,1.9,0.4,Iris-setosa
46	4.8,3.0,1.4,0.3,Iris-setosa
47	5.1,3.8,1.6,0.2,Iris-setosa
48	4.6,3.2,1.4,0.2,Iris-setosa
49	5.3,3.7,1.5,0.2,Iris-setosa
50	5.0,3.3,1.4,0.2,Iris-setosa
51	7.0,3.2,4.7,1.4,Iris-versicolor
52	6.4,3.2,4.5,1.5,Iris-versicolor
53	6.9,3.1,4.9,1.5,Iris-versicolor
54	5.5,2.3,4.0,1.3,Iris-versicolor
55	6.5,2.8,4.6,1.5,Iris-versicolor
56	5.7,2.8,4.5,1.3,Iris-versicolor
57	6.3,3.3,4.7,1.6,Iris-versicolor
58	4.9,2.4,3.3,1.0,Iris-versicolor
59	6.6,2.9,4.6,1.3,Iris-versicolor
60	5.2,2.7,3.9,1.4,Iris-versicolor
61	5.0,2.0,3.5,1.0,Iris-versicolor
62	5.9,3.0,4.2,1.5,Iris-versicolor
63	6.0,2.2,4.0,1.0,Iris-versicolor
64	6.1,2.9,4.7,1.4,Iris-versicolor
65	5.6,2.9,3.6,1.3,Iris-versicolor
66	6.7,3.1,4.4,1.4,Iris-versicolor
67	5.6,3.0,4.5,1.5,Iris-versicolor
68	5.8,2.7,4.1,1.0,Iris-versicolor
69	6.2,2.2,4.5,1.5,Iris-versicolor
70	5.6,2.5,3.9,1.1,Iris-versicolor
71	5.9,3.2,4.8,1.8,Iris-versicolor
72	6.1,2.8,4.0,1.3,Iris-versicolor
73	6.3,2.5,4.9,1.5,Iris-versicolor
74	6.1,2.8,4.7,1.2,Iris-versicolor
75	6.4,2.9,4.3,1.3,Iris-versicolor
76	6.6,3.0,4.4,1.4,Iris-versicolor
77	6.8,2.8,4.8,1.4,Iris-versicolor
78	6.7,3.0,5.0,1.7,Iris-versicolor
79	6.0,2.9,4.5,1.5,Iris-versicolor
80	5.7,2.6,3.5,1.0,Iris-versicolor
81	5.5,2.4,3.8,1.1,Iris-versicolor
82	5.5,2.4,3.7,1.0,Iris-versicolor
83	5.8,2.7,3.9,1.2,Iris-versicolor
84	6.0,2.7,5.1,1.6,Iris-versicolor
85	5.4,3.0,4.5,1.5,Iris-versicolor
86	6.0,3.4,4.5,1.6,Iris-versicolor
87	6.7,3.1,4.7,1.5,Iris-versicolor
88	6.3,2.3,4.4,1.3,Iris-versicolor
89	5.6,3.0,4.1,1.3,Iris-versicolor
90	5.5,2.5,4.0,1.3,Iris-versicolor
91	5.5,2.6,4.4,1.2,Iris-versicolor
92	6.1,3.0,4.6,1.4,Iris-versicolor
93	5.8,2.6,4.0,1.2,Iris-versicolor
94	5.0,2.3,3.3,1.0,Iris-versicolor
95	5.6,2.7,4.2,1.3,Iris-versicolor
96	5.7,3.0,4.2,1.2,Iris-versicolor
97	5.7,2.9,4.2,1.3,Iris-versicolor
98	6.2,2.9,4.3,1.3,Iris-versicolor
99	5.1,2.5,3.0,1.1,Iris-versicolor
100	5.7,2.8,4.1,1.3,Iris-versicolor
101	6.3,3.3,6.0,2.5,Iris-virginica
102	5.8,2.7,5.1,1.9,Iris-virginica
103	7.1,3.0,5.9,2.1,Iris-virginica
104	6.3,2.9,5.6,1.8,Iris-virginica
105	6.5,3.0,5.8,2.2,Iris-virginica
106	7.6,3.0,6.6,2.1,Iris-virginica
107	4.9,2.5,4.5,1.7,Iris-virginica
108	7.3,2.9,6.3,1.8,Iris-virginica
109	6.7,2.5,5.8,1.8,Iris-virginica
110	7.2,3.6,6.1,2.5,Iris-virginica
111	6.5,3.2,5.1,2.0,Iris-virginica
112	6.4,2.7,5.3,1.9,Iris-virginica
113	6.8,3.0,5.5,2.1,Iris-virginica
114	5.7,2.5,5.0,2.0,Iris-virginica
115	5.8,2.8,5.1,2.4,Iris-virginica
116	6.4,3.2,5.3,2.3,Iris-virginica
117	6.5,3.0,5.5,1.8,Iris-virginica
118	7.7,3.8,6.7,2.2,Iris-virginica
119	7.7,2.6,6.9,2.3,Iris-virginica
120	6.0,2.2,5.0,1.5,Iris-virginica
121	6.9,3.2,5.7,2.3,Iris-virginica
122	5.6,2.8,4.9,2.0,Iris-virginica
123	7.7,2.8,6.7,2.0,Iris-virginica
124	6.3,2.7,4.9,1.8,Iris-virginica
125	6.7,3.3,5.7,2.1,Iris-virginica
126	7.2,3.2,6.0,1.8,Iris-virginica
127	6.2,2.8,4.8,1.8,Iris-virginica
128	6.1,3.0,4.9,1.8,Iris-virginica
129	6.4,2.8,5.6,2.1,Iris-virginica
130	7.2,3.0,5.8,1.6,Iris-virginica
131	7.4,2.8,6.1,1.9,Iris-virginica
132	7.9,3.8,6.4,2.0,Iris-virginica
133	6.4,2.8,5.6,2.2,Iris-virginica
134	6.3,2.8,5.1,1.5,Iris-virginica
135	6.1,2.6,5.6,1.4,Iris-virginica
136	7.7,3.0,6.1,2.3,Iris-virginica
137	6.3,3.4,5.6,2.4,Iris-virginica
138	6.4,3.1,5.5,1.8,Iris-virginica
139	6.0,3.0,4.8,1.8,Iris-virginica
140	6.9,3.1,5.4,2.1,Iris-virginica
141	6.7,3.1,5.6,2.4,Iris-virginica
142	6.9,3.1,5.1,2.3,Iris-virginica
143	5.8,2.7,5.1,1.9,Iris-virginica
144	6.8,3.2,5.9,2.3,Iris-virginica
145	6.7,3.3,5.7,2.5,Iris-virginica
146	6.7,3.0,5.2,2.3,Iris-virginica
147	6.3,2.5,5.0,1.9,Iris-virginica
148	6.5,3.0,5.2,2.0,Iris-virginica
149	6.2,3.4,5.4,2.3,Iris-virginica
150	5.9,3.0,5.1,1.8,Iris-virginica

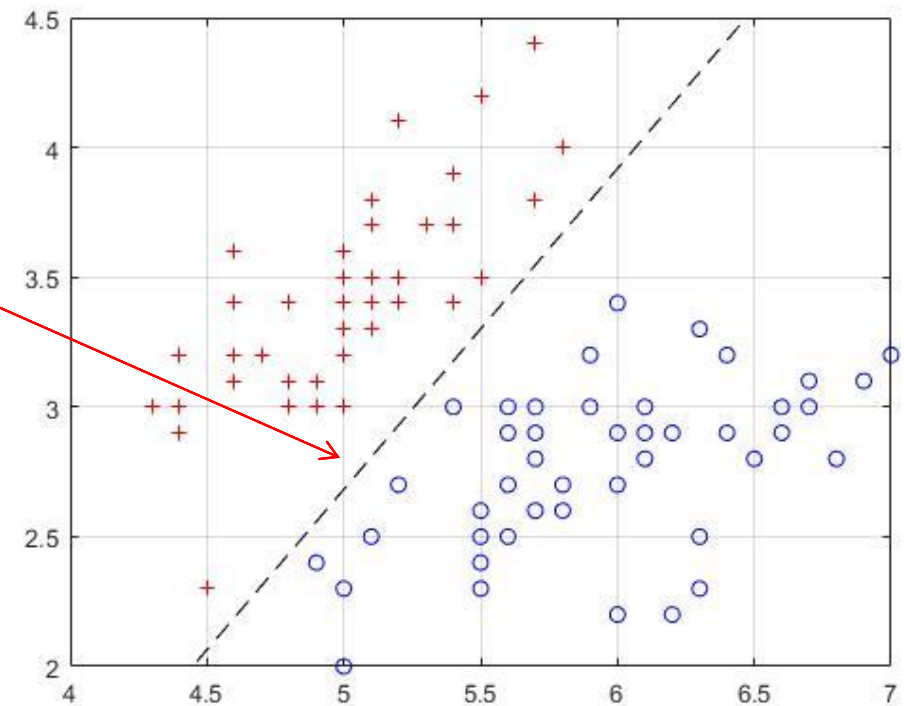
22	5.1,3.7,1.5,0.4,Iris-setosa
23	4.6,3.6,1.0,0.2,Iris-setosa
24	5.1,3.3,1.7,0.5,Iris-setosa
25	4.8,3.4,1.9,0.2,Iris-setosa



取iris数据的第1维和第2维特征，观察第1类和第2类样本在空间的分布。

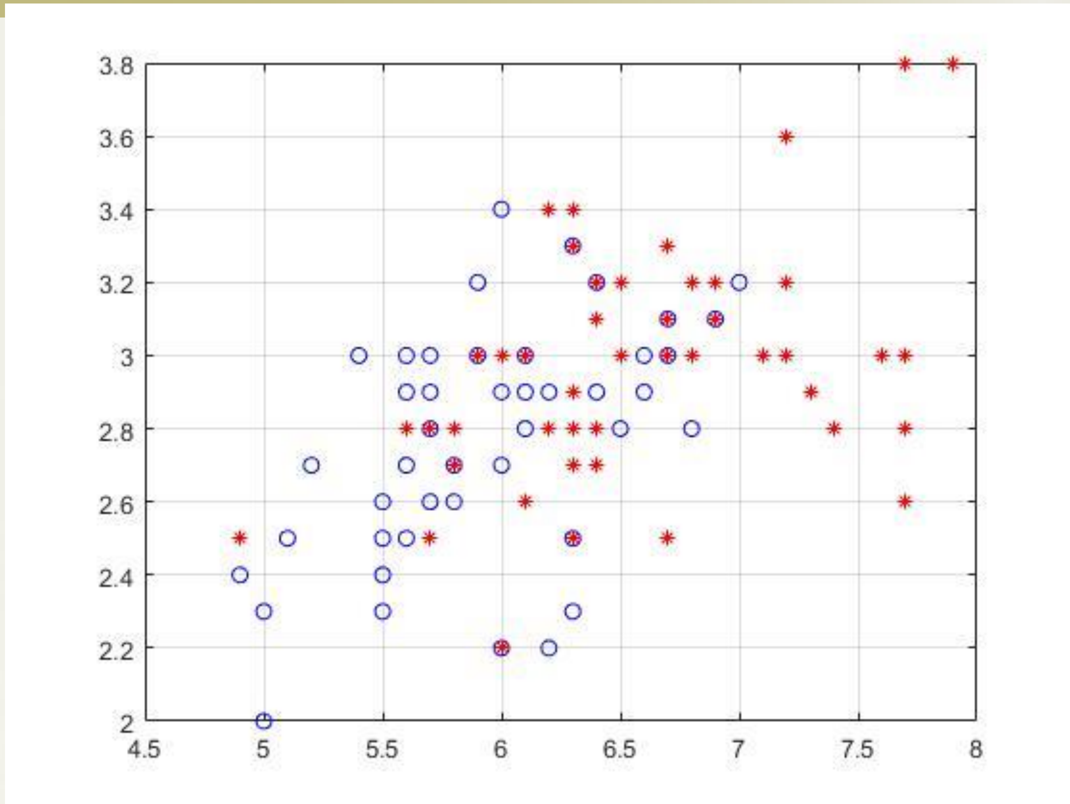
模式识别的分类问题是根据识别对象特征的观察值，并常以空间划分的方法将其分到某个类别中去。（傅京孙）

线性可分



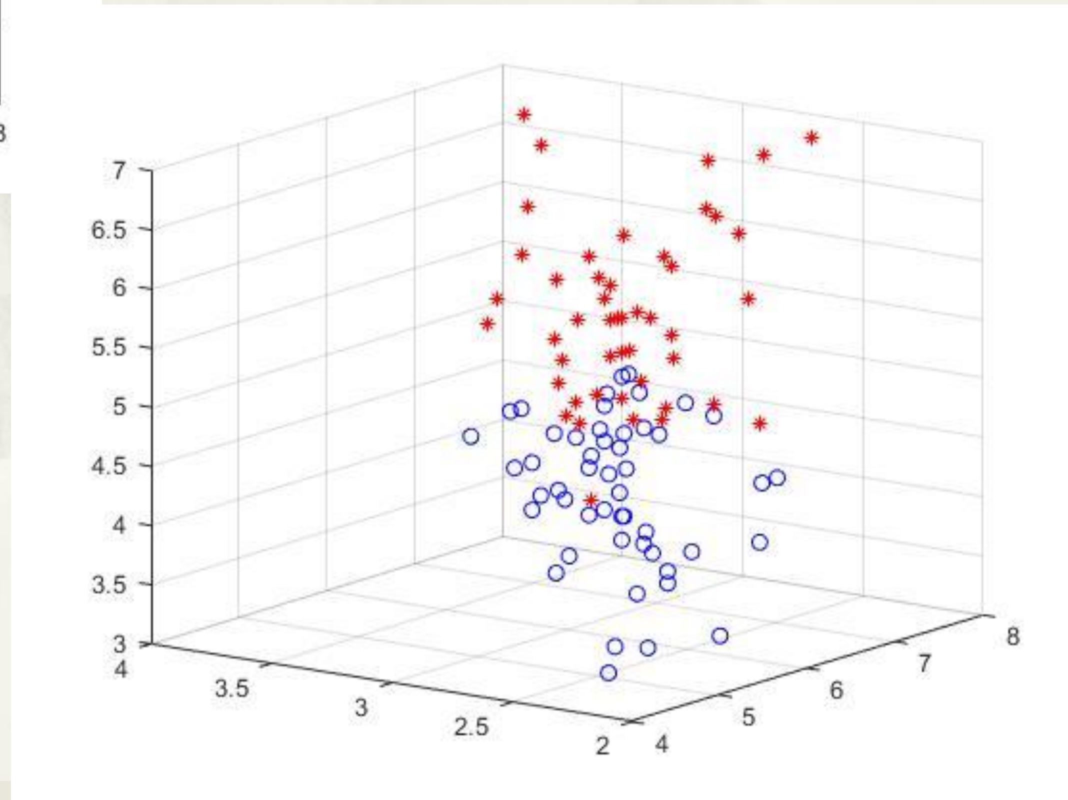
寻找到一个判别函数 $f(x)$，使得对于“+”类样本的值代入$f(x)$都有$f(x)>0$，“o”类样本值代入$f(x)$都有$f(x)<0$。或者，

$$\begin{cases} if\ f(x) > 0, then\ x \in \omega_+ \\ if\ f(x) < 0, then\ x \in \omega_o \end{cases}$$



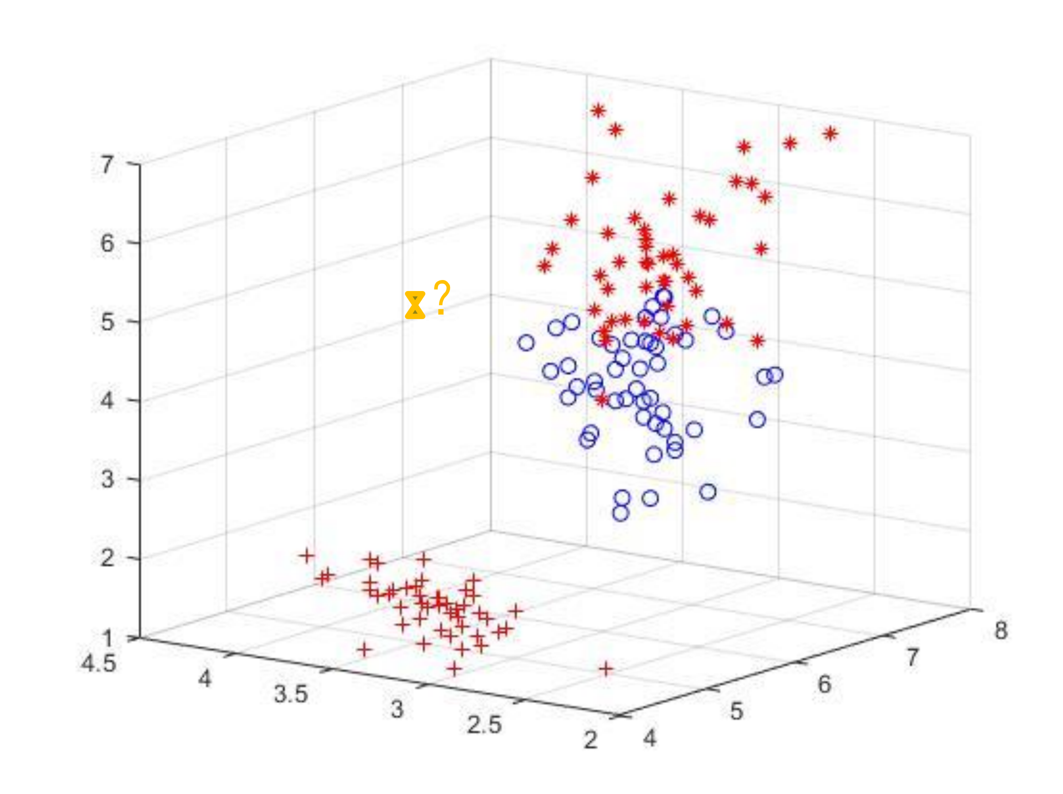
取iris数据的第1维和第2维特征，观察第2类和第3类样本在空间的分布。

在2维空间中，第2类和第3类样本有重叠，不可分。



取iris数据的第1，2，3维特征，观察第2类和第3类样本在空间的分布。

在3维空间中，第2类和第3类样本线性不可分（非线性）。



取iris数据的第1，2，3维特征，观察3类样本在3维空间的分布。

讨论：
1）对于分类问题，机器学习的目标是找到各类样本之间的分解面？
2）特征映射和特征空间的维度。
3）样本的空间分布：几何？概率？距离？
4）分类和预测，以及正确率（or错误率）。

设想：如果我们利用已知样本（已经发生的事件/训练样本）估计出一种可用于判断的概率函数，那么，我们对于未发生事件（待识别样本）就可以做出一种概率意义上的判断（分类）。

要求：
（1）各类别总体的概率分布是已知的，即$P(\omega_+), P(\omega_-)$已知。
（2）要决策分类的类别数是一定的。

1.基于最小错误率的Bayes决策

- Bayes公式：

$$P(A|B)P(B) = P(B|A)P(A)$$

- 换符号：

$$A \to x; B \to \omega$$

$$P(x|\omega)P(\omega) = P(\omega|x)P(x)$$

- 对于二分类问题：

$$P(x|\omega_1)P(\omega_1) = P(\omega_1|x)P(x) \quad \text{和} \quad P(x|\omega_2)P(\omega_2) = P(\omega_2|x)P(x)$$

示例:

例：癌细胞识别。
假设要识别的细胞已经做过处理。抽取了d个表示细胞基本特性的特征，成为一个d维空间向量x $\left(\mathrm{x} \in R^d\right)$。识别的目的是要将x分类为正常细胞或异常细胞。令$\omega_1$表示正常，$\omega_2$表示异常。

根据医院细胞病理检查的大量统计资料，可以对某一地区正常细胞出现的比例做出估计。这相当于在识别前已知正常状态概率$\mathrm{P}(\omega_1)$和异常状态概率$\mathrm{P}(\omega_2)$ 。

$\mathrm{p}(x|\omega_1)$是正常状态下细胞特征（光密度）观察x的类条件概率密度；

$\mathrm{p}(x|\omega_2)$是异常状态下细胞特征（光密度）观察x的类条件概率密度。

那么，由Bayes公式

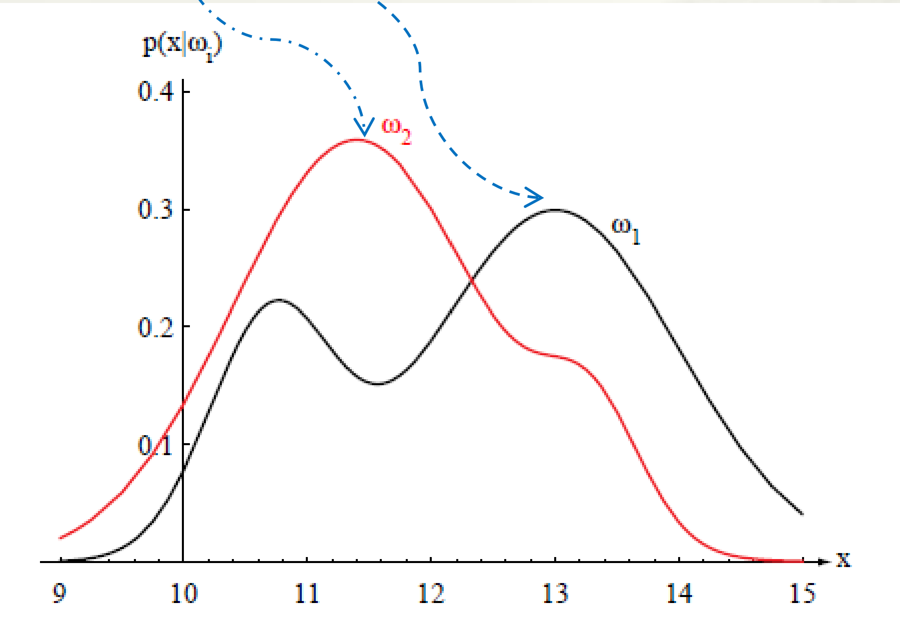
$$P(\omega_i|x) = \frac{p(x|\omega_i)P(\omega_i)}{\sum_{j=1}^{C} p\left(x\middle|\omega_j\right)P(\omega_j)}$$

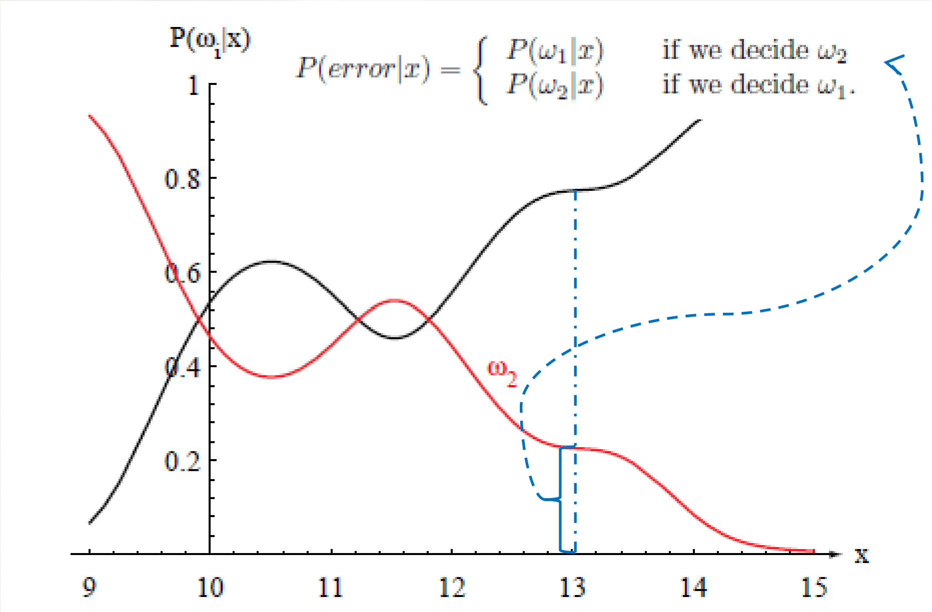
可得到条件概率$\mathrm{P}(\omega_i|x)$，即后验概率。

2.Bayes决策规则和错误率

可观察

若$\mathrm{P}(\omega_1|x) > \mathrm{P}(\omega_2|x)$，则将$x$归类为$\omega_1$，反之，把$x$归类于异常状态$\omega_2$。





If $P(\omega_i|x) = \max\limits_j P(\omega_j|x)$, then $x \in \omega_i$

If $p(x|\omega_i)P(\omega_i) = \max\limits_j p(x|\omega_j)P(\omega_j)$, then $x \in \omega_i$

If $l(x) = \dfrac{p(x|\omega_1)}{p(x|\omega_2)} \gtrless \dfrac{P(\omega_2)}{P(\omega_1)}$ then $x \in \begin{cases}\omega_1 \\ \omega_2\end{cases}$

If $h(x) = -ln[l(x)] = -lnp(x|\omega_1) + lnp(x|\omega_2)$

$\gtrless \ln\left(\dfrac{P(\omega_1)}{P(\omega_2)}\right)$, then $x \in \begin{cases}\omega_1 \\ \omega_2\end{cases}$

错误率

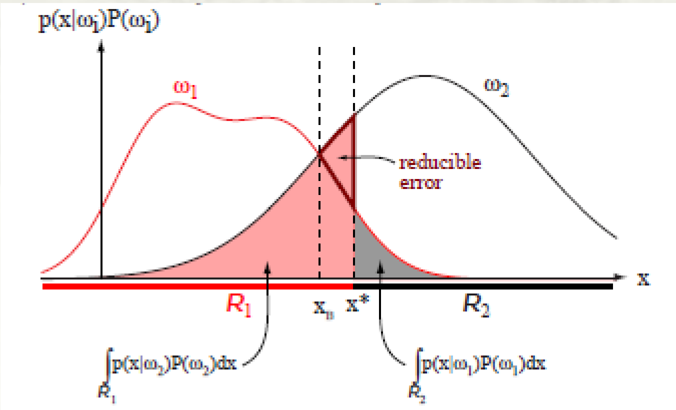
以1维情况来演示：按照这种Bayes决策规则进行分类可使错误率最小。

所谓错误率是指平均错误率，以P(e)表示，其定义为

$$P(e)=\int_{-\infty}^{+\infty}p(e,x)dx=\int_{-\infty}^{+\infty}p(e|x)p(x)dx$$

对于2类问题，由决策规则可知

$$p(e|x)=\begin{cases}p(\omega_1|x),\text{当}p(\omega_2|x)>p(\omega_1|x)\\ p(\omega_2|x),\text{当}p(\omega_1|x)>p(\omega_2|x)\end{cases}$$



令t为两类分界面。当特征x是一维时，t为x轴上一个点，且t将x轴分为2个区域R1和R2。于是

$$\mathrm{P}(e)=\int_{-\infty}^{t}p(\omega_2|x)p(x)dx+\int_{t}^{+\infty}p(\omega_1|x)p(x)dx$$

$$=\int_{-\infty}^{t}p(x|\omega_2)p(\omega_2)dx+\int_{t}^{+\infty}p(x|\omega_1)p(\omega_1)dx$$

$$=P(\omega_2)P_2(e)+P(\omega_1)P_1(e)$$

决策规则实际上是对每个x都使$p(e|x)$取最小，即$P(e)$最小。

3.基于最小风险的Bayes决策

使错误率$P(e)$达到最小是重要的，但实际上需要考虑一个比错误率更为广泛的概念——风险。而风险又是和损失函数紧密相连的。

在决策论中，称采取的决定为决策或行动。所有可能采取的决策组合的集合称决策空间或行动空间，以$\mathcal{A}$表示。

每个决策或行动都将带来一定的损失，它通常是决策和自然状态的函数。

状态 / 损失 / 决策	自然状态					
	ω_1	ω_2	…	ω_j	…	ω_c
α_1	$\lambda(\alpha_1,\omega_1)$	$\lambda(\alpha_1,\omega_2)$	…	$\lambda(\alpha_1,\omega_j)$	…	$\lambda(\alpha_1,\omega_c)$
α_2	$\lambda(\alpha_2,\omega_1)$	$\lambda(\alpha_2,\omega_2)$	…	$\lambda(\alpha_2,\omega_j)$	…	$\lambda(\alpha_2,\omega_c)$
…	…	…	…	…	…	…
α_i	$\lambda(\alpha_i,\omega_1)$	$\lambda(\alpha_i,\omega_2)$	…	$\lambda(\alpha_i,\omega_j)$	…	$\lambda(\alpha_i,\omega_c)$
…	…	…	…	…	…	…
α_a	$\lambda(\alpha_a,\omega_1)$	$\lambda(\alpha_a,\omega_2)$	…	$\lambda(\alpha_a,\omega_j)$	…	$\lambda(\alpha_a,\omega_c)$

最小风险Bayes决策规则

令期望风险为

$$\mathrm{R} = \int R(\alpha(x)|x)\, p(x)dx$$

我们希望：要求采取的一系列决策行动$\alpha(x)$使期望风险R最小。最小风险Bayes决策规则为

$$\text{如果 } R(\alpha_k|x) = \min_{i=1,2,\ldots,a} R(\alpha_i|x)\text{，则} \qquad \alpha = \alpha_k$$

对于实际问题，最小风险 Bayes决策可按下列步骤进行：

1) 根据$Bayes$公式计算出后验概率$\mathrm{P}(\omega_j|x)$
2) 利用后验概率及决策表，计算条件风险$R(\alpha_i|x) = \sum_{i=1}^{C} \lambda(\alpha_i|\omega_j)P(\omega_j|x)$
3) 对得到的a个条件风险值$R(\alpha_i|x)$进行比较，找出使条件风险最小的决策α_k，即

$$R(\alpha_k|x) = \min_{i=1,2,\ldots,a} R(\alpha_i|x)$$

则α_k就是最小风险Bayes决策。

4.参数估计决策

对于类条件概率密度函数$p(x|\omega_i)$，若已经知道，则直接采用Bayes决策规则；若不知道，但知道$p(x|\omega_i)$的分布形式，则可采用参数估计的方法。

在连续概率密度函数中，研究较透彻的是多元正态分布。其原因是

1）物理上的合理性；

2）数学上比较简便。

单变量正态分布概率密度函数：

$$p(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

多元正态分布的概率密度函数：

$$p(X)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}exp\left\{-\frac{1}{2}(X-\vec{\mu})^T\Sigma^{-1}(X-\vec{\mu})\right\}$$

Σ是$d\times d$维协方差矩阵，$|\Sigma|$是Σ的行列式，T表示转置，$\vec{\mu}=\mathrm{E}\{X\}$，

$\Sigma=\mathrm{E}\{（X-\vec{\mu}）(X-\vec{\mu})^T\}$

多元正态概率型下最小错误率Bayes判别函数和决策面

似然函数的对数：

$$\boldsymbol{g}_i(X) = -\frac{1}{2}(X - U_i)^T \boldsymbol{\Sigma}_i{}^{-1}(X - U_i) - \frac{d}{2}\ln(2\pi) - \frac{1}{2}ln|\boldsymbol{\Sigma_i}| + \ln[P(\omega_i)]$$

决策面方程：

$$\boldsymbol{g}_i(X) = \boldsymbol{g}_j(X)$$

我们分三种情况予以讨论：

① $\boldsymbol{\Sigma}_i = \sigma^2 I, \qquad i = 1,2, \dots, C$

② $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$

③ 各类协方差阵不相等。

第一种情况：$\boldsymbol{\Sigma}_i = \sigma^2 I,\ i = 1,2,\dots,C$

即，每类的协方差矩阵都相等，而且类内各特征相互独立，具有相等的方差σ^2。

1）若先验概率$P(\omega_i) \neq P(\omega_j)$，此时 $\boldsymbol{\Sigma}_i = \begin{pmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{pmatrix}$; $|\boldsymbol{\Sigma}_i| = \sigma^{2d}$; $\boldsymbol{\Sigma}_i^{-1} = \frac{1}{\sigma^2}\boldsymbol{I}$

$$\boldsymbol{g}_i(X) = -\frac{1}{2\sigma^2}(X-U_i)^T(X-U_i) \boxed{-\frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma^{2d}} + \ln[P(\omega_i)]$$

与类别无关

$$\boldsymbol{g}_i(X) = -\frac{1}{2\sigma^2}\boxed{(X-U_i)^T(X-U_i)} + \ln[P(\omega_i)]$$

$$\boxed{(X-U_i)^T(X-U_i)} = \|X-U_i\|^2 = \boxed{\sum_{j=1}^{d}(x_j-\mu_{ij})^2} \qquad i = 1,2,\dots,C$$

欧氏距离平方

为X的二次函数，但X^TX与i无关，可忽略。故判决函数为线性函数。

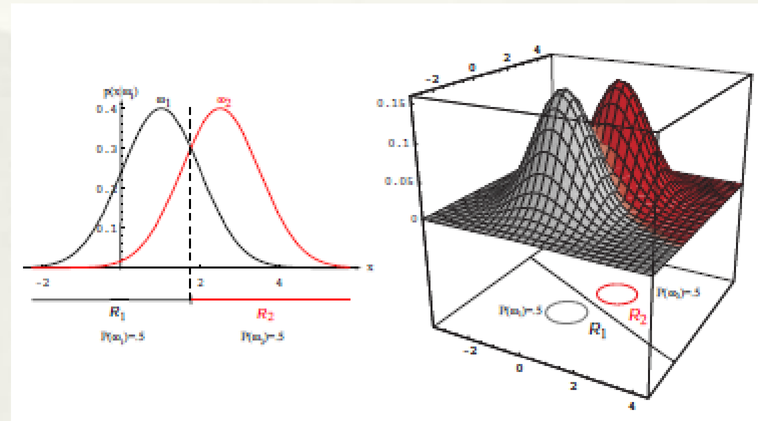
2）若先验概率$P(\omega_i) = P(\omega_j)$，此时

$$\boldsymbol{g}_i(X) = -\frac{1}{2\sigma^2}(X-U_i)^T(X-U_i)$$

只需要计算$\|X-U_i\|^2$，取min即可。称为最小距离分类器。

第二种情况： $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$ 即，各类协方差矩阵都相等，$\boldsymbol{\Sigma}$与i无关。

此时，$\boldsymbol{g}_i(X)$可简化为：



$$\boldsymbol{g}_i(X) = -\frac{1}{2}(X - U_i)^T \boldsymbol{\Sigma}^{-1}(X - U_i) + \ln[P(\omega_i)]$$

若先验概率都相等，则为Mahalanobis距离平方。

忽略与i无关的$X^T\boldsymbol{\Sigma}^{-1}X$项，则判别函数可写成

$$\boldsymbol{g}_i(X) = \boldsymbol{W}_i{}^T X + \boldsymbol{W}_{i0}$$

其中， $\boldsymbol{W}_i = \boldsymbol{\Sigma}^{-1} U_i$

$$\boldsymbol{W}_{i0} = -\frac{1}{2} U_i{}^T \boldsymbol{\Sigma}^{-1} U_i + \ln[P(\omega_i)]$$

仍是X的线性函数。因此，决策面仍是一个超平面。

第三种情况：各类协方差阵不相等。

$$\begin{aligned}\boldsymbol{g}_i(X) &= -\frac{1}{2}(X-U_i)^T\boldsymbol{\Sigma}_i{}^{-1}(X-U_i)\boxed{-\frac{d}{2}\ln(2\pi)}-\frac{1}{2}ln|\boldsymbol{\Sigma_i}|+\ln[P(\omega_i)]\\ &= -\frac{1}{2}(X-U_i)^T\boldsymbol{\Sigma}_i{}^{-1}(X-U_i)-\frac{1}{2}ln|\boldsymbol{\Sigma_i}|+\ln[P(\omega_i)]\\ &= X^T\boldsymbol{W}_iX+\boldsymbol{W}_{i1}{}^TX+\boldsymbol{W}_{i0}\end{aligned}$$

其中，

$$\boldsymbol{W}_i = -\frac{1}{2}\boldsymbol{\Sigma}_i{}^{-1}$$

$$\boldsymbol{W}_{i1} = \boldsymbol{\Sigma}_i{}^{-1}U_i$$

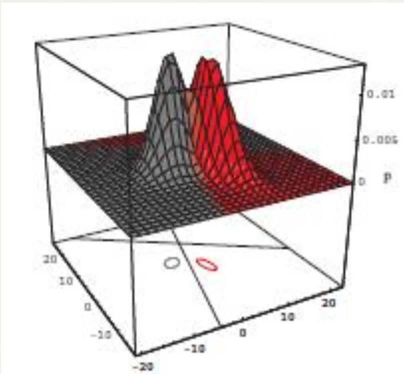
$$\boldsymbol{W}_{i0} = -\frac{1}{2}U_i{}^T\boldsymbol{\Sigma}_i{}^{-1}U_i-\frac{1}{2}ln|\boldsymbol{\Sigma}_i|+lnP(\omega_i)$$

此时，判决函数为X的二次型。若决策区域$\mathcal{R}_1$与$\mathcal{R}_2$相邻，则判决面应满足

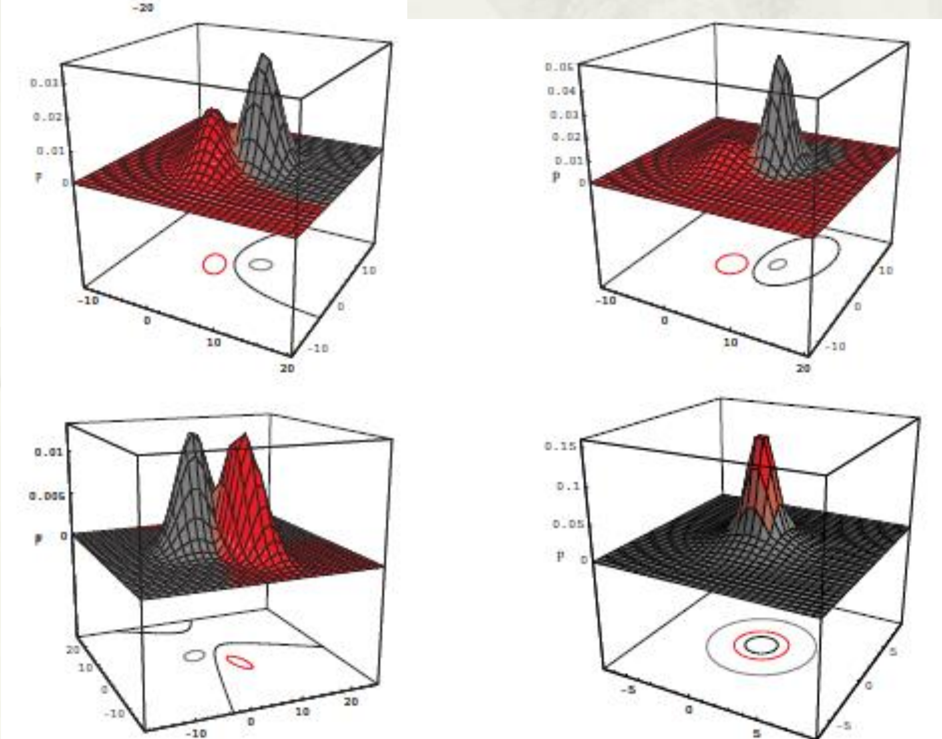
$$\boldsymbol{g}_i(X)-\boldsymbol{g}_j(X)=0$$

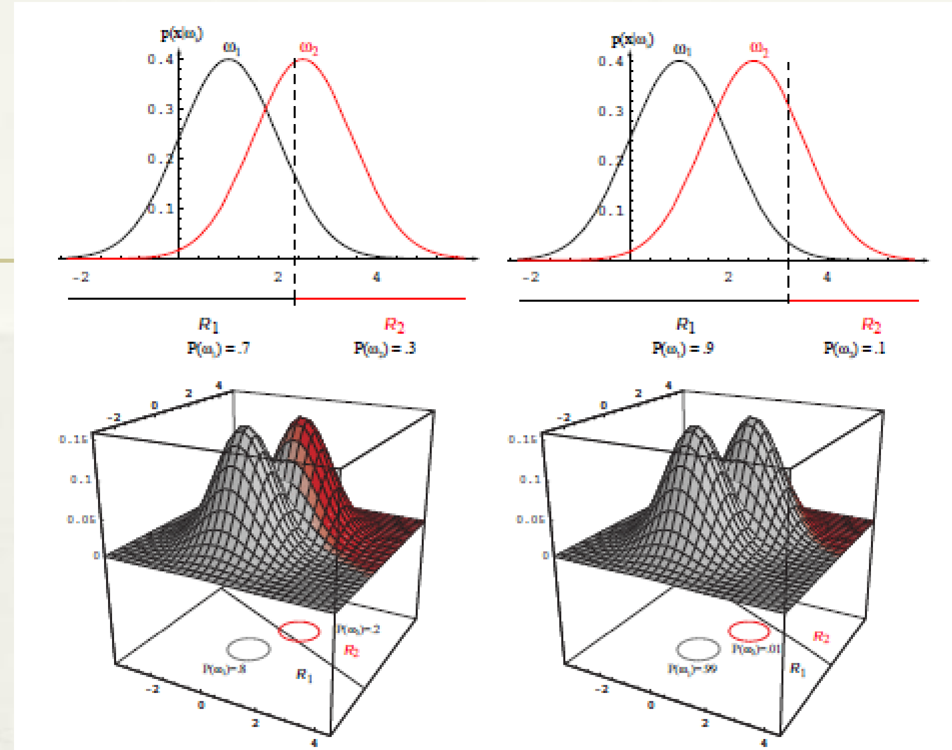
即，$X^T(\boldsymbol{W}_i - W_j)X+\left(\boldsymbol{W}_{i1}-\boldsymbol{W}_{j1}\right)^TX+\boldsymbol{W}_{i0}-\boldsymbol{W}_{j0}=0$

为超二次曲面。随$\boldsymbol{\Sigma}_i, U_i, P(\omega_i)$不同而是超球面，超椭球面，超抛物面，超双曲面，或超平面。

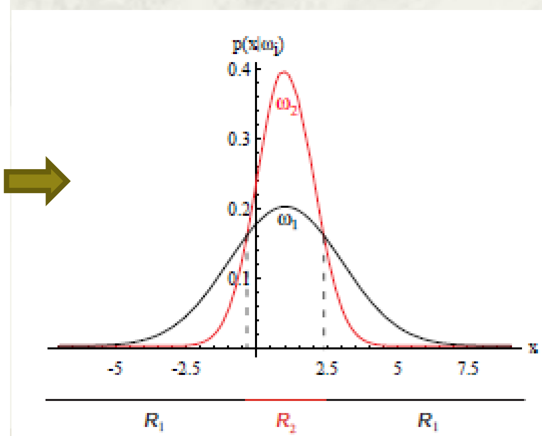


第三种情况下的边界示意





先验概率变化，决策边界会移动。



5.错误率估计和边界

Chernoff上界：

利用不等式：$\min[a,b] \le a^{\beta} b^{1-\beta}$

$$(a, b \ge 0, 且\ 0 \le \beta \le 1)$$

有

$$P(e) \le P^{\beta}(\omega_1)P^{1-\beta}(\omega_2)\int p^{\beta}(x|\omega_1)p^{1-\beta}(x|\omega_2)dx$$

解得

$$\int p^{\beta}(x|\omega_1)p^{1-\beta}(x|\omega_2)dx = e^{-k(\beta)}$$

正态分布时，

$$k(\beta) = \frac{\beta(1-\beta)}{2}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t[\beta\boldsymbol{\Sigma}_1 + (1-\beta)\boldsymbol{\Sigma}_2]^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) + \frac{1}{2}ln\frac{|\beta\Sigma_1 + (1-\beta)\Sigma_2|}{|\Sigma_1|^{\beta}|\Sigma_2|^{1-\beta}}$$

Bhattacharyya上界：

取$\beta = \frac{1}{2}$

$$P(e) \le \sqrt{P(\omega_1)P(\omega_2)}\int\sqrt{p(x|\omega_1)p(x|\omega_2)}\,dx$$

$$= \sqrt{P(\omega_1)P(\omega_2)}e^{-k(1/2)}$$

正态分布时，

$$k(1/2) = \frac{1}{8}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t\left[\frac{\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2}{2}\right]^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) + \frac{1}{2}ln\frac{\left|\frac{\Sigma_1 + \Sigma_2}{2}\right|}{\sqrt{|\Sigma_1||\Sigma_2|}}$$

注：对于偏离高斯分布太远的分布，这些上界不能说明什么问题。