

模式识别

第6讲 统计决策分类法 (II) (概率密度估计)

2018~2019学年



内容安排

一、绪论、数学基础（第1讲）

二、聚类分析（第2讲）

三、判别函数分类法（几何分类法）（第3、4讲）

四、统计决策分类法（概率分类法）（第5、6讲）

五、特征提取与选择（第7讲）

六、模糊模式识别（第8讲）

七、神经网络模式识别（第9讲）

期末考试（平时作业：40%，期末考试：60%）

四、统计决策分类法

- 4.1 研究对象及相关概率
- 4.2 贝叶斯决策
- 4.3 贝叶斯分类器的错误率
- 4.4 耐曼-皮尔逊决策
- 4.5 概率密度函数的参数估计
- 4.6 概率密度函数的非参数估计
- 4.7 后验概率密度分类的势函数方法

4.4 耐曼-皮尔逊 (Neyman-Pearson) 决策

1. 基本思想

在某些实际问题中，可能存在以下几种情况：

- (1) 不知道各类的先验概率 $P(\omega_i)$;
- (2) 难于确定误判损失 $L_{ij}(X)$;
- (3) 某一类错误较另一种错误更为重要。

针对(1)，可以采用最小最大风险准则或令各类先验概率相等的办法克服；

针对(2)，如果允许的话，可以不使用损失函数而采用最小化最大误判概率准则(Minimax)；

针对(3)，可以采用最小后验风险准则判决。

主要针对 (3)，我们采用N-P准则。

基本思想： 限制某一类错误率为一个确定值，求取使得另一类错误率最小的判决规则与判决阈值（二类问题）。

在两类问题贝叶斯决策的错误率公式中：

$$\begin{aligned} P(e) &\triangleq \Pr(\mathbf{X} \in R_2, \mathbf{X} \in \omega_1) + \Pr(\mathbf{X} \in R_1, \mathbf{X} \in \omega_2) \\ &= P(\mathbf{X} \in R_2 | \omega_1)P(\mathbf{X} \in \omega_1) + P(\mathbf{X} \in R_1 | \omega_2)P(\mathbf{X} \in \omega_2) \\ &= P_1(e)P(\omega_1) + P_2(e)P(\omega_2) \end{aligned}$$

式中，

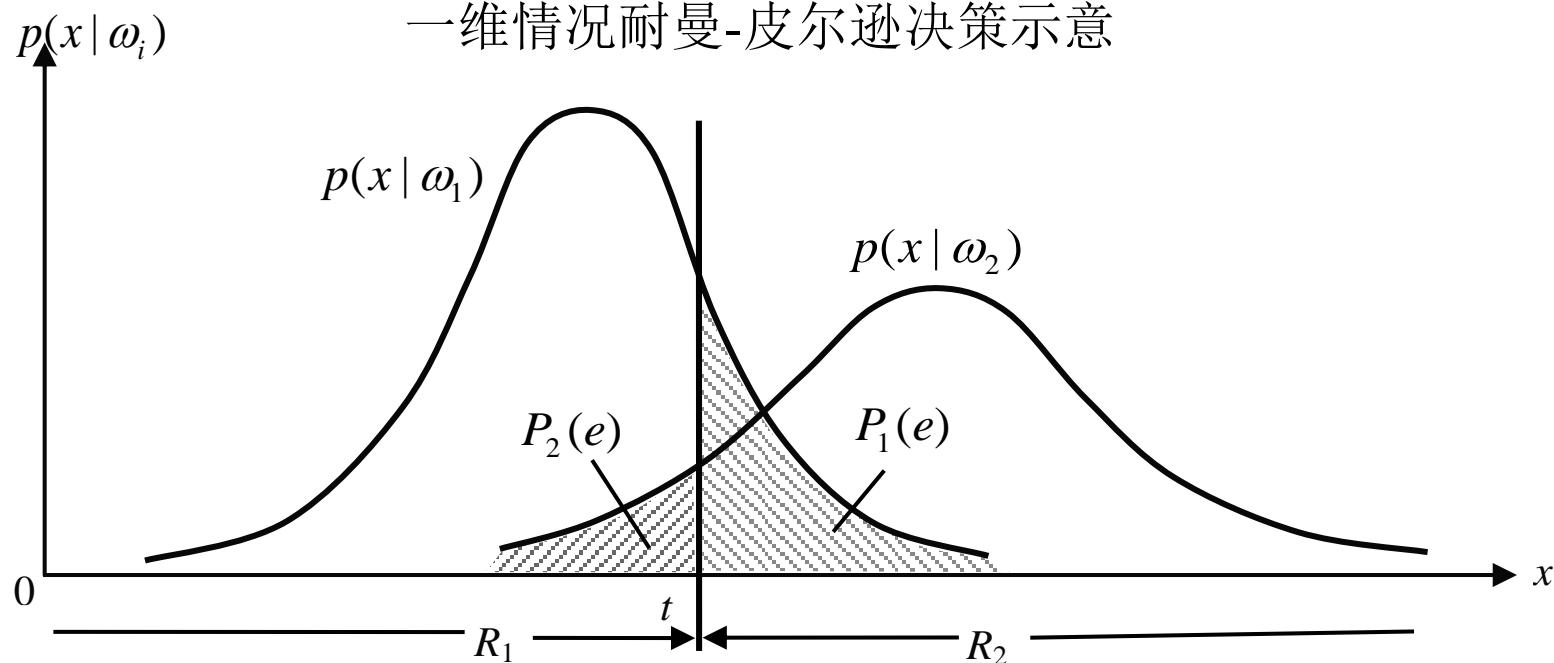
$$P_1(e) \triangleq \int_{R_2} p(\mathbf{X} | \omega_1) d\mathbf{X}$$
$$P_2(e) \triangleq \int_{R_1} p(\mathbf{X} | \omega_2) d\mathbf{X}$$

先验概率通常为常数，故一般也称 $P_1(e)$ 和 $P_2(e)$ 为两类错误率：

$P_1(e)$ ： ω_1 类模式被误判为 ω_2 类的错误率；

$P_2(e)$ ： ω_2 类模式被误判为 ω_1 类的错误率。

一维情况耐曼-皮尔逊决策示意



耐曼-皮尔逊决策出发点：在 $P_2(e)$ 等于常数的条件下，使 $P_1(e)$ 为最小，以此确定判决阈值 t 。

例：在信号检测中， ω_1 代表“有警情”， ω_2 代表“无警情”。

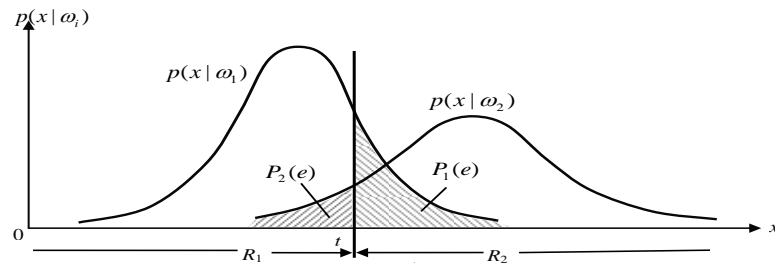
因此： $P_2(e)$ 代表虚警概率；

$P_1(e)$ 代表漏报概率 $= 1 - P_D$ (P_D ：检测概率)

此时耐曼-皮尔逊决策含义：在虚警概率 $P_2(e)$ 是一个可以承受的常数值的情况下，使漏报概率为最小。

2. 判别式推导

求解问题：



在 $P_2(e)$ 等于常数的条件下，求 $P_1(e)$ 极小值的条件极值问题。
 $P_2(e)$ 的值一般很小。

构造辅助函数 $Q = P_1(e) + \mu P_2(e)$

式中： μ ——待定常数； $P_2(e)$ ——常数。

求判决规则和阈值使 $P_1(e)$ 或 Q 最小。



$$P_1(e) = \int_{R_2} p(\mathbf{X} | \omega_1) d\mathbf{X}$$

$$P_2(e) = \int_{R_1} p(\mathbf{X} | \omega_2) d\mathbf{X}$$

$$\begin{aligned} Q &= \int_{R_2} p(\mathbf{X} | \omega_1) d\mathbf{X} + \mu \int_{R_1} p(\mathbf{X} | \omega_2) d\mathbf{X} \\ &= \left(1 - \int_{R_1} p(\mathbf{X} | \omega_1) d\mathbf{X} \right) + \mu \int_{R_1} p(\mathbf{X} | \omega_2) d\mathbf{X} \\ &= 1 + \int_{R_1} [\mu p(\mathbf{X} | \omega_2) - p(\mathbf{X} | \omega_1)] d\mathbf{X} \end{aligned} \quad (4-57)$$

要使 Q 最小，积分项至少应为负值，即在 R_1 区域内，至少应保证

$$\mu p(\mathbf{X} | \omega_2) < p(\mathbf{X} | \omega_1)$$

$$\mu p(\mathbf{X} | \omega_2) < p(\mathbf{X} | \omega_1)$$

即

$$\frac{p(\mathbf{X} | \omega_1)}{p(\mathbf{X} | \omega_2)} > \mu \Rightarrow \mathbf{X} \in \omega_1 \quad (4-58)$$

同理由式(4-57) 有:

$$\begin{aligned} Q &= \int_{R_2} p(\mathbf{X} | \omega_1) d\mathbf{X} + \mu \left(1 - \int_{R_2} p(\mathbf{X} | \omega_2) d\mathbf{X} \right) \\ &= \mu + \int_{R_2} [p(\mathbf{X} | \omega_1) - \mu p(\mathbf{X} | \omega_2)] d\mathbf{X} \end{aligned}$$

在 R_2 区域内至少应保证

$$p(\mathbf{X} | \omega_1) < \mu p(\mathbf{X} | \omega_2)$$

即

$$\frac{p(\mathbf{X} | \omega_1)}{p(\mathbf{X} | \omega_2)} < \mu \Rightarrow \mathbf{X} \in \omega_2 \quad (4-59)$$

得决策规则: 若 $\frac{p(\mathbf{X} | \omega_1)}{p(\mathbf{X} | \omega_2)} \gtrless \mu$, 则 $\mathbf{X} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$



$$Q = \int_{R_2} p(\mathbf{X} | \omega_1) d\mathbf{X} + \mu \int_{R_1} p(\mathbf{X} | \omega_2) d\mathbf{X} \quad (4-57)$$

$$\text{若 } \frac{p(\mathbf{X} | \omega_1)}{p(\mathbf{X} | \omega_2)} \gtrless \mu, \text{ 则 } \mathbf{X} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

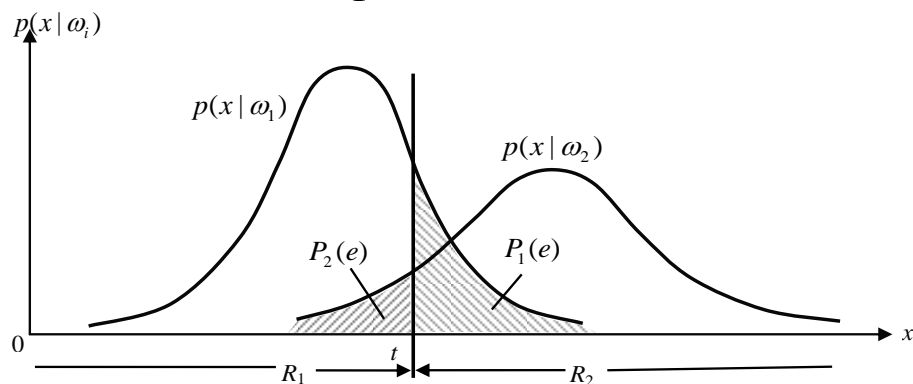
当 $\frac{p(\mathbf{X} | \omega_1)}{p(\mathbf{X} | \omega_2)} = \mu$ 时，此时的 \mathbf{X} 为模式的判决阈值 t ，

是似然比的判决阈值 μ 的函数 $\mathbf{X} = t(\mu)$ 。

由于 $p(\mathbf{X} | \omega_1)$ 和 $p(\mathbf{X} | \omega_2)$ 是已知的，所以耐曼-皮尔逊决策最终归结为寻找似然比阈值 μ 。

求解 μ 值从常数 $P_2(e)$ 入手，这时由 $P_2(e) = \int_{R_1} p(\mathbf{X} | \omega_2) d\mathbf{X}$ 有

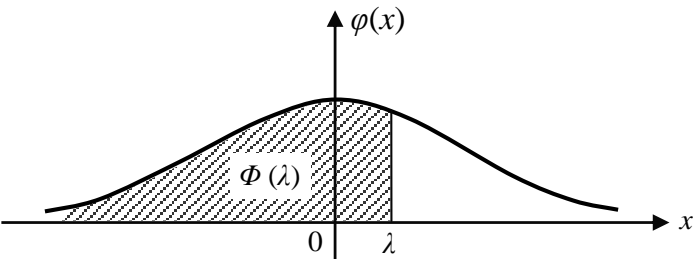
$$P_2(e) = \int_{-\infty}^{t(\mu)} p(\mathbf{X} | \omega_2) d\mathbf{X}$$



即 μ 是 $P_2(e)$ 的函数，通过查标准正态分布表可以求得 μ 的值。

复习

1. 标准正态分布表



$$\Phi(\lambda) = \int_{-\infty}^{\lambda} \varphi(x) dx$$

① 表中末行系函数值：
 $\Phi(3.0), (3.1), \dots, \Phi(3.9)$.

② 表中为 $\lambda \geq 0$ 时， $\Phi(\lambda)$ 的值。 $\lambda \leq 0$ ：使用公式。

③ 纵向值： λ 的整数部分和小数点后第一位。

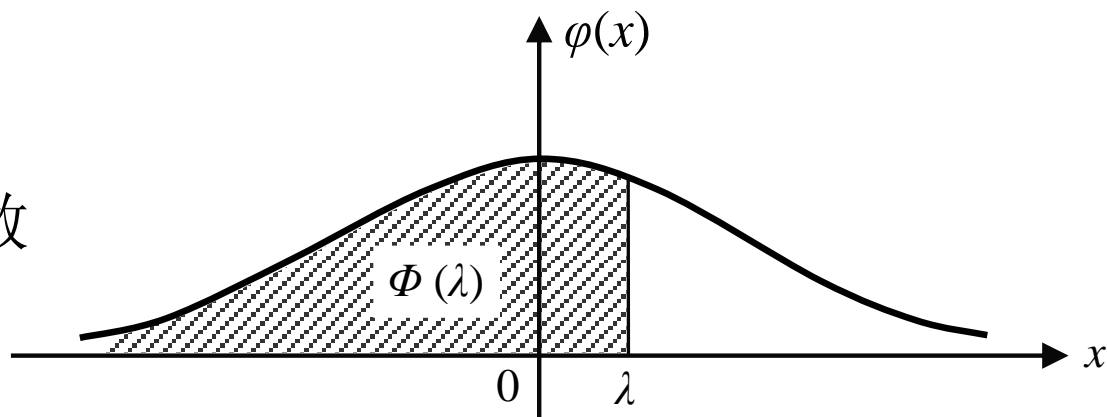
④ 横向值： λ 的小数点后第二位。

λ	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.5026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.6054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7039	0.7967	0.7995	0.7023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8689	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.5406	0.9418	0.9430	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9159	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9648	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9700	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9939	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

2. 正态分布的概率计算

① 左边阴影部分的面积表示为概率。即分布函数

$$\Phi(\lambda) = \int_{-\infty}^{\lambda} \varphi(x) dx$$



② 当 $\lambda < 0$ 时，可以使用 $\Phi(-\lambda) = 1 - \Phi(\lambda)$ ；

③ $\Phi(\lambda) = \int_{-\infty}^{\infty} \varphi(x) dx = 1$

④ 在任一区间 (λ_1, λ_2) 内取值的概率：

$$P(\lambda_1 < \lambda < \lambda_2) = \int_{\lambda_1}^{\lambda_2} \varphi(x) dx = \Phi(\lambda_2) - \Phi(\lambda_1)$$

例 利用标准正态分布表，求标准正态分布在下面区间内取值的概率。(1) $(-0.5, 1.5)$; (2) $(-1.96, 1.96)$; (3) $(-3, 3)$

解： (1) $P(-0.5 < x < 1.5) = \Phi(1.5) - \Phi(-0.5)$
 $= \Phi(1.5) - [1 - \Phi(0.5)]$
 $= 0.9332 - [1 - 0.6915] = 0.6274$

(2) $P(-1.96 < x < 1.96) = \Phi(1.96) - [1 - \Phi(1.96)]$
 $= 2 \times 0.9750 - 1 = 0.9500$

(3) $P(-3 < x < 3) = \Phi(3) - [1 - \Phi(3)]$
 $= 2\Phi(3) - 1 = 2 \times 0.9987 - 1 = 0.9974$

例4.4 一两类问题，模式分布为二维正态，其分布参数

$$\mathbf{M}_1 = [-1, 0]^T \quad \mathbf{M}_2 = [1, 0]^T$$

协方差矩阵为 $\mathbf{C}_1 = \mathbf{C}_2 = \mathbf{I}$ ，设 $P_2(e) = 0.046$ ，求耐曼-皮尔逊决策规则下，似然比的判决阈值 μ ，以及模式的判别阈值（界面）。

解：(1) 求类概率密度函数

正态分布的类概率密度函数为

$$p(\mathbf{X} | \omega_i) = \frac{1}{(2\pi)^{n/2} |\mathbf{C}_i|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{X} - \mathbf{M}_i)^T \mathbf{C}_i^{-1} (\mathbf{X} - \mathbf{M}_i) \right\} \\ i=1, 2$$

已知 $\mathbf{M}_1 = [-1, 0]^T$, $\mathbf{M}_2 = [1, 0]^T$, 又计算得:

$$|\mathbf{C}_i|^{\frac{1}{2}} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}^{\frac{1}{2}} = 1 \quad \mathbf{C}_i^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}$$

$$\therefore p(\mathbf{X} | \omega_1) = \frac{1}{2\pi} \exp \left\{ -\frac{(\mathbf{X} - \mathbf{M}_1)^T (\mathbf{X} - \mathbf{M}_1)}{2} \right\} = \frac{1}{2\pi} \exp \left\{ -\frac{(x_1 + 1)^2 + x_2^2}{2} \right\}$$

$$p(\mathbf{X} | \omega_2) = \frac{1}{2\pi} \exp \left\{ -\frac{(\mathbf{X} - \mathbf{M}_2)^T (\mathbf{X} - \mathbf{M}_2)}{2} \right\} = \frac{1}{2\pi} \exp \left\{ -\frac{(x_1 - 1)^2 + x_2^2}{2} \right\}$$

$$p(\mathbf{X} | \omega_1) = \frac{1}{2\pi} \exp \left\{ -\frac{(\mathbf{X} - \mathbf{M}_1)^T (\mathbf{X} - \mathbf{M}_1)}{2} \right\} = \frac{1}{2\pi} \exp \left\{ -\frac{(x_1 + 1)^2 + x_2^2}{2} \right\}$$

$$p(\mathbf{X} | \omega_2) = \frac{1}{2\pi} \exp \left\{ -\frac{(\mathbf{X} - \mathbf{M}_2)^T (\mathbf{X} - \mathbf{M}_2)}{2} \right\} = \frac{1}{2\pi} \exp \left\{ -\frac{(x_1 - 1)^2 + x_2^2}{2} \right\}$$

(2) 求似然比

$$\frac{p(\mathbf{X} | \omega_1)}{p(\mathbf{X} | \omega_2)} = \exp \left\{ -\frac{1}{2} (x_1^2 + 2x_1 + 1) + \frac{1}{2} (x_1^2 - 2x_1 + 1) \right\} = \exp \{-2x_1\}$$

(3) 求基于模式的判别式

决策规则： 若 $\exp\{-2x_1\} \gtrless \mu$ ，则 $X \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$

两边取自然对数，有

$$-2x_1 \gtrless \ln \mu$$

得判别式

$$\text{若 } x_1 \gtrless -\frac{1}{2} \ln \mu, \text{ 则 } X \in \begin{cases} \omega_1 \\ \omega_2 \end{cases} \quad (4-62)$$

书中公式
方向错了！

(4) 求似然比判决阈值 μ

由 $P_2(e)$ 与 μ 的关系有

$$\begin{aligned} P_2(e) &= \int_{R_1} p(\mathbf{X} | \omega_2) d\mathbf{X} \\ &= \int_{-\infty}^{T_1} \int_{-\infty}^{T_2} \frac{1}{2\pi} \exp \left\{ -\frac{(x_1 - 1)^2 + x_2^2}{2} \right\} dx_2 dx_1 \end{aligned}$$

分离积分，向正态分布表的标准形式

$$\Phi(\lambda) = \int_{-\infty}^{\lambda} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad (\lambda \geq 0)$$

变换，有（ x_2 没有限制，故 T_2 为正无穷）

$$P_2(e) = \int_{-\infty}^{-\frac{1}{2} \ln \mu} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{(x_1 - 1)^2}{2} \right\} dx_1 \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{x_2^2}{2} \right\} dx_2$$

$$\text{令 } x_1 - 1 = y \text{ 有: } P_2(e) = \int_{-\infty}^{-\frac{1}{2} \ln \mu - 1} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{y^2}{2} \right\} dy$$

$$\Phi(\lambda') = 1 - \Phi(-\lambda')$$

查正态分布数值表，要求 $P_2(e)=0.046$ 。记 $\lambda = -\lambda'$

$$\therefore \lambda' = -\frac{1}{2} \ln \mu - 1 < 0 \quad \therefore \text{在表上查 } \Phi(\lambda) = 1 - 0.046 = 0.954 \text{。}$$

x	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9648	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9700	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767

对应 $\lambda = ?$ 对应 $\lambda = 1.69$ ，即 $\lambda' = -1.69$

$$\text{有 } -\frac{1}{2} \ln \mu - 1 = -1.69 \Rightarrow -\frac{1}{2} \ln \mu = -0.69$$

计算得 $\mu = e^{1.38} = 3.98$ （这一步其实没必要计算）

$$\mu = e^{1.38} = 3.98$$



$$\text{若 } x_1 \lesseqgtr -\frac{1}{2} \ln \mu, \text{ 则 } \mathbf{X} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases} \quad (4-62)$$

由 (4-62) 式得判别界面: $x_1 = -\frac{1}{2} \ln \mu = -0.69$

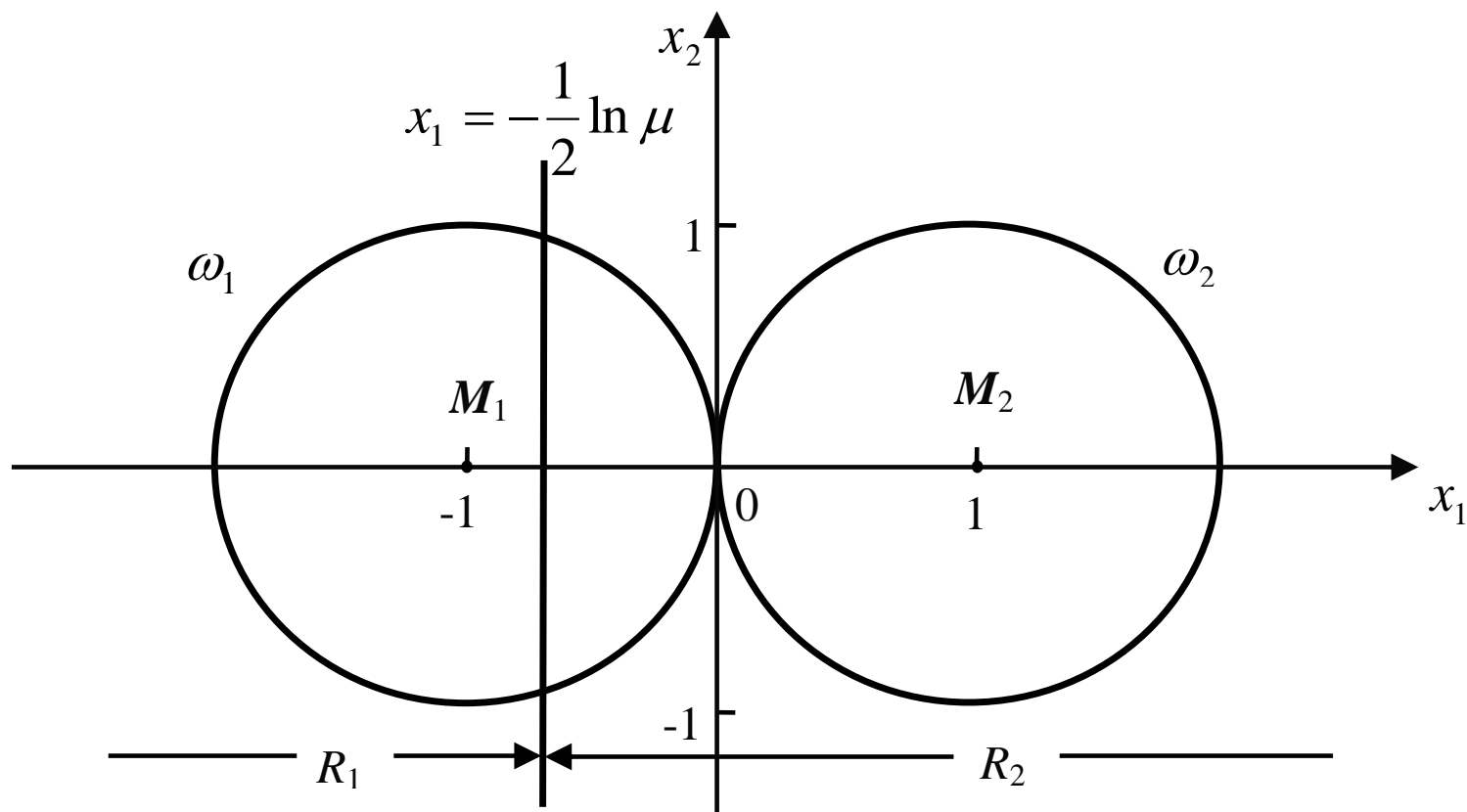


图4.12 耐曼-皮尔逊决策结果

总结分析：研究算法的三种思路

① 使总错误率最小：→ 最小错误率Bayes决策

② 使风险（错误引起的平均损失）最小：

{ → 最小后验风险Bayes决策
→ (0-1)损失最小风险Bayes决策

③ 限制一类错误概率，追求另一类错误率最小：
→ Neyman-Pearson决策

4.5 概率密度函数的参数估计方法

在前面推导分类判决公式中, 我们一直假设类模式的概率密度函数 $p(X|\omega_i)$ 是已知的, 然后去设计贝叶斯分类器。但在实际中, 这些知识往往是不知道的, 这就需要用已知的样本进行学习或训练。也就是说利用统计推断理论中的估计方法, 从样本集数据中估计这些密度。

以下讨论:

已知类别的样本 (训练样本) →

学习或训练 →

获得类模式的概密 $p(X|\omega_i)$

概率密度的两类估计方法：

* 参数估计方法：

已知概率密度函数的形式而函数的有关参数未知，通过估计参数来估计概率密度函数的方法。

* 非参数估计方法：

概率密度函数的形式未知，直接估计概率密度函数的方法。

两种主要参数估计法：

最大似然估计、贝叶斯估计。

4.5.1 最大似然估计

设： ω_i 类的类模式的概率密度函数具有某种确定的函数形式； θ 是该函数的一个未知参数或参数集。

最大似然估计把 θ 当作确定的未知量进行估计。

1. 似然函数

从 ω_i 类中独立地抽取 N 个样本： $X^N = \{X_1, X_2, \dots, X_N\}$

称这 N 个样本的联合概率密度函数 $p(X^N | \theta)$ 为**相对于样本集 X^N 的 θ 的似然函数**。

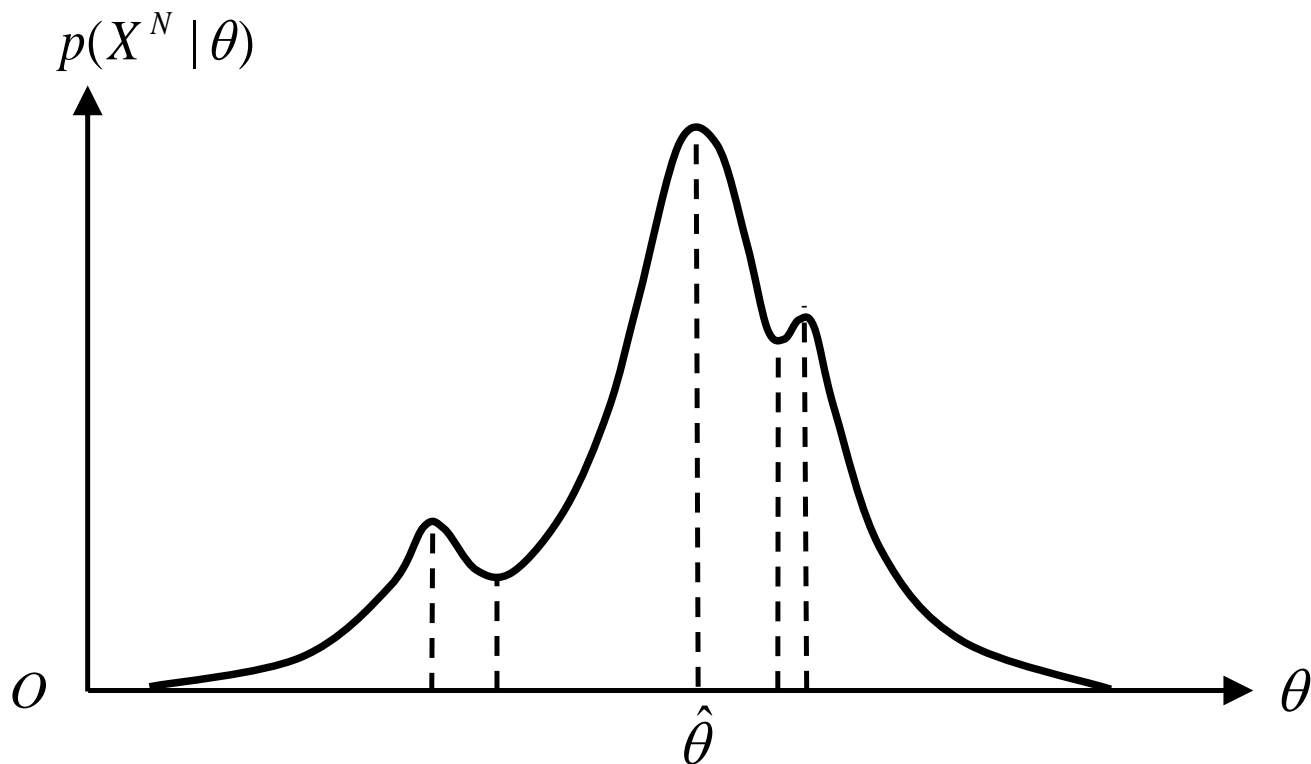
$$h(\theta | X^N) \equiv p(X^N | \theta) = p(X_1, X_2, \dots, X_N | \theta) = \prod_{k=1}^N p(X_k | \theta)$$

——在参数 θ 下观测到的样本集 X^N 的概率(联合分布)密度

2. 最大似然估计

不同参数值对应不同的密度函数，相应地，产生这 N 个样本的概率值也各不相同，因此**最可能产生这些样本的pdf参数值**，就应该选使得**概率值最大的那一个**！——不证自明的**似然原理**。

找到这样的 θ ，它使似然函数 $h(\theta / X^N) \equiv p(X^N | \theta)$ 极大。



θ 为一维时的最大似然估计示意图

θ 的**最大似然估计量 $\hat{\theta}$** 就是**使似然函数达到最大值的估计量**。

由 $\frac{dp(X^N | \theta)}{d\theta} = 0$ 求得。

为便于分析，定义似然函数的对数（对数似然函数）为

$$H(\boldsymbol{\theta}) = \ln p(X^N | \boldsymbol{\theta})$$

$\boldsymbol{\theta}$ 的最大似然估计是下面微分方程的解：

$$\frac{dH(\boldsymbol{\theta})}{d\boldsymbol{\theta}} = 0$$

设 ω_i 类模式的概率密度函数有 p 个未知参数，记为 p 维向量

$$\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_p]^T$$

此时

$$H(\boldsymbol{\theta}) = \ln p(X^N | \boldsymbol{\theta}) = \sum_{k=1}^N \ln p(X_k | \boldsymbol{\theta})$$

$$\frac{\partial}{\partial \boldsymbol{\theta}} \left[\sum_{k=1}^N \ln p(X_k | \boldsymbol{\theta}) \right] = 0$$

$$\begin{cases} \sum_{k=1}^N \frac{\partial}{\partial \theta_1} \ln p(X_k | \boldsymbol{\theta}) = 0 \\ \sum_{k=1}^N \frac{\partial}{\partial \theta_2} \ln p(X_k | \boldsymbol{\theta}) = 0 \\ \vdots \\ \sum_{k=1}^N \frac{\partial}{\partial \theta_p} \ln p(X_k | \boldsymbol{\theta}) = 0 \end{cases} \quad (4-69)$$

解以上微分方程即可得到 $\boldsymbol{\theta}$ 的最大似然估计值。

3. 正态分布情况举例

设 ω_i 类：正态分布、**一维模式**、概率密度函数为

$$p(\mathbf{X} | \omega_i) \sim N(\mu, \sigma^2)$$

待估计参数为 μ, σ^2 。

$p(\mathbf{X} | \omega_i)$ 可表示为 $p(\mathbf{X} | \boldsymbol{\theta}) \sim N(\mu, \sigma^2)$ 。

其中， $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$ ， $\theta_1 = \mu$ ， $\theta_2 = \sigma^2$ 。

若 \mathbf{X}^N 表示从 ω_i 中独立抽取的 N 个样本，则 $\boldsymbol{\theta}$ 的似然函数为

$$p(\mathbf{X}^N | \boldsymbol{\theta}) = \prod_{k=1}^N p(\mathbf{X}_k | \boldsymbol{\theta})$$

其中， $p(\mathbf{X}_k | \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(\mathbf{X}_k - \mu)^2}{2\sigma^2}\right]$

$$\ln p(\mathbf{X}_k | \boldsymbol{\theta}) = -\frac{1}{2} \ln(2\pi\sigma^2) - \frac{(\mathbf{X}_k - \mu)^2}{2\sigma^2}$$

得
$$\begin{cases} \sum_{k=1}^N \frac{\partial}{\partial \theta_1} \ln p(\mathbf{X}_k | \boldsymbol{\theta}) = \sum_{k=1}^N \frac{\mathbf{X}_k - \theta_1}{\theta_2} = 0 \\ \sum_{k=1}^N \frac{\partial}{\partial \theta_2} \ln p(\mathbf{X}_k | \boldsymbol{\theta}) = \sum_{k=1}^N \left[\frac{-1}{2\theta_2} + \frac{(\mathbf{X}_k - \theta_1)^2}{2\theta_2^2} \right] = 0 \end{cases}$$

由以上方程组解得均值和方差的估计量为

$$\hat{\mu} = \hat{\theta}_1 = \frac{1}{N} \sum_{k=1}^N \mathbf{X}_k$$

$$\hat{\sigma}^2 = \hat{\theta}_2 = \frac{1}{N} \sum_{k=1}^N (\mathbf{X}_k - \hat{\mu})^2$$

类似地，多元正态分布情况：

$$\hat{\mathbf{M}}_i = \frac{1}{N} \sum_{k=1}^N \mathbf{X}_k$$

$$\hat{\mathbf{C}}_i = \frac{1}{N} \sum_{k=1}^N (\mathbf{X}_k - \hat{\mathbf{M}}_i)(\mathbf{X}_k - \hat{\mathbf{M}}_i)^T$$

最大似然估计结果：

- 均值向量的最大似然估计是样本均值；
- 协方差矩阵的最大似然估计是N个单秩矩阵的平均。

4.5.2 贝叶斯估计与贝叶斯学习

贝叶斯**估计**和**学习**都是将未知参数看作随机变量的情况下，求取 ω_i 类模式 X 的分布密度 $p(X|\omega_i)$ 。【注意定义！】

1. 贝叶斯估计和贝叶斯学习的概念

1) 贝叶斯估计

通过参数 θ 关于样本 X^N 的似然函数 $p(X^N|\theta)$ 并利用贝叶斯公式，将随机变量 θ 的先验概率密度 $p(\theta)$ 转为后验概率密度 $p(\theta|X^N)$ ，然后根据 θ 的后验概率密度求出估计量 $\hat{\theta}$ 。

步骤：

- (1) 确定 θ 的先验概率密度 $p(\theta)$ 。
- (2) 由样本集 $X^N = \{X_1, X_2, \dots, X_N\}$ 求出 $p(X^N|\theta)$ 。
- (3) 利用贝叶斯公式求出 θ 的后验概率密度

$$p(\theta|X^N) = \frac{p(X^N|\theta)p(\theta)}{\int p(X^N|\theta)p(\theta)d\theta}$$

(4) 求贝叶斯估计量: $\hat{\theta}^N = \int \theta p(\theta | X^N) d\theta$ 。(4-70)

(5) 求贝叶斯估计出的类概率密度 (教材未给出):

$$\hat{p}(x | X^N) = p(x | \hat{\theta}^N) = p(x | \int \theta p(\theta | X^N) d\theta)$$

2) 贝叶斯学习

利用 θ 的先验概率密度及样本提供的信息, 求出 θ 的后验概率密度, 根据后验概率密度直接求出类概率密度函数 $p(X | \omega_i)$ 。

迭代计算式的推导:

$p(X | \omega_i)$ 由未知参数 θ 确定, 这里讨论只针对 ω_i 类, 可写为

$$p(X | \omega_i) = p(X | \theta) \quad (4-71)$$

假定 $X^N = \{X_1, X_2, \dots, X_N\}$ 是独立抽取的 ω_i 类的一组样本,

设 θ 的后验概率密度函数为 $p(\theta | X^N)$, 根据贝叶斯公式有

$$p(\theta | X^N) = \frac{p(X^N | \theta) p(\theta)}{\int p(X^N | \theta) p(\theta) d\theta} \quad (4-72)$$

$$p(\theta | X^N) = \frac{p(X^N | \theta) p(\theta)}{\int p(X^N | \theta) p(\theta) d\theta} \quad (4-72)$$

式中 $p(X^N | \theta) = p(\mathbf{X}_N | \theta) p(X^{N-1} | \theta)$ (4-73)

将 (4-73) 式代入 (4-72) 式得

除样本 \mathbf{X}_N 以外
其余样本的集合

$$p(\theta | X^N) = \frac{p(\mathbf{X}_N | \theta) \underline{\underline{p(X^{N-1} | \theta) p(\theta)}}}{\int \underline{\underline{p(\mathbf{X}_N | \theta) p(X^{N-1} | \theta) p(\theta) d\theta}} \quad (4-74)$$

类似地,

$$p(\theta | X^{N-1}) p(X^{N-1}) = \underline{\underline{p(X^{N-1} | \theta) p(\theta)}} \quad (4-75)$$

将 (4-75) 式代入 (4-74) 式, 消去 $p(X^{N-1})$ 得

$$p(\theta | X^N) = \frac{p(\mathbf{X}_N | \theta) p(\theta | X^{N-1})}{\int p(\mathbf{X}_N | \theta) p(\theta | X^{N-1}) d\theta} \quad (4-76)$$

参数估计的递推
贝叶斯方法,
迭代过程即是贝
叶斯学习的过程

—— 利用 X^N 估计 $p(\theta | X^N)$ 的迭代计算式

$$p(\theta | X^N) = \frac{p(X_N | \theta) p(\theta | X^{N-1})}{\int p(X_N | \theta) p(\theta | X^{N-1}) d\theta}$$

迭代式的使用： $p(\theta | X^{N-1}) \rightarrow p(\theta | X^N)$

- * 根据先验知识得到 $p(\theta)$ 的初始估计： $p(\theta) = p(\theta | X^0)$
相当于 $N=0$ 时 ($X^N = X^0$) 密度函数的一个估计。

- * 用 X_1 对初始的 $p(\theta)$ 进行修改。

$$p(\theta | X^1) = p(\theta | X_1) = \frac{p(X_1 | \theta) p(\theta)}{\int p(X_1 | \theta) p(\theta) d\theta}$$

- * 给出 X_2 ，对用 $X^1 = \{X_1\}$ 估计的结果进行修改。

$$p(\theta | X^2) = p(\theta | X_1, X_2) = \frac{p(X_2 | \theta) p(\theta | X^1)}{\int p(X_2 | \theta) p(\theta | X^1) d\theta}$$

- * 给出 X_3 ，对用 $X^2 = \{X_1, X_2\}$ 估计的结果进行修改。

$$p(\theta | X^3) = p(\theta | X_1, X_2, X_3) = \frac{p(X_3 | \theta) p(\theta | X^2)}{\int p(X_3 | \theta) p(\theta | X^2) d\theta}$$

* 逐次给出 $\mathbf{X}_3, \mathbf{X}_4, \dots, \mathbf{X}_N$, 得到

$$p(\theta | \mathbf{X}^N) = \frac{p(\mathbf{X}_N | \theta) p(\theta | \mathbf{X}^{N-1})}{\int p(\mathbf{X}_N | \theta) p(\theta | \mathbf{X}^{N-1}) d\theta}$$

由于 $p(\mathbf{x} | \mathbf{X}^N, \theta) = p(\mathbf{x} | \theta)$, 于是

$$\begin{aligned} p(\mathbf{x} | \mathbf{X}^N) &= \int p(\mathbf{x}, \theta | \mathbf{X}^N) d\theta = \int p(\mathbf{x} | \theta, \mathbf{X}^N) p(\theta | \mathbf{X}^N) d\theta \\ &= \int p(\mathbf{x} | \theta) p(\theta | \mathbf{X}^N) d\theta \end{aligned}$$

2. 正态分布密度函数的贝叶斯估计和贝叶斯学习

单变量正态分布、方差 σ^2 已知, 待估计量为均值 μ 。

1) 贝叶斯估计 (特指: 估计概率密度函数的参数)

设 ω_i 类: $p(x | \mu) \sim N(\mu, \sigma^2)$, μ 为未知随机参数。

可以合理地假定 μ 服从正态分布, 设 μ 的先验概率密度 $p(\mu)$:

$$p(\mu) \sim N(\mu_0, \sigma_0^2)$$

设 $\mathbf{x}^N = \{x_1, x_2, \dots, x_N\}$ 是 ω_i 类的 N 个独立抽取的样本, 有

$$p(\mu | x^N) = \frac{p(x^N | \mu) p(\mu)}{\int p(x^N | \mu) p(\mu) d\mu} \quad (4-79)$$

式中, $p(x^N | \mu) = \prod_{k=1}^N p(x_k | \mu)$

有 $p(\mu | x^N) = \alpha \prod_{k=1}^N p(x_k | \mu) p(\mu)$

式中, $\alpha = 1 / \int p(x^N | \mu) p(\mu) d\mu$, 与 μ 无关的比例因子。

由于 $p(x | \mu) \sim N(\mu, \sigma^2)$ $p(\mu) \sim N(\mu_0, \sigma_0^2)$

有 $p(\mu | x^N) = \alpha \prod_{k=1}^N p(x_k | \mu) p(\mu)$

$$= \alpha \prod_{k=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_k - \mu)^2}{2\sigma^2}\right] \cdot \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}\right]$$

$$\begin{aligned}
&= \alpha' \exp \left\{ -\frac{1}{2} \left[\sum_{k=1}^N \frac{(\mu - x_k)^2}{\sigma^2} + \frac{(\mu - \mu_0)^2}{\sigma_0^2} \right] \right\} \\
&= \alpha'' \exp \left\{ -\frac{1}{2} \left[\left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2} \right) \mu^2 - 2 \left(\frac{1}{\sigma^2} \sum_{k=1}^N x_k + \frac{\mu_0}{\sigma_0^2} \right) \mu \right] \right\}
\end{aligned}$$

式中， α' 和 α'' 收入与 μ 无关的项。

将 $p(\mu | x^N)$ 写为正态分布密度函数的标准形式 $N(\mu_N, \sigma_N^2)$ ：

$$p(\mu | x^N) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp \left\{ -\frac{1}{2} \left(\frac{\mu - \mu_N}{\sigma_N} \right)^2 \right\}$$

可求得 μ_N 和 σ_N^2 分别为

$$\mu_N = \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2} m_N + \frac{\sigma^2}{N\sigma_0^2 + \sigma^2} \mu_0$$

$$\sigma_N^2 = \frac{\sigma_0^2 \sigma^2}{N\sigma_0^2 + \sigma^2}$$

$$\text{式中, } m_N = \frac{1}{N} \sum_{k=1}^N x_k$$

由式 (4-70) $\hat{\theta} = \int \theta p(\theta | X^N) d\theta$ 计算 μ 的贝叶斯估计为

$$\begin{aligned}\hat{\mu} &= \int \mu p(\mu | x^N) d\mu = \int \mu \frac{1}{\sqrt{2\pi}\sigma_N} \exp\left[-\frac{1}{2}\left(\frac{\mu - \mu_N}{\sigma_N}\right)^2\right] d\mu \\ &= \mu_N \\ &= \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2} m_N + \frac{\sigma^2}{N\sigma_0^2 + \sigma^2} \mu_0\end{aligned}$$

当 $N(\mu_0, \sigma_0^2) = N(0, 1)$ 且 $\sigma^2 = 1$ (合理要求) 时, 有

$$\hat{\mu} = \frac{N}{N+1} m_N = \frac{1}{N+1} \sum_{k=1}^N x_k$$

——与最大似然估计形式类似

2) 贝叶斯学习（特指：求取概率密度函数pdf）

递推求解出均值的后验概率密度 $p(\mu | X^N)$ 后，直接计算模型平均后的类概率密度函数 $p(x) \equiv p(x | X^N) \equiv p(x | X^N)$ 。

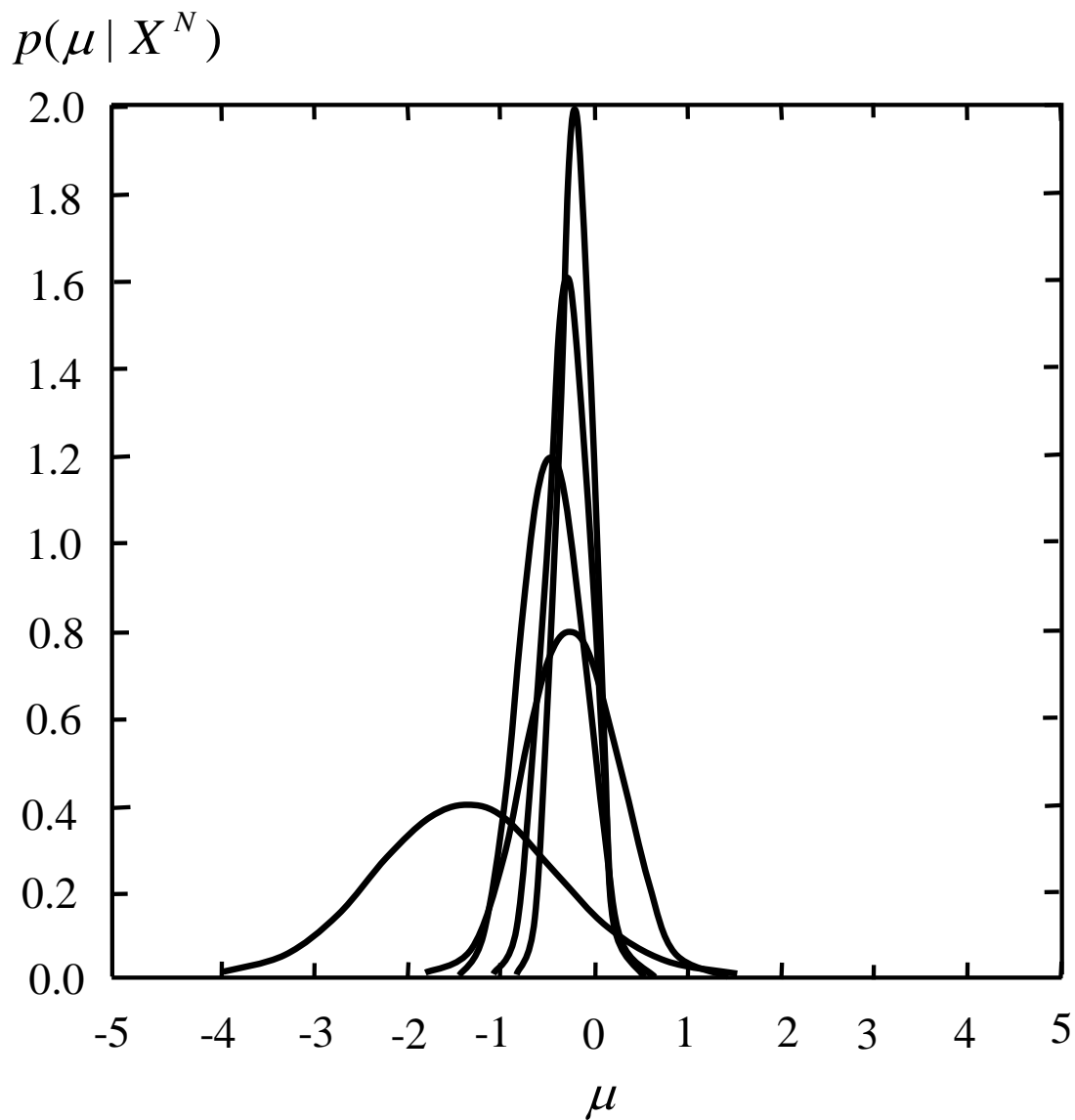
$$\text{首先，由前有： } p(\mu | x^N) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp \left\{ -\frac{1}{2} \left(\frac{\mu - \mu_N}{\sigma_N} \right)^2 \right\}$$

$$\text{式中， } \mu_N = \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2} m_N + \frac{\sigma^2}{N\sigma_0^2 + \sigma^2} \mu_0$$

$$\sigma_N^2 = \frac{\sigma_0^2 \sigma^2}{N\sigma_0^2 + \sigma^2}$$

μ_N ：观察了 N 个样本后对 μ 的最好估计；

σ_N^2 ：估计的不确定性。



当 N 增大时：
 $p(\mu | X^N)$ 越来越尖峰突起，
 σ_N^2 单调减小；

当 N 趋于无穷时：
 $p(\mu | X^N)$ 趋于 δ 函数，
 σ_N^2 趋于零。

图4.14 均值的贝叶斯学习过程示意图

接着，计算模型平均后的类概率密度函数 $p(x | x^N)$ ：

$$\begin{aligned} p(x | x^N) &= \int p(x | \mu) p(\mu | x^N) d\mu \text{——全概率公式} \\ &= \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \cdot \frac{1}{\sqrt{2\pi}\sigma_N} \exp\left[-\frac{(\mu-\mu_N)^2}{2\sigma_N^2}\right] d\mu \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2 + \sigma_N^2}} \exp\left[-\frac{(x-\mu_N)^2}{2(\sigma^2 + \sigma_N^2)}\right] \end{aligned}$$

可见：

$p(x | x^N)$ 是正态分布；

均值为 μ_N ；

方差为 $(\sigma^2 + \sigma_N^2)$ 。

全概率公式

$$p(x) = \int p(x | \mu) p(\mu) d\mu$$

$$\begin{aligned} p(x | \underline{x^N}) &= \int p(x | \underline{\mu}, \underline{x^N}) p(\mu | \underline{x^N}) d\mu \\ &= \int p(x | \underline{\mu}) p(\mu | x^N) d\mu \end{aligned}$$

多维正态模式分布: $p(\mathbf{X} | \omega_i) = p(\mathbf{X} | \mathbf{M}) \sim N(\mathbf{M}, \mathbf{C})$

其中: \mathbf{C} 已知, \mathbf{M} 未知。

\mathbf{M} 的先验概率密度函数 $p(\mathbf{M})$: $p(\mathbf{M}) \sim N(\mathbf{M}_0, \mathbf{C}_0)$

则利用贝叶斯估计得到的 \mathbf{M} 的后验概率密度函数为

$$p(\mathbf{M} | \mathbf{X}^N) \sim N(\mathbf{M}_N, \mathbf{C}_N)$$

$$\text{其中, } \mathbf{M}_N = \mathbf{C}_0(\mathbf{C}_0 + \frac{1}{N}\mathbf{C})^{-1}\hat{\mathbf{M}} + \frac{1}{N}\mathbf{C}(\mathbf{C}_0 + \frac{1}{N}\mathbf{C})^{-1}\mathbf{M}_0$$

$$\mathbf{C}_N = \frac{1}{N}\mathbf{C}(\mathbf{C}_0 + \frac{1}{N}\mathbf{C})^{-1}\mathbf{C}_0$$

$$\hat{\mathbf{M}} = \frac{1}{N} \sum_{k=1}^N \mathbf{X}_k$$

根据贝叶斯学习得到的类概率密度函数为

$$p(\mathbf{X} | \mathbf{X}^N) = \int p(\mathbf{X} | \mathbf{M}) p(\mathbf{M} | \mathbf{X}^N) d\mathbf{M}$$

均值为 \mathbf{M}_N , 协方差矩阵为 $(\mathbf{C} + \mathbf{C}_N)$ 。

4.6 概率密度函数的非参数估计方法

根据样本直接估计类概率密度函数的方法。

4.6.1 基本方法

1. 出发点：基于事实

随机向量 \mathbf{X} 落入区域 R 的概率 P 为： $P = \int_R p(\mathbf{X}) d\mathbf{X}$ 。

$p(\mathbf{X})$ ：类概率密度函数。

设从密度为 $p(\mathbf{X})$ 的总体中独立抽取的样本 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$ 。

若观察到： N 个样本中有 k_N 个落入区域 R 中，则可以合理认为

$$\hat{P} \approx k_N / N$$

其中 \hat{P} ： \mathbf{X} 落入区域 R 中概率 P 的估计。

类概率密度函数 $p(\mathbf{X})$ 的估计:

设 $p(\mathbf{X})$ 连续, 区域 R 足够小且体积为 V , $p(\mathbf{X})$ 在 R 中没有变化, \mathbf{X} 是 R 中的点。设 $\hat{p}(\mathbf{X})$ 为 \mathbf{X} 点概率密度的估计

$$P = \int_R p(\mathbf{X}) d\mathbf{X} = p(\mathbf{X})V$$

$$\text{因此有: } \hat{P} \triangleq \int_R \hat{p}(\mathbf{X}) d\mathbf{X} \approx \hat{p}(\mathbf{X})V,$$

$$\text{前面有: } \hat{P} \approx \frac{k}{N}$$

$$\text{于是有: } \hat{p}(\mathbf{X}) \approx \frac{k_N/N}{V} \quad (4-91)$$

2. 存在的两个问题

- 1) 固定 V , 样本数增多, 则 k_N/N 以概率1收敛。但只能得到在某一体积 V 中的平均估计。
- 2) N 固定, V 趋于零, $p(\mathbf{X}) \approx 0$ 或发散到无穷大。没有意义。

必须注意 V 、 k_N 、 k_N/N 随 N 变化的趋势和极限, 保持合理性。

3. 估计的步骤:

- * 构造一串包含 \mathbf{X} 的区域 $R_1, R_2, \dots, R_N, \dots$
- * 对 R_1 采用一个样本估计, 对 R_2 采用两个样本,
- * 假定 V_N 是 R_N 的体积, k_N 是落入 R_N 内的样本数目, $\hat{p}_N(\mathbf{X})$ 是 $p(\mathbf{X})$ 的第 N 次估计, 有

$$\hat{p}_N(\mathbf{X}) = \frac{k_N/N}{V_N} \quad (4-92)$$

4. 为保证估计合理性应满足的三个条件

1) $\lim_{N \rightarrow \infty} V_N = 0$

$\hat{p}_N(\mathbf{X})$ 能代表 \mathbf{X} 点的密度 $p(\mathbf{X})$ 。

2) $\lim_{N \rightarrow \infty} k_N = \infty$

使式右边能以概率1收敛于 $p(\mathbf{X})$

3) $\lim_{N \rightarrow \infty} k_N/N = 0$

落入 R_N 中的样本数始终是总数中的极小部分

5. 两种非参数估计法: Parzen窗法、 k_N 近邻估计法。

4.6.2 Parzen窗法

1. Parzen窗估计的基本概念

设区域 R_N : d 维超立方体, 棱长: h_N , 则

$$V_N = h_N^d$$

定义窗函数 $\phi(u)$:

$$\phi(u) = \begin{cases} 1, & \text{当 } |u_j| \leq \frac{1}{2}; j = 1, 2, \dots, d \\ 0, & \text{其它} \end{cases}$$

以原点为中心的
超立方体

当 \mathbf{X}_i 落入以 \mathbf{X} 为中心, 体积为 V_N 的超立方体时:

$$\phi[(\mathbf{X} - \mathbf{X}_i)/h_N] = 1$$

否则

$$\phi[(\mathbf{X} - \mathbf{X}_i)/h_N] = 0$$

落入超立方体内的样本 \mathbf{X}_i 的个数为

$$k_N = \sum_{i=1}^N \phi\left(\frac{\mathbf{X} - \mathbf{X}_i}{h_N}\right) \quad (4-95)$$

代入 $\hat{p}_N(\mathbf{X}) = \frac{k_N/N}{V_N}$ 得

$$\hat{p}_N(\mathbf{X}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{V_N} \phi\left(\frac{\mathbf{X} - \mathbf{X}_i}{h_N}\right)$$

—— Parzen窗法基本公式

实质：

窗函数的作用是内插（平滑），样本对估计所起的作用取决于它到 \mathbf{X} 的距离。

为使 $\hat{p}_N(\mathbf{X})$ 成为密度函数, $\phi(u)$ 应满足的两个条件:

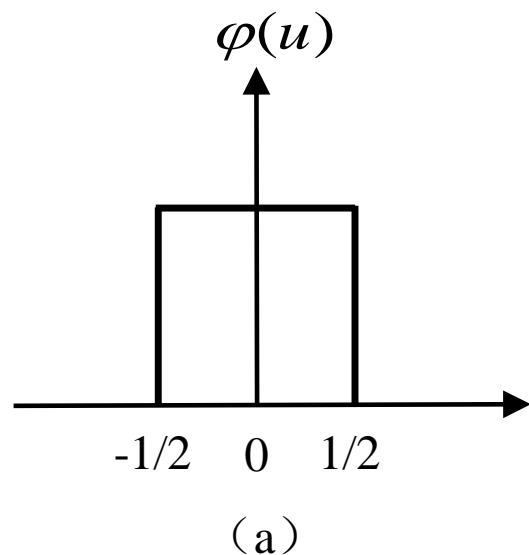
- 1) $\phi(u) \geq 0$;
- 2) $\int \phi(u) du = 1$ 。

2. 窗函数的选择

一维形式

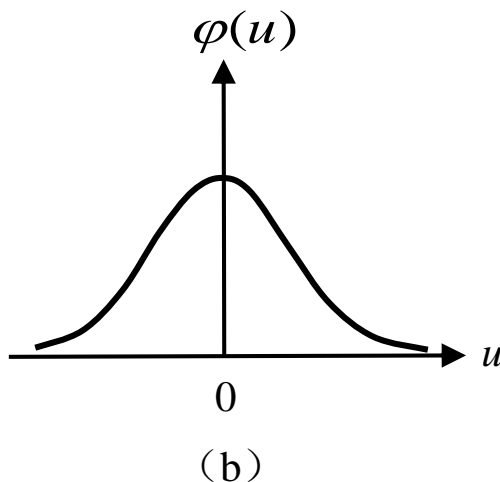
1) 方窗函数

$$\phi(u) = \begin{cases} 1, & |u| \leq \frac{1}{2} \\ 0, & \text{其它} \end{cases}$$



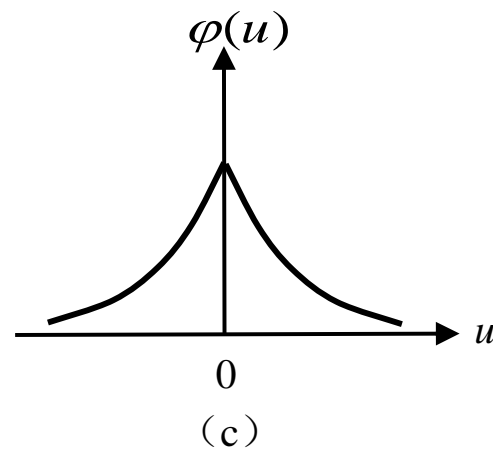
2) 正态窗函数

$$\phi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$$



3) 指数窗函数

$$\phi(u) = \exp(-|u|)$$



满足条件 $\phi(u) \geq 0$ 和 $\int \phi(u) du = 1$ 的都可以作为窗函数。

最终估计效果的好坏与样本情况、窗函数以及窗函数参数的选择有关。

3. 窗宽 h_N 对估计量 $\hat{p}_N(\mathbf{X})$ 的影响

定义
$$\delta_N(\mathbf{X}) = \frac{1}{V_N} \phi\left(\frac{\mathbf{X}}{h_N}\right)$$

有
$$\hat{p}_N(\mathbf{X}) = \frac{1}{N} \sum_{i=1}^N \delta_N(\mathbf{X} - \mathbf{X}_i)$$

h_N 影响 $\delta_N(\mathbf{X})$ 的幅度，对 $\hat{p}_N(\mathbf{X})$ 的影响很大。

如何选取根据经验折中考虑。

4. 估计量 $\hat{p}_N(\mathbf{X})$ 的统计性质

满足某些限制条件时， $\hat{p}_N(\mathbf{X})$ 渐近无偏和平方误差一致。

限制条件：

1) 总体密度函数 $p(\mathbf{X})$ 在 \mathbf{X} 点连续；

2) 窗函数满足以下条件:

$$\phi(u) \geq 0$$

$$\int \phi(u) du = 1$$

保证 $\hat{p}_N(\mathbf{X})$ 有
密度函数的性质

$$\sup_u \phi(u) < \infty$$

保证 $\phi(u)$ 有界

$$\lim_{\|u\| \rightarrow \infty} \phi(u) \prod_{i=1}^d u_i = 0$$

使 $\phi(u)$ 随 u 的增加
较快趋于零

3) 窗函数受下列条件的约束:

$$\lim_{N \rightarrow \infty} V_N = 0$$

$$\lim_{N \rightarrow \infty} NV_N = \infty$$

使体积随 N 的增大
趋于零时, 速度
低于 N 增加的速度

例 4.5 设待估计的 $p(\mathbf{X})$ 是均值为零，方差为 1 的正态密度函数。随机地抽取含有 1 个、16 个、256 个学习样本 X_i 的样本集。试分别根据这三个样本集用 Parzen 窗法估计 $p(\mathbf{X})$ ，即求估计式 $\hat{p}_N(\mathbf{X})$ 。

解：考虑 \mathbf{X} 是一维模式向量的情况。选择正态窗函数

$$\phi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$$

并设 $h_N = h_1/\sqrt{N}$ ， h_1 为可调节的参数。

有

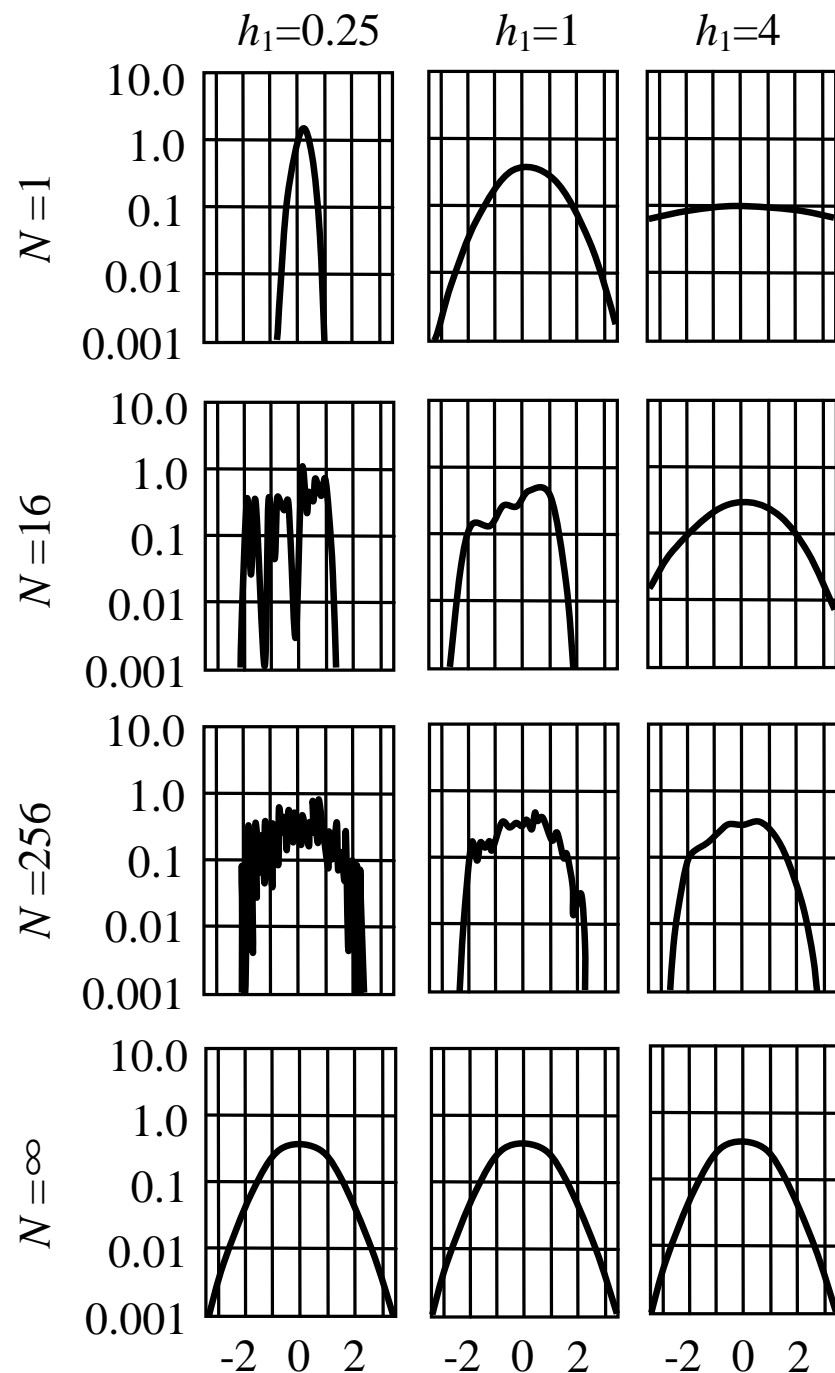
$$\begin{aligned}\phi\left(\frac{x - x_i}{h_N}\right) &= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - x_i}{h_N}\right)^2\right] \\ &= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - x_i}{h_1/\sqrt{N}}\right)^2\right]\end{aligned}$$

$$\hat{p}_N(x) = \frac{1}{N} \sum_{i=1}^N \frac{1}{h_N} \phi\left(\frac{x - x_i}{h_N}\right)$$

$$= \frac{1}{N} \sum_{i=1}^N \frac{\sqrt{N}}{h_1} \phi\left(\frac{x - x_i}{h_N}\right)$$

$$= \frac{1}{h_1 \sqrt{N}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{x - x_i}{h_1 / \sqrt{N}}\right)^2\right]$$

估计结果：

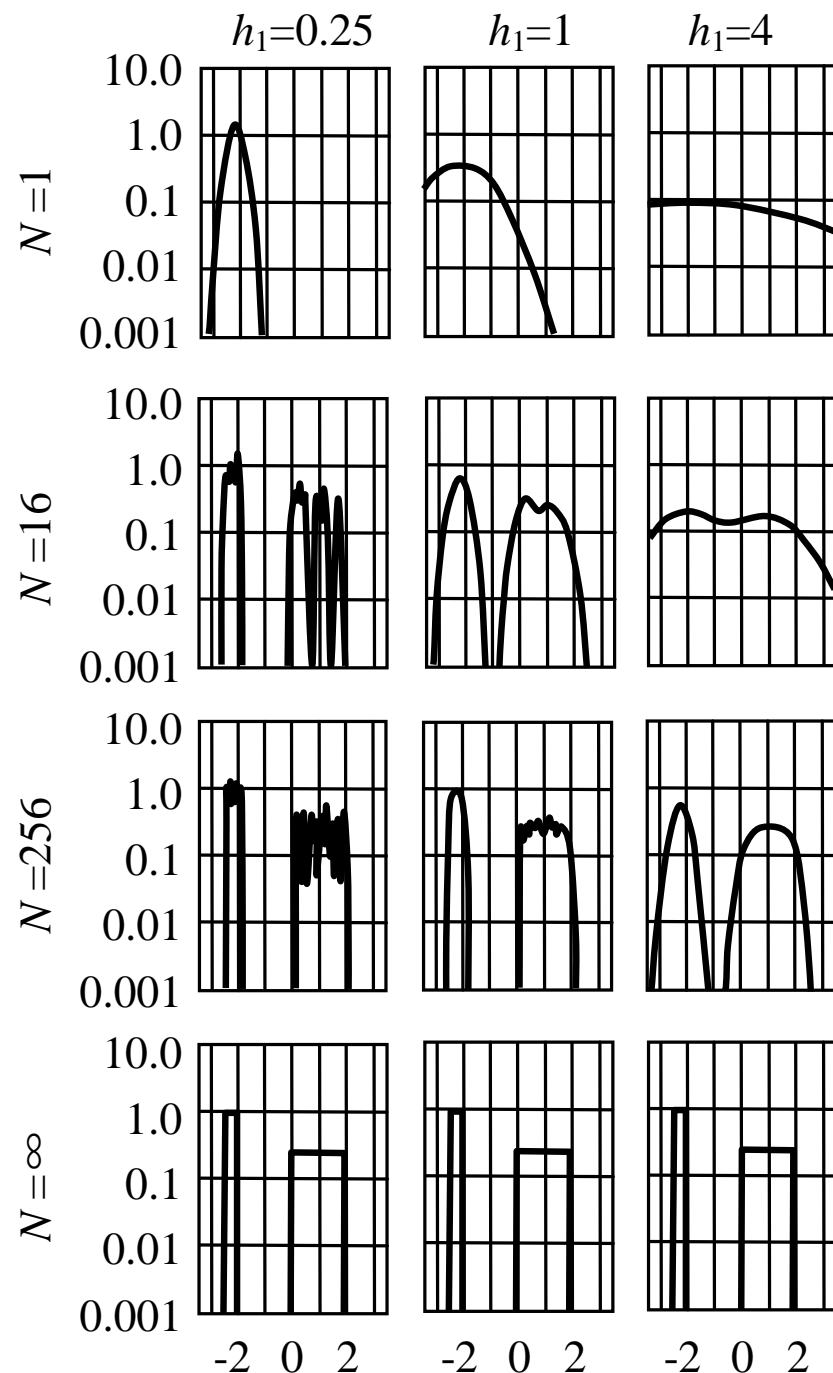


例 4.6 仍以一维情况为例，假定待估计的概率密度函数 $p(x)$ 为两个均匀分布密度的混合

$$p(x) = \begin{cases} 1, & -2.5 < x < -2 \\ 0.25, & 0 < x < 2 \\ 0, & \text{其它} \end{cases}$$

随机抽取含 1 个、16 个、256 个学习样本的样本集，求 $p(x)$ 的估计 $\hat{p}_N(x)$ 。

解：估计结果



Parzen窗法特点：

- * 具有一般性，适用于单峰、**多峰**形式。
- * 要得到较精确的估计**必须抽取大量的样本**。

（一般非参数估计法的共同问题）

比参数估计法多得多；

样本数目随模式维数一般按指数规律增长。

4.6.3 k_N -近邻估计法

基本思想：

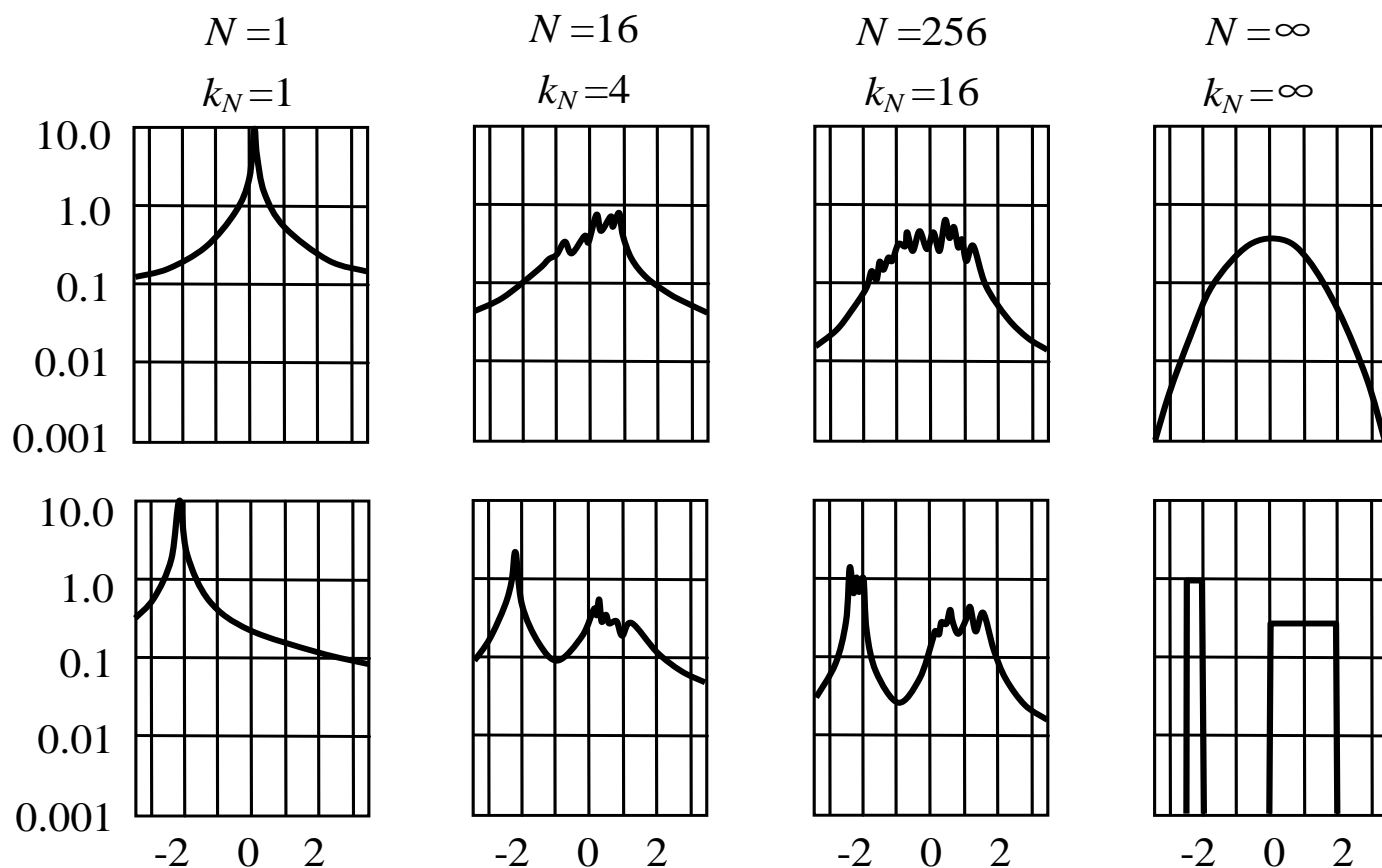
使体积为样本密度的函数，而不是样本数 N 的函数。

$$\hat{p}_N(\mathbf{X}) = \frac{k_N/N}{V_N}$$

限制条件仍然是：

$$\lim_{N \rightarrow \infty} V_N = 0, \quad \lim_{N \rightarrow \infty} k_N = \infty, \quad \lim_{N \rightarrow \infty} k_N / N = 0。$$

例4.5和4.6中，用 k_N -近邻法估计的 $p(\mathbf{X})$ 的结果：



4.7 后验概率密度函数的势函数估计法

方法： 直接采用后验概率密度 $p(\omega_i | \mathbf{X})$ 进行分类。

关键： 利用所给的训练样本估计概率密度函数， M 个。

$$\{p_i(\omega_i | \mathbf{X}), i = 1, 2, \dots, M\}$$

判别准则：

如果 $p(\omega_i | \mathbf{X}) > p(\omega_j | \mathbf{X}) \quad \forall j \neq i$, 则 $\mathbf{X} \in \omega_i$

采用势函数直接由训练样本集来迭代估计 $p(\omega_i | \mathbf{X})$ 的方法：

迭代估计 $p(\omega_i | \mathbf{X})$ 的势函数法

从上面判别式可看出后验概率密度 $p(\omega_i | \mathbf{X})$ 直接做判别函数 $d_i(\mathbf{X})$ ，采用的方法相当于线性判别函数中讲的第三种情况，只是此处的判别函数为非线性。

$p(\omega_i | \mathbf{X})$ 可用 $\hat{f}_k(\mathbf{X})$ 近似代表, $\hat{f}_k(\mathbf{X}) \in [0, 1]$, k -迭代次数。

$\hat{f}_k(\mathbf{X})$ 可由合适的势函数 $K(\mathbf{X}, \mathbf{X}_k)$ 组成。

势函数的确定方法有两种方法:

同“非线性判别函数”中的类似

(1) 用对称的有限项多项式展开:

$$K(\mathbf{X}, \mathbf{X}_k) = \sum_{i=1}^m \varphi_i(\mathbf{X}_k) \varphi_i(\mathbf{X})$$

其中 $\{\varphi_i(\mathbf{X})\}$ 在模式定义域内为正交函数集。

(2) 一些“双变量 \mathbf{X} 和 \mathbf{X}_k 的对称函数”, 如:

$$K(\mathbf{X}, \mathbf{X}_k) = \exp \left\{ -\alpha \|\mathbf{X} - \mathbf{X}_k\|^2 \right\}$$

$$K(\mathbf{X}, \mathbf{X}_k) = \frac{1}{1 + \alpha \|\mathbf{X} - \mathbf{X}_k\|^2}$$

式中 α 为正常数。

第*i*类判别函数的迭代算法：

从取 $\hat{f}_0(\mathbf{X})=0$ 开始。

第一步：将模式样本 \mathbf{X}_1 送入分类器，产生势函数 $K(\mathbf{X}, \mathbf{X}_k)$ ，有

$$\hat{f}_1(\mathbf{X}) = \begin{cases} \hat{f}_0(\mathbf{X}) + r_1 K(\mathbf{X}, \mathbf{X}_1) = r_1 K(\mathbf{X}, \mathbf{X}_1) & , \text{ 若 } \mathbf{X}_1 \in \omega_i \\ \hat{f}_0(\mathbf{X}) - r_1 K(\mathbf{X}, \mathbf{X}_1) = -r_1 K(\mathbf{X}, \mathbf{X}_1) & , \text{ 若 } \mathbf{X}_1 \notin \omega_i \end{cases}。$$

第二步：将 \mathbf{X}_2 送入分类器，它对近似 $p(\omega_i | \mathbf{X})$ 的函数 $\hat{f}_k(\mathbf{X})$ 的影响有三种可能情况：

① 若 $\mathbf{X}_2 \in \omega_i$ 且 $\hat{f}_1(\mathbf{X}_2) > 0$ ，或 $\mathbf{X}_2 \notin \omega_i$ 且 $\hat{f}_1(\mathbf{X}_2) < 0$ ，

分类正确， $\hat{f}_2(\mathbf{X}) = \hat{f}_1(\mathbf{X})$ 。

② 若 $\mathbf{X}_2 \in \omega_i$ ，但 $\hat{f}_1(\mathbf{X}_2) \leq 0$ ，

错误分类， $\hat{f}_2(\mathbf{X}) = \hat{f}_1(\mathbf{X}) + r_2 K(\mathbf{X}, \mathbf{X}_2)$

③ 若 $\mathbf{X}_2 \notin \omega_i$ ，但 $\hat{f}_1(\mathbf{X}_2) \geq 0$ ，

错误分类， $\hat{f}_2(\mathbf{X}) = \hat{f}_1(\mathbf{X}) - r_2 K(\mathbf{X}, \mathbf{X}_2)$

.....

第 $k+1$ 步：送入训练样本 \mathbf{X}_{k+1} ，相应的势函数为 $K(\mathbf{X}, \mathbf{X}_{k+1})$ ，亦有

①若 $\mathbf{X}_{k+1} \in \omega_i$ 且 $\hat{f}_k(\mathbf{X}_{k+1}) > 0$ 或 $\mathbf{X}_{k+1} \notin \omega_i$ 且 $\hat{f}_k(\mathbf{X}_{k+1}) < 0$ ，

正确分类， $\hat{f}_{k+1}(\mathbf{X}) = \hat{f}_k(\mathbf{X})$ 。

② 若 $\mathbf{X}_{k+1} \in \omega_i$ ，且 $\hat{f}_k(\mathbf{X}_{k+1}) \leq 0$ ，错误分类，

$$\hat{f}_{k+1}(\mathbf{X}) = \hat{f}_k(\mathbf{X}) + r_{k+1}K(\mathbf{X}, \mathbf{X}_{k+1})$$

③ 若 $\mathbf{X}_{k+1} \notin \omega_i$ ，且 $\hat{f}_k(\mathbf{X}_{k+1}) \geq 0$ ，错误分类，

$$\hat{f}_{k+1}(\mathbf{X}) = \hat{f}_k(\mathbf{X}) - r_{k+1}K(\mathbf{X}, \mathbf{X}_{k+1})$$

迭代系数 r_k , $k = 1, 2, \dots$ ，为正实数序列，应满足：

$$\lim_{k \rightarrow \infty} r_k = 0, \quad \sum_{k=1}^{\infty} r_k = \infty \quad \text{和} \quad \sum_{k=1}^{\infty} r_k^2 < \infty$$

例，可采用调和序列 $\left\{ \frac{1}{k} \right\}$ 。

说明

1. 与确定性分类器中的势函数法相比较：

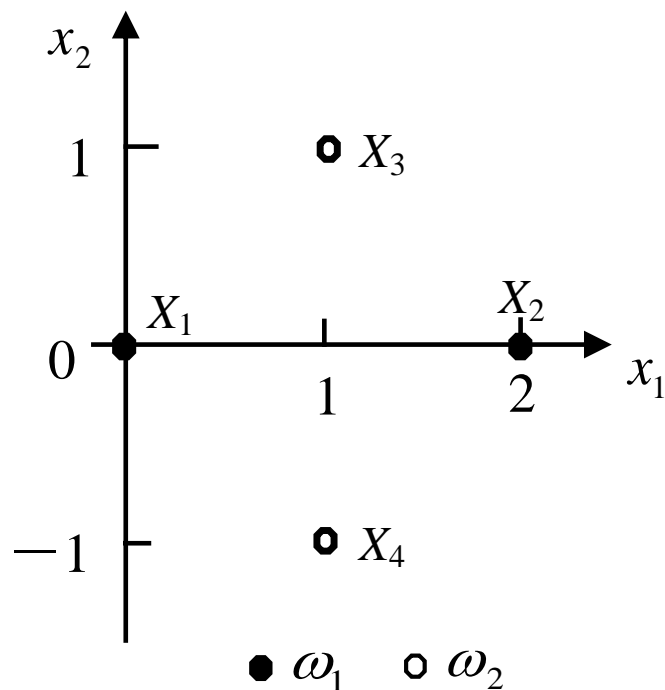
相同点：迭代算法的思路完全相同。

不同点：所取的 r_k 不同。

几何分类法中： $r_k = 0, 1, -1$

概率分类法中： $\lim_{k \rightarrow \infty} r_k = 0$ 、 $\sum_{k=1}^{\infty} r_k = \infty$ 和 $\sum_{k=1}^{\infty} r_k^2 < \infty$

2. 近似函数 $\hat{f}_k(\mathbf{X})$ 的产生与训练样本集密切相关，而训练样本是随机出现的，所以 $\hat{f}_k(\mathbf{X})$ 也是随机函数，它随着迭代次数的增加收敛于判别函数 $p(\omega_i | \mathbf{X})$ 。



例4.7 给定训练样本

$$\omega_1 : \mathbf{X}_1 = [0,0]^T, \quad \mathbf{X}_2 = [2,0]^T,$$

$$\omega_2 : \mathbf{X}_3 = [1,1]^T, \quad \mathbf{X}_4 = [1,-1]^T,$$

选择合适的势函数对模式进行分类。

解：① 从图上可看出两类模式不是线性可分的，

选择指数型二维势函数 ($\alpha = 1$) :

$$K(\mathbf{X}, \mathbf{X}_k) = \exp \left\{ -\|\mathbf{X} - \mathbf{X}_k\|^2 \right\} = \exp \left\{ -[(x_1 - x_{k1})^2 + (x_2 - x_{k2})^2] \right\}$$

② 设初值 $\hat{f}_0(\mathbf{X}) = 0$, $r_k = \frac{1}{k}$

③ 迭代

第一步：送入 $\mathbf{X}_1 \in \omega_1$,

$$\hat{f}_1(\mathbf{X}) = \hat{f}_0(\mathbf{X}) + r_1 K_1(\mathbf{X}, \mathbf{X}_1) = \exp\left\{-[(x_1 - 0)^2 + (x_2 - 0)^2]\right\} = \exp\left\{-[x_1^2 + x_2^2]\right\}$$

第二步：送入 $\mathbf{X}_2 \in \omega_1$, $\hat{f}_1(\mathbf{X}_2) = e^{-4} > 0$, 正确分类, $\therefore \hat{f}_2(\mathbf{X}) = \hat{f}_1(\mathbf{X})$

第三步：送入 $\mathbf{X}_3 \in \omega_2$, $\hat{f}_2(\mathbf{X}_3) = e^{-2} > 0$, 分类错误,

$$\begin{aligned}\hat{f}_3(\mathbf{X}) &= \hat{f}_2(\mathbf{X}) - r_3 K(\mathbf{X}, \mathbf{X}_3) \\ &= \exp\left\{-[x_1^2 + x_2^2]\right\} - \frac{1}{3} \exp\left\{-[(x_1 - 1)^2 + (x_2 + 1)^2]\right\} \\ &= \exp\left\{-[x_1^2 + x_2^2]\right\} - \frac{1}{3} \exp\left\{-[x_1^2 + x_2^2 - 2x_1 + 2x_2 + 2]\right\}\end{aligned}$$

第四步：送入 $\mathbf{X}_4 \in \omega_2$ ， $\hat{f}_3(\mathbf{X}_4) = e^{-2} - \frac{1}{3}e^{-4} > 0$ ，分类错误，

$$\begin{aligned}\therefore \hat{f}_4(\mathbf{X}) &= \hat{f}_3(\mathbf{X}) - r_4 K(\mathbf{X}, \mathbf{X}_4) \\ &= \hat{f}_3(\mathbf{X}) - \frac{1}{4} \exp\{ -[(x_1 - 1)^2 + (x_2 + 1)^2] \} \\ &= \hat{f}_3(\mathbf{X}) - \frac{1}{4} \exp\{ -[x_1^2 + x_2^2 - 2x_1 - 2x_2 + 2] \}\end{aligned}$$

取 $\mathbf{X}_5 = \mathbf{X}_1$ ， $\mathbf{X}_6 = \mathbf{X}_2$ ，到全部训练样本完成一次无错分类的迭代为止，算法已收敛于解：

$$\begin{aligned}\hat{f}(\mathbf{X}) &= \exp\{ -[x_1^2 + x_2^2] \} - \frac{1}{3} \exp\{ -[x_1^2 + x_2^2 - 2x_1 + 2x_2 + 2] \} \\ &\quad - \frac{1}{4} \exp\{ -[x_1^2 + x_2^2 - 2x_1 - 2x_2 + 2] \} + \frac{1}{6} \exp\{ -[x_1^2 + x_2^2 - 4x_1 + 4] \}\end{aligned}$$

判别函数 $d(\mathbf{X}) = p(\omega_1 | \mathbf{X}) = \hat{f}(\mathbf{X})$

课后作业

- 见另文。
- 下次上课前提交。
- 最好使用电子档。

End of This Part