率为

$$P_N(e|\boldsymbol{x}) = \sum_{j=1}^{c}\sum_{i\neq j}P(\omega_j|\boldsymbol{x})P(\omega_i|\boldsymbol{x}') \tag{6-96}$$

当 $N\to\infty$ 时渐近条件错误率为

$$P(e|\boldsymbol{x}) = \lim_{N\to\infty}P_N(e|\boldsymbol{x}) = \sum_{j=1}^{c}\sum_{i\neq j}P(\omega_j|\boldsymbol{x})P(\omega_i|\boldsymbol{x}) \tag{6-97}$$

两式相减并经化简后可得

$$|P_N(e|\boldsymbol{x}) - P(e|\boldsymbol{x})| = \left|\sum_{j=1}^{c}P(\omega_j|\boldsymbol{x})[P(\omega_j|\boldsymbol{x}) - P(\omega_j|\boldsymbol{x}')]\right| \tag{6-98}$$

我们再次在 $\boldsymbol{x}$ 的局部邻近区域 A 中对 $P(\omega_j|\boldsymbol{x}')$ 取线性近似可得

$$|P_N(e|\boldsymbol{x}) - P(e|\boldsymbol{x})| \doteq \left|\sum_{j=1}^{c}P(\omega_j|\boldsymbol{x})\nabla_1 P(\omega_j|\boldsymbol{x})^T(\boldsymbol{x}-\boldsymbol{x}')\right| \tag{6-99}$$

在这种情况下，令

$$\nabla_1 = \sum_{j=1}^{c}P(\omega_j|\boldsymbol{x})\nabla_1 P(\omega_j|\boldsymbol{x}) \tag{6-100}$$

则与两类情况一样，极小化

$$E\{|P_N(e|\boldsymbol{x}) - P(e|\boldsymbol{x})|^2|\boldsymbol{x}\}$$

等价于极小化

$$E\{|\nabla_1^T(\boldsymbol{x}-\boldsymbol{x}')|^2|\boldsymbol{x}\}$$

和(6-90)式一样，我们可将 ∇_1 写成局部均值向量的形式，

$$\nabla_1 = \sum_{i=1}^{c}P(\omega_i|\boldsymbol{x})\nabla P(\omega_i|\boldsymbol{x}) \doteq \frac{1}{c}\sum_{i=1}^{c}P(\omega_i|\boldsymbol{x})^2[\mu_i(\boldsymbol{x}) - \mu_0(\boldsymbol{x})] \tag{6-101}$$

利用局部样本可估计 $P(\omega_i|\boldsymbol{x})$ 为

$$P(\omega_i|\boldsymbol{x}) = \frac{N_i}{N} \tag{6-102}$$

同样可用局部样本均值向量 M_i 代替 μ_i，这样式(6-101)可以写成

$$\nabla_1 = \frac{1}{c}\sum_{i=1}^{c}\left(\frac{N_i}{N}\right)^2[M_i(\boldsymbol{x}) - M_0(\boldsymbol{x})] \tag{6-103}$$

求出 ∇_1 后，下面的问题完全和两类情况一样，也就是说最佳距离度量也用

$$D(\boldsymbol{x},\boldsymbol{x}^l) = |\nabla_1^T(\boldsymbol{x}-\boldsymbol{x}^l)| \tag{6-104}$$

表示，并应用类似算法，可求得任意 $\boldsymbol{x}$ 的最近邻 $\boldsymbol{x}'$，然后按最近邻法分类。

从上面分析可见，采用现在定义的距离度量，在不增加样本数 N 的情况下，可以使最近邻法的性能与 $N\to\infty$ 时最近邻法的性能的平方差别最小，因此我们称这样定义的距离是最佳的。

最后还要指出，上面的结果可以推广到 k-近邻法，只不过推导和距离计算都较复杂，这里不再讨论。

习　　题

6.1 举例说明最近邻决策面是分段线性的。

6.2 证明式(6-14)～式(6-18)。

6.3 说明在什么情况下，最近邻法平均错误率 P 达到其上界。

6.4 已知在 $P^*=0$ 或 $P^*=(c-1)/c$ 时最近邻法平均错误率 P 等于贝叶斯错误率 P^*。问当 P^* 界于这两种极端情况之间时，是否还可能使 $P=P^*$。

(1)当 $P(\omega_i)=1/c$

$$p(x|\omega_i)=\begin{cases}1, 0\leqslant x\leqslant \dfrac{cr}{c-1}\\ 1, i\leqslant x\leqslant i+1-\dfrac{cr}{c-1}\\ 0, \text{其它}\end{cases}$$

时，证明一维问题的贝叶斯错误率 $P^*=r$。

(2)证明此时 $P=P^*$。

6.5 有七个二维向量：$\boldsymbol{x}_1^T=(1,0)$、$\boldsymbol{x}_2^T=(0,1)$、$\boldsymbol{x}_3^T=(0,-1)$、$\boldsymbol{x}_4^T=(0,0)$、$\boldsymbol{x}_5^T=(0,2)$、$\boldsymbol{x}_6^T=(0,-2)$、$\boldsymbol{x}_7^T=(-2,0)$，假定前三个为 ω_1 类，后四个为 ω_2 类。

(1)画出最近邻法决策面；

(2)求样本均值 m_1，m_2。若按离样本均值距离的大小进行分类，试画出决策面。

6.6 画出 k-近邻法程序框图。

6.7 对于有限样本情况，重复剪辑是否比两分剪辑的特性要好？

6.8 证明“6.3.1 近邻法的快速算法”中规则 2。