

Chapter2

Author: 中山大学 17数据科学与计算机学院 YSY

<https://github.com/ysyisyourbrother>

引言

如果是1类被判定为2类就是一个错判error:

$$P(error|x) = \begin{cases} P(\omega_1|x) & \text{如果判定 } \omega_2 \\ P(\omega_2|x) & \text{如果判定 } \omega_1 \end{cases}$$

对于某一给定x, 我们可以在最小化误差的概率下去判决。

如果 $P(\omega_1|x) > P(\omega_2|x)$, 判别为 ω_1 ; 否则判别为 ω_2

平均的误差概率可以表示为, 对每个可能的x, 判错的概率均值为:

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

误差先验概率就是: x和error的共同出现概率, 然后对x求和。

贝叶斯决策论——连续特征

引入风险的概念。令 $\omega_1 \dots \omega_c$ 是c个类别, 而 $\alpha_1 \dots \alpha_a$ 是有限的a种可能采取的行为, 可能是判定某个数据为哪一类。**风险 (损失) 函数** $\lambda(\alpha_i|\omega_j)$ 表示实际类别为j的时候采取行动i的风险。

假如对某个模式x采取行动i, 而他的真实类别为j, 损失为:

$$R(\alpha_i|\mathbf{x}) = \sum_{j=1}^c \lambda(\alpha_i|\omega_j) P(\omega_j|\mathbf{x})$$

一个预期的损失被称为**风险**。上面这个是条件风险

我们可以通过**最小化条件风险**的行为来使得预期的损失最小化

我们通过判决函数 $\alpha(x)$ 来确定要采取 $\alpha_1 \dots \alpha_a$ 中的哪一个行动。

定义总风险R, 它是选择的判决函数决定的**预期总风险**:

$$R = \int R(\alpha(\mathbf{x})|\mathbf{x}) p(\mathbf{x}) d\mathbf{x}$$

目标是选择一个最好的判决函数, 让每一个模式的 $R(\alpha_i|x)$ 尽可能小, 这样总风险将被最小化。

对每个行为i都计算一次 $R(\alpha_i|x)$, 取最小风险的行为。最小化后的风险值就是**贝叶斯风险**。

两分类问题

对于两分类问题，行为 α_1 就是判决为 ω_1

设 $\lambda_{ij} = \lambda(\alpha_i|\omega_j)$ 也就是实际类别为 j ，判成 i 的损失。

$$R(\alpha_1|\mathbf{x}) = \lambda_{11}P(\omega_1|\mathbf{x}) + \lambda_{12}P(\omega_2|\mathbf{x})$$

$$R(\alpha_2|\mathbf{x}) = \lambda_{21}P(\omega_1|\mathbf{x}) + \lambda_{22}P(\omega_2|\mathbf{x})$$

有大量的方式来表述最小风险决策规则，基本规则就是 $R_1 < R_2$ 就采取行动1。

$$(\lambda_{21} - \lambda_{11})P(\omega_1|\mathbf{x}) > (\lambda_{12} - \lambda_{22})P(\omega_2|\mathbf{x})$$

上面的情况下就会判成 ω_1

利用贝叶斯公式可以把条件概率转换成后验概率：

$$(\lambda_{21} - \lambda_{11})p(\mathbf{x}|\omega_1)P(\omega_1) > (\lambda_{12} - \lambda_{22})p(\mathbf{x}|\omega_2)P(\omega_2)$$

合理假设 $\lambda_{21} < \lambda_{11}$ 因为错判的损失肯定比没错判要大，因此可以写成：

$$\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)}$$

这个式子也可以理解成：如果 x 在 ω_1 的出现概率大于在 ω_2 出现概率一定程度，就可以判为 ω_1 ，这个程度由两者的先验和损失决定。

最小误差分类

可以把**风险（损失）函数** $\lambda(\alpha_i|\omega_j)$ 定义如下：

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \quad i, j = 1, \dots, c$$

也就是错了有1个风险，对了0风险

因此可以化简，行动 α_1 的风险为：

$$\begin{aligned} R(\alpha_i|\mathbf{x}) &= \sum_{j=1}^c \lambda(\alpha_i|\omega_j)P(\omega_j|\mathbf{x}) \\ &= \sum_{j \neq i} P(\omega_j|\mathbf{x}) \\ &= 1 - P(\omega_i|\mathbf{x}) \end{aligned}$$

因此我们的目标是最小化行动 α_i 损失，也就是最大化 $P(\omega_i|x)$ 因此，就是判别为 $P(\omega_i|x)$ 最大的。

Neyman-Pearson准则

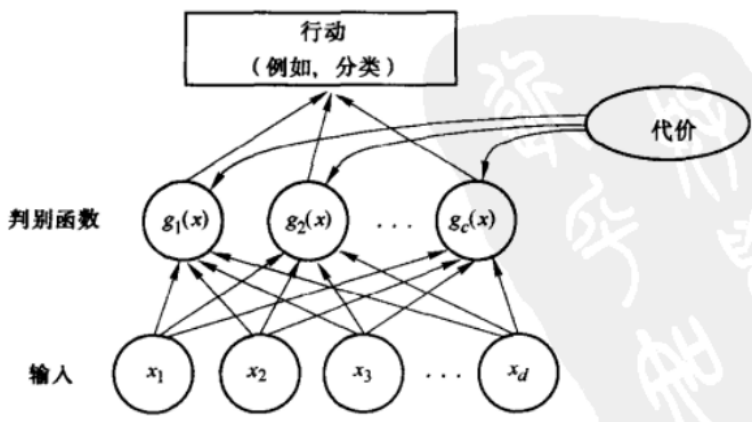
我们可能希望最小化在某个约束条件下的总风险 R ，比如对某一个特定行为 α_i 要求这个行为的风险值小于某一个常数 $\int R(\alpha_1|x)dx < \epsilon$ ，就是**不限制所有的类别误判率，但要限制某一种的误判率**。

分类器、判决函数及判定面

多类情况

若 $g_i(x) > g_j(x)$ 则分类器将特征向量 x 判决为 ω_i 类

图 2-5 一个包含 d 个输入 c 个判别函数 $g_i(\mathbf{x})$ 的一般的统计模式分类器的体系结构。接下来的步骤即是确定哪个判别函数值最大,并相应的对输入作分类。箭头表示信息流的方向,当信息流动的方向比较明显时箭头可以省略



我们可以让 $g_i(x) = -R(\alpha_i|x)$, 在最小误差概率的情况下, 我们可以让 $g_i(x) = P(\omega_i|x)$

基于这个我们可以用贝叶斯条件概率得到更多的判决函数表示方法:

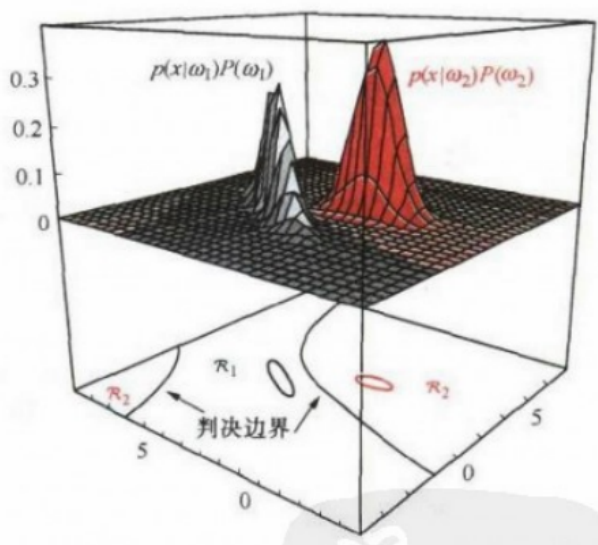
$$g_i(\mathbf{x}) = P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{\sum_{j=1}^c p(\mathbf{x}|\omega_j)P(\omega_j)}$$

$$g_i(\mathbf{x}) = p(\mathbf{x}|\omega_i)P(\omega_i)$$

$$g_i(\mathbf{x}) = \ln p(\mathbf{x}|\omega_i) + \ln P(\omega_i)$$

尽管判决函数形式不同,但判决规则相同。都是将特征空间分为 c 个判决区域。区域由判决边界分割。注意图中纵坐标的含义。

图 2-6 在这个二维的两类问题的分类器中,概率密度为高斯分布,判决边界由两个双曲线构成,因此判决区域 \mathcal{R}_2 并非简单的连通的。椭圆轮廓线标记出 $1/e$ 乘以概率密度的峰值



两类情况

两类分类器定义为 $g(\mathbf{x}) \equiv g_1(\mathbf{x}) - g_2(\mathbf{x})$

如果 $g(x)>0$ 判决为1类，否则为2类

使用上面提到的判决函数可以将 $g(x)$ 写为：

$$g(\mathbf{x}) = P(\omega_1|\mathbf{x}) - P(\omega_2|\mathbf{x})$$
$$g(\mathbf{x}) = \ln \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} + \ln \frac{P(\omega_1)}{P(\omega_2)}$$

正态密度

单变量密度函数

正态分布在左右给定均值和方差的分布中具有最大熵

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

多元密度函数

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]$$

服从正态分布的随机变量的线性变换，也是一个正态分布。

服从正态分布的随机变量的线性组合，不管这些随机变量是独立还是非独立的，也是一个正态分布。特别是，如果 $p(\mathbf{x}) \sim N(\boldsymbol{\mu}, \Sigma)$ ， \mathbf{A} 是一 $d \times k$ 的矩阵且 $\mathbf{y} = \mathbf{A}'\mathbf{x}$ 是一 k 维向量，那么 $p(\mathbf{y}) \sim N(\mathbf{A}'\boldsymbol{\mu}, \mathbf{A}'\Sigma\mathbf{A})$ ，如图 2-8 所示。在 $k=1$ 且 \mathbf{A} 是一单位向量 \mathbf{a} 的特殊情况下， $y = \mathbf{a}'\mathbf{x}$ 是一标量，表示 \mathbf{x} 到沿 \mathbf{a} 方向的一条直线的投影；此种情况下， $\mathbf{a}'\Sigma\mathbf{a}$ 是 \mathbf{x} 向 \mathbf{a} 投影的方差。那么通常，对于协方差矩阵的知识允许我们可以计算数据沿任何方向或任意子空间的分散程度。

可以将一个任意多元正态分布转换到一个球坐标系会比较方便处理，利用特征向量的原理，我们可以构造一个矩阵 \mathbf{A}_w 让变换后的正态分布协方差为单位矩阵。

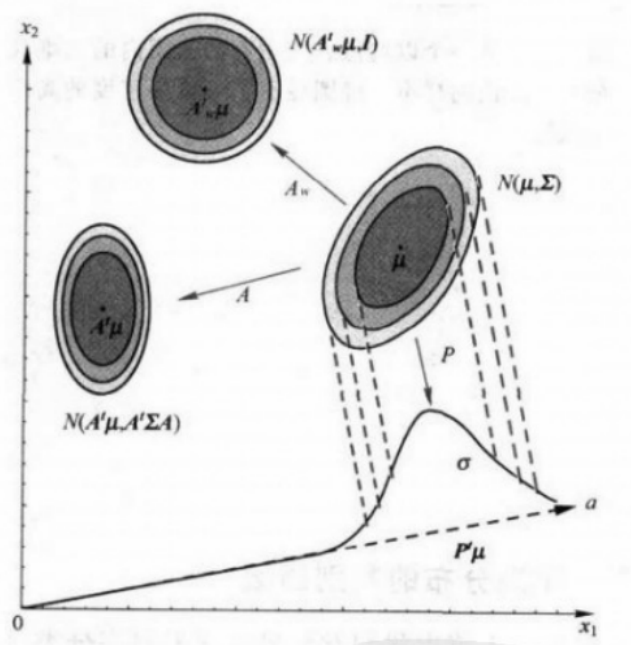
注意下图中的倒V矩阵是对角矩阵，对角矩阵的逆相当于对角线元素的-1次方。

有时将一个任意的多元正态分布的坐标转换到一个球坐标系会比较方便处理，比如，某个分布的协方差矩阵与单位矩阵 \mathbf{I} 成比例。如果定义矩阵 Φ ，其列向量是 Σ 的正交本征向量，及 Λ 为与相应本征值对应的对角矩阵，那么变换

$$\mathbf{A}_w = \Phi \Lambda^{-1/2} \quad (44)$$

将使变换后的分布的协方差矩阵成为单位阵。在信号处理中，由于此变换使转换后的分布的本征向量谱具有均匀性，因此 \mathbf{A}_w 也被称为白化变换。

图 2-8 特征空间中的一个线性变换将一个任意正态分布变换成另一个正态分布。一个变换, A , 将原分布变成分布 $N(A'\mu, A'\Sigma A)$; 另一个线性变换, 即由向量 a 决定的向某条直线的投影 P , 产生沿该直线方向的 $N(\mu, \sigma^2)$ 分布。尽管这些变换产生一个不同空间中的分布, 我们还是将它们显示在原 x_1 - x_2 空间中。一种白化变换, A_w , 将产生一个圆周对称的高斯分布



白化变化让协方差矩阵变成单位矩阵的原理是:

$$p(A_\omega^T \mathbf{x}) \sim N(A_\omega^T \mu, A_\omega^T \Sigma A_\omega)$$

因为特征向量是单位正交的, 因此乘积为单位矩阵

$$\begin{aligned} A_\omega^T \Sigma A_\omega &= (\phi \Lambda^{-1/2})^T \phi \Lambda \phi^T (\phi \Lambda^{-1/2}) \\ &= \Lambda^{-1/2} (\phi^T (\phi \Lambda \phi^T) \phi) \Lambda^{-1/2} \\ &= I \end{aligned}$$

矩阵 A_ω 可以让协方差矩阵变成单位矩阵, 如果要得到标准的多元正态分布还需要对均值进行变换, 因此我们可以减去均值得到:

$$A_\omega^T (X - \mu) \sim N(0, I)$$

马氏距离

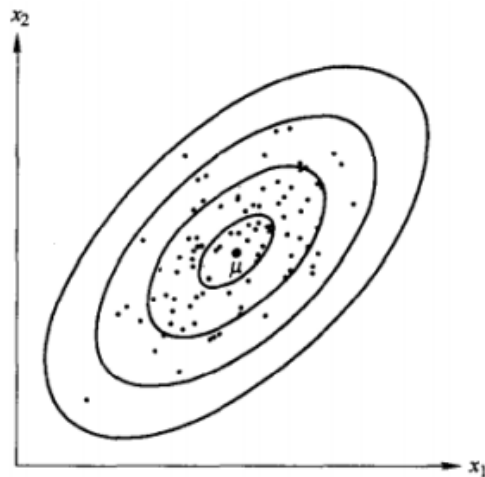
多元正态分布的中心由均值向量决定, 形状由协方差矩阵决定。分布的方向由协方差矩阵的特征向量决定, 像下图中的2个正交特征向量决定了椭圆的方向。同时特征值决定了椭圆轴的长度。在更高维度的时候就是一个超椭球体, 包含多个特征向量。

多元正态密度完全由 $d+d(d+1)/2$ 个参数——均值向量 μ 的元素及协方差矩阵 Σ 中的独立元素——来决定。从一正态分布中所抽取的样本点趋向于落在一个单一的云团或聚类中(图 2-9); 聚类中心由均值向量决定, 聚类的形状由协方差矩阵决定。由式(38)可知, 由于其二次型 $(\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu)$ 为一常量, 因此等密度的点的轨迹为一超椭球体。这些超椭球体的主轴由 Σ 的本征向量(由 Φ 表示)给出, 本征值(由 Λ 表示)决定这些轴的长度。下式

$$r^2 = (\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu) \quad (45)$$

有时被称为从 \mathbf{x} 到 μ 的平方 Mahalanobis 距离(或称为马氏距离)。因此, 等密度分布的边界是一些到 μ 的恒定马氏距离的超椭球体, 且这些超椭球体的体积决定了均值附近的样本的离散程度。

图 2-9 从一个以均值 μ 为中心的云团内的二维高斯分布中取出的样本。椭圆显示了等概率密度的高斯分布轨迹



正态分布判决函数

使用最小误差概率分类的方法设置判决函数为：

$$g_i(\mathbf{x}) = \ln p(\mathbf{x}|\omega_i) + \ln P(\omega_i)$$

如果条件概率是多元概率分布： $p(x|\omega_i) \sim N(\mu_i, \Sigma_i)$ ，由多元正态分布概率密度函数可以得：

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_i)' \Sigma_i^{-1} (\mathbf{x} - \mu_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| + \ln P(\omega_i)$$

情况1： $\Sigma_i = \sigma^2 I$

这种情况就是协方差矩阵为简单的对角矩阵，各特征统计独立，并且具有相同方差。

小的超球体聚类中的情况相对应，第 i 类的聚类以均值向量 μ_i 为中心。 Σ_i 的绝对值及其逆矩阵的计算尤其简单： $|\Sigma_i| = \sigma^{2d}$ 及 $\Sigma_i^{-1} = (1/\sigma^2)/I$ 。既然式(49)中的 $|\Sigma_i|$ 和 $(d/2) \ln 2\pi$ 都与 i 无关，因此它们是无关紧要的附加常量，可以被省略。因此我们得到简单的判别函数

$$g_i(\mathbf{x}) = -\frac{\|\mathbf{x} - \mu_i\|^2}{2\sigma^2} + \ln P(\omega_i) \quad (50)$$

我们还可以进一步化简这个式子，将二次型展开得：

$$g_i(\mathbf{x}) = -\frac{1}{2\sigma^2} [\mathbf{x}'\mathbf{x} - 2\mu_i'\mathbf{x} + \mu_i'\mu_i] + \ln P(\omega_i)$$

它看上去像是 \mathbf{x} 的一个二次函数，但是，二次项 $\mathbf{x}'\mathbf{x}$ 对于所有 i 是相等的，使得它成为一个可省略的附加常量，因此，我们得到了等价的线性判别函数，

$$g_i(\mathbf{x}) = \mathbf{w}_i'\mathbf{x} + w_{i0} \quad (53)$$

其中

$$\mathbf{w}_i = \frac{1}{\sigma^2} \mu_i \quad (54)$$

且

$$w_{i0} = \frac{-1}{2\sigma^2} \mu_i'\mu_i + \ln P(\omega_i) \quad (55)$$

我们称 w_{i0} 为第 i 个方向的阈值或偏置。

于是判决函数就变成了一个线性分类器，我们可以利用第五章的知识，计算两个判决函数的判定面，也就是 $g_i(x) = g_j(x)$ 的地方：

$$\mathbf{w}'(\mathbf{x} - \mathbf{x}_0) = 0 \quad (56)$$

其中

$$\mathbf{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j \quad (57)$$

且

$$\mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\sigma^2}{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2} \ln \frac{P(\omega_i)}{P(\omega_j)} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) \quad (58)$$

这个方程定义了一个经过 \mathbf{x}_0 的与向量 \mathbf{w} 垂直的超平面。

如果先验相同，则58式右边第二项为0，则过的 \mathbf{x}_0 就是均值连线的中心。此时就可以简单的计算特征向量 \mathbf{x} 到 c 个均值向量的直线距离，离哪个近就去哪一类

情况2: $\Sigma_i = \Sigma$

第二种情况是所有类的协方差矩阵都相同，但此时每类中的属性不是无关的了。这种情况对应于样本落在相同大小和形状的超椭球体聚类中。

$\boldsymbol{\mu}_i$ 附近。由于式(49)中 $|\Sigma_i|$ 和 $(d/2)\ln 2\pi$ 两项与 i 无关，它们可作为多余的附加常量而被省略。这种简化可推出判别函数为

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) + \ln P(\omega_i) \quad (59)$$

如果先验相同，问题又被简化为，对每个特征向量 \mathbf{x} ，计算和 c 个均值向量的马氏距离。如果先验不同，判定面会偏向先验小的，这样更多的 \mathbf{x} 会被判决为先验大的类

同样继续化简，将马氏距离的公式展开：

将二次型 $(\mathbf{x} - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)$ 展开可得一与 i 无关的二次项 $\mathbf{x}' \boldsymbol{\Sigma}^{-1} \mathbf{x}$ ，将式(59)中的此项去掉后可再次得到线性判别函数：

$$g_i(\mathbf{x}) = \mathbf{w}_i' \mathbf{x} + w_{i0} \quad (60)$$

其中

$$\mathbf{w}_i = \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i \quad (61)$$

且

$$w_{i0} = -\frac{1}{2} \boldsymbol{\mu}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i + \ln P(\omega_i) \quad (62)$$

同样可以计算超平面 $g_i(\mathbf{x}) = g_j(\mathbf{x})$ ：

$$\mathbf{w}'(\mathbf{x} - \mathbf{x}_0) = 0$$

其中

$$\mathbf{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

且

$$\mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\ln[P(\omega_i)/P(\omega_j)]}{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

由于通常 $\mathbf{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$ 并非朝着 $\boldsymbol{\mu}_i - \boldsymbol{\mu}_j$ 的方向，因而通常分离 \mathcal{R}_i 和 \mathcal{R}_j 的超平面也并非与均值间的连线垂直正交。但是，如果先验概率相等，其判定面确实是与均值连线交于其中点 \mathbf{x}_0 处的。如果先验概率不等，最优边界超平面将远离可能性较大的均值(图 2-12)。同前，如果偏移量足够大，判定面可以不落在两个均值向量之间。

情况3: $\Sigma_i = \text{任意}$

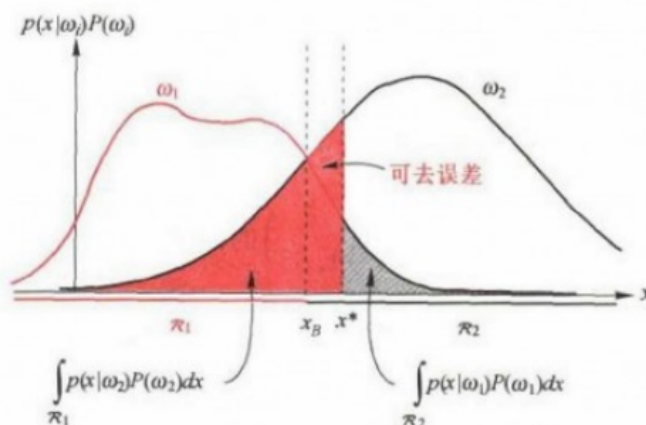
因为此时的协方差矩阵不是都相同了，因此对49式只有一项可以忽略。且其判别函数显然也是二次型。

误差概率和误差积分

在两类问题下计算错误概率为：

$$\begin{aligned} P(\text{error}) &= P(\mathbf{x} \in \mathcal{R}_2, \omega_1) + P(\mathbf{x} \in \mathcal{R}_1, \omega_2) \\ &= P(\mathbf{x} \in \mathcal{R}_2 | \omega_1) P(\omega_1) + P(\mathbf{x} \in \mathcal{R}_1 | \omega_2) P(\omega_2) \\ &= \int_{\mathcal{R}_2} p(\mathbf{x} | \omega_1) P(\omega_1) d\mathbf{x} + \int_{\mathcal{R}_1} p(\mathbf{x} | \omega_2) P(\omega_2) d\mathbf{x} \end{aligned}$$

图 2-17 相等先验概率情况下的误差概率组成以及(非最优)判决点 x^* 。粉红区域对应于实际类别为 ω_2 而判为 ω_1 的误差概率，灰色区域相反，如式(70)。如果判决边界在相等后验概率点 x_B 处，那么此“可去误差”区将消失，总的有阴影的区域将减到最小——这就是贝叶斯判决导致的贝叶斯误差率



在多类问题下，计算正确分类的概率更简单：

$$\begin{aligned} P(\text{correct}) &= \sum_{i=1}^c P(\mathbf{x} \in \mathcal{R}_i, \omega_i) \\ &= \sum_{i=1}^c P(\mathbf{x} \in \mathcal{R}_i | \omega_i) P(\omega_i) \\ &= \sum_{i=1}^c \int_{\mathcal{R}_i} p(\mathbf{x} | \omega_i) P(\omega_i) d\mathbf{x} \end{aligned}$$

正态密度的误差上界

贝叶斯判决规则确保了最低的误差概率，并且知道如何计算判决边界。然而，这些结论并没有告诉我们实际的误差概率是多少。高斯情况下的整个误差率计算过程相当复杂，尤其是高维情况，这是因为式(71)的积分范围中的判决区域不连续。但是，两类情况下式(5)的一般误差积分公式可近似的给出一个误差率的上界(upper bound)。

Chernoff 界

为获得误差的界(bound),需利用下列不等式

$$\min[a, b] \leq a^\beta b^{1-\beta} \quad (a, b \geq 0 \text{ 且 } 0 \leq \beta \leq 1) \quad (72)$$

可以得到二分类问题误差的公式为:

$$P(\text{error}) \leq P^\beta(\omega_1) P^{1-\beta}(\omega_2) \int p^\beta(\mathbf{x}|\omega_1) p^{1-\beta}(\mathbf{x}|\omega_2) d\mathbf{x} \quad (0 \leq \beta \leq 1)$$

如果条件概率是正态的,式(73)的积分结果可用解析法计算出(习题 36),

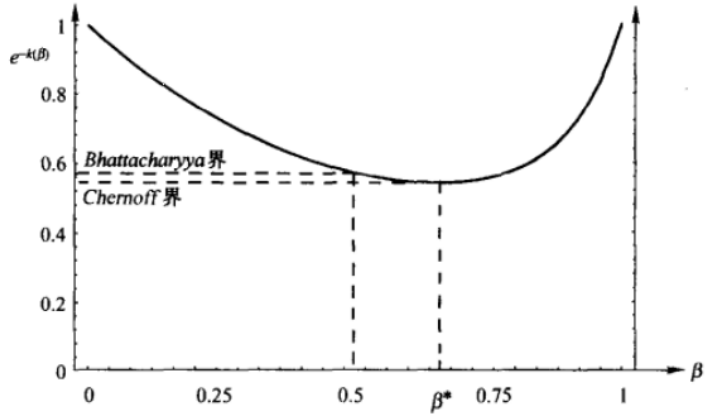
$$\int p^\beta(\mathbf{x}|\omega_1) p^{1-\beta}(\mathbf{x}|\omega_2) d\mathbf{x} = e^{-k(\beta)} \quad (74)$$

其中

$$k(\beta) = \frac{\beta(1-\beta)}{2} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)' [\beta \boldsymbol{\Sigma}_1 + (1-\beta) \boldsymbol{\Sigma}_2]^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) + \frac{1}{2} \ln \frac{|\beta \boldsymbol{\Sigma}_1 + (1-\beta) \boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|^\beta |\boldsymbol{\Sigma}_2|^{1-\beta}} \quad (75)$$

图 2-18 的例子显示了 $e^{-k(\beta)}$ 如何随 β 的变化而变化的典型情形。 $P(\text{error})$ 的 Chernoff 界可通过数值分析或直接查找使 $e^{-k(\beta)}$ 最小的 β 值求出,然后把这个 β 代入式(73)。这里很关键的一点是此优化过程是在一维 β 空间中进行的,尽管分布本身可能位于任意高维的空间。

图 2-18 Chernoff 界不会比 Bhattacharyya 界松弛。此例中,Chernoff 界取在 $\beta^* = 0.66$ 处,比 Bhattacharyya 界 ($\beta = 0.5$) 稍微紧致一些



Bhattacharyya界

图 2-18 所示的 Chernoff 界对 β 的依赖性在很多问题中都普遍存在,即对于处于极值(即 $\beta \rightarrow 1$ 和 $\beta \rightarrow 0$)处的界较松弛,而中间较紧致。优化后的 β 的精确值取决于分布参数和先验概率,一种计算较简单但稍微松一点的界可以通过仅仅取 $\beta = 1/2$ 处的值获得。这就给出所谓的 Bhattacharyya 误差的界,于是式(73)的形式为

$$\begin{aligned} P(\text{error}) &\leq \sqrt{P(\omega_1)P(\omega_2)} \int \sqrt{p(\mathbf{x}|\omega_1)p(\mathbf{x}|\omega_2)} d\mathbf{x} \\ &= \sqrt{P(\omega_1)P(\omega_2)} e^{-k(1/2)} \end{aligned} \quad (76)$$

其中通过式(75)可得高斯分布的情况如下:

$$k(1/2) = 1/8 (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)' \left[\frac{\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2}{2} \right]^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) + \frac{1}{2} \ln \frac{\left| \frac{\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2}{2} \right|}{\sqrt{|\boldsymbol{\Sigma}_1||\boldsymbol{\Sigma}_2|}} \quad (77)$$