自动化科学与电气工程学院
School of Automation Science and Electrical Engineering

北京航空航天大學
BEIHANG UNIVERSITY

贝叶斯分类器设计

院（系）名称 自动化科学与电气工程学院

专业名称 模式识别与智能系统

学生学号 15031184

学生姓名 李思奇

2018 年 5 月 2 日

1. 实验原理

1.1 实验背景

贝叶斯分类器是一种从一定的概率模型出发，根据各类特征的概率模型来估算后验概率并做出决策的分类器。实质是将模式识别的问题转化成了概率模型估计的问题。在许多场合，贝叶斯分类器可以与决策树和神经网络分类算法相媲美，该算法能运用到大型数据库中，而且方法简单、分类准确率高、速度快。

1.2 理论推导

根据概率论中的贝叶斯公式，我们知道：

$$P(\omega_i \mid x)=\frac{p(x,\omega_i)}{p(x)}=\frac{p(x\mid\omega_i)P(\omega_i)}{p(x)}, i=1,2 \tag{1-1}$$

其中，$P(\omega_i)$ 为先验概率，$p(x\mid\omega_i)$ 是 x 与 ω_i 的联合概率密度函数，$p(x)$ 是两类所有样本的概率密度，$p(x\mid\omega_i)$ 是第 i 类样本的概率密度（类条件密度）。

对于最小错误率贝叶斯决策，我们希望尽量减少分类中的错误，即目标是追求最小错误率。根据定义，最小错误率决策是求解一种决策规则，使得错误率最小化，即：

$$\min P(e)=\int P(e\mid\mathbf{x})p(\mathbf{x})d\mathbf{x} \tag{1-2}$$

由于对于所有的 $\mathbf{x}$，有 $P(e\mid\mathbf{x})\ge 0, p(\mathbf{x})\ge 0$，所以上式等价于对于所有 $\mathbf{x}$ 最小化 $P(e\mid\mathbf{x})$。因此，对于两类问题，得到最小错误率决策如下：

如果 $P(\omega_1\mid\mathbf{x})>P(\omega_2\mid\mathbf{x})$，则 $\mathbf{x}\in\omega_1$，反之，则 $\mathbf{x}\in\omega_2$

若引入损失函数 $\lambda(\alpha_i,\omega_j)$，则可得出最小风险贝叶斯决策。设损失函数 $\lambda(\alpha_i,\omega_j)$ 为样本应属于 ω_j 类而被判为 α_i 时引起的损失。则对于某个样本 $\mathbf{x}$，它属于各个状态的后验概率为 $P(\omega_j\mid\mathbf{x}), j=1,2,...,c$，对它采取决策 $\alpha_i, i=1,2,...,k$ 的期望损失是：

$$R(\alpha_i\mid\mathbf{x})=E\left[\lambda(\alpha_i,\omega_j)\mid\mathbf{x}\right]=\sum_{j=1}^{c}\lambda(\alpha_i,\omega_j)P(\omega_j\mid\mathbf{x}), i=1,2,...,k \tag{1-3}$$

则，最小风险贝叶斯决策为：

$$\alpha_k=\mathrm{Arg}\min_i R(\alpha_i\mid\mathbf{x}), i=1,2,...,c \tag{1-4}$$

1.3 决策步骤

①利用贝叶斯公式计算后验概率

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j) P(\omega_j)}{\sum_{i=1}^{c} p(x \mid \omega_i) P(\omega_i)}, j = 1, 2, \ldots, c \tag{1-5}$$

②利用损失函数，计算条件风险

$$R(\alpha_i \mid \mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i, \omega_j) P(\omega_j \mid \mathbf{x}), i = 1, 2, \ldots, k \tag{1-6}$$

③决策：在各种决策中选择风险最小的决策，即

$$\alpha = \arg\min_{i=1,\ldots,k} R(\alpha_i \mid \mathbf{x}) \tag{1-7}$$

2. 基础实验

2.1 类条件概率计算

给定两类样本集：

$$\Omega_1 = \{-3.9847, -3.5549, -1.2401, -0.9780, -0.7932, -2.8531, -2.7605, -3.7287, -3.5414, -2.2692, -3.4549, -3.0752, -3.9934, -0.9780, -1.5799, -1.4885, -0.7431, -0.4221, -1.1186, -2.3462, -1.0826, -3.4196, -1.3193, -0.8367, -0.6579, -2.9683\}$$

$$\Omega_2 = \{2.8792, 0.7932, 1.1882, 3.0682, 4.2532, 0.3271, 0.9846, 2.7648, 2.6588\}$$

已知两组数据均满足正态分布，根据概率论，应有：

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{2-1}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} {x_i}^2 - \left(\frac{1}{n} \sum_{i=1}^{n} x_i\right)^2 \tag{2-2}$$

根据式(2-1)和式(2-2)，求出对应的 $\hat{\mu}$ 和 $\hat{\sigma}$，并做出类条件概率密度图如图 2.1 所示。

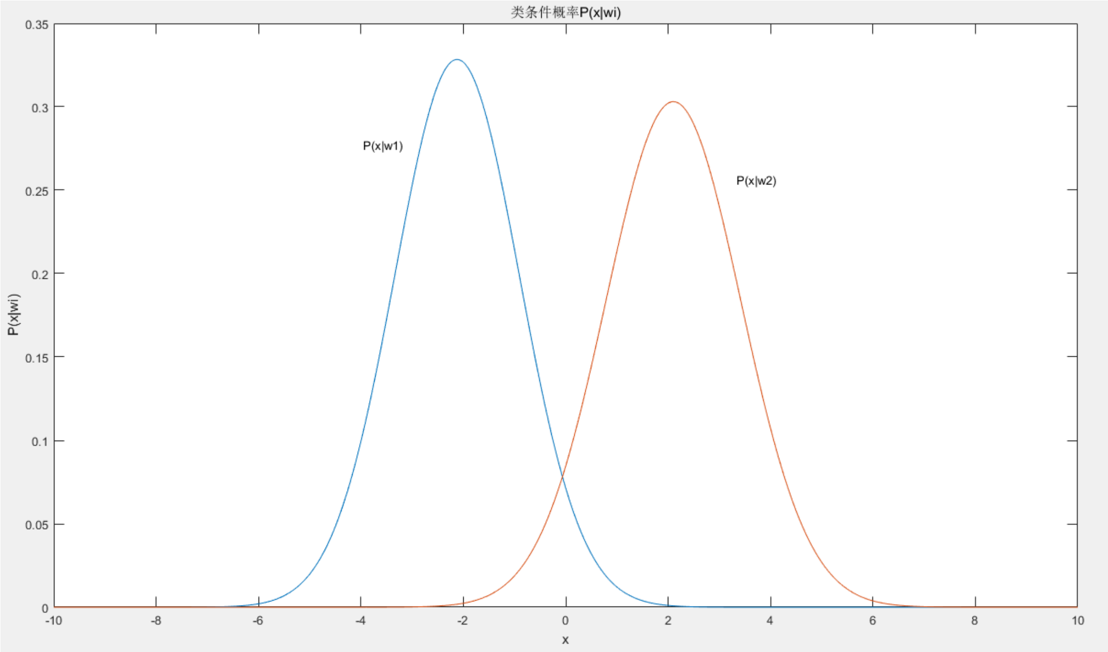

图 2.1 类条件概率密度

2.2 后验概率

根据式(1-5)，可求出两类的类条件概率密度函数，如图 2.2.1 所示。

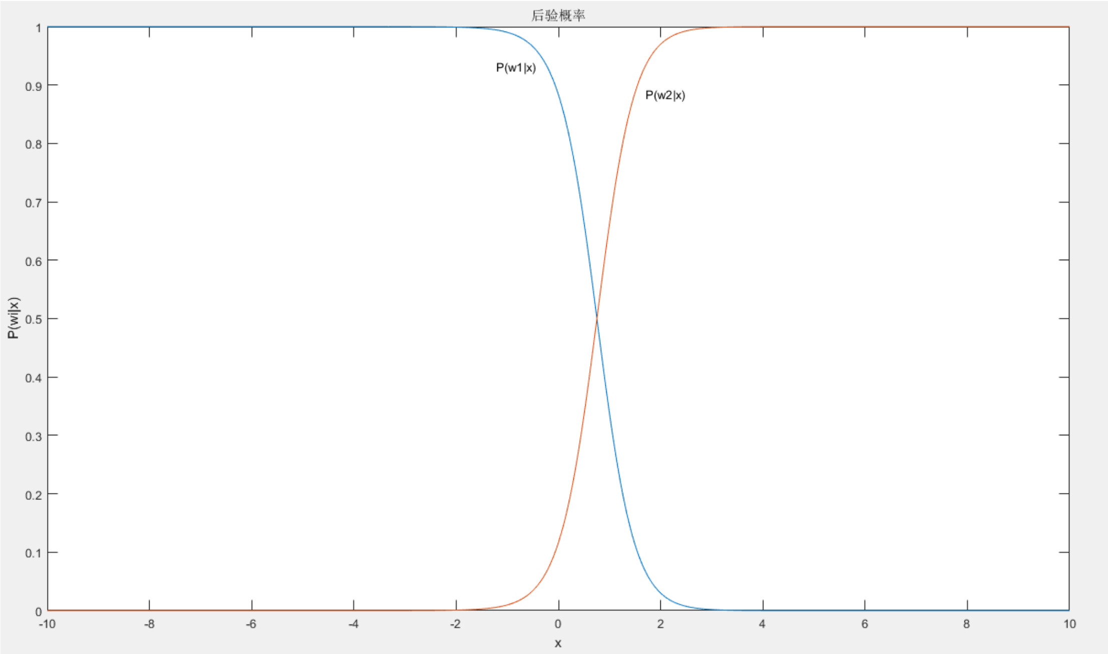

图 2.2.1 后验概率

若不考虑损失函数，即采用最小错误率贝叶斯决策，可求出决策边界方程为：

$$p(x|\omega_1)p(\omega_1)=p(x|\omega_2)p(\omega_2)$$

即：

$$\frac{0.9}{\sqrt{2\pi}\sigma_1}\exp(-\frac{(x-\mu_1)^2}{2\sigma_1{}^2})=\frac{0.1}{\sqrt{2\pi}\sigma_2}\exp(-\frac{(x-\mu_2)^2}{2\sigma_2{}^2})$$

做出最小错误率贝叶斯决策边界如图 2.2.2 所示。

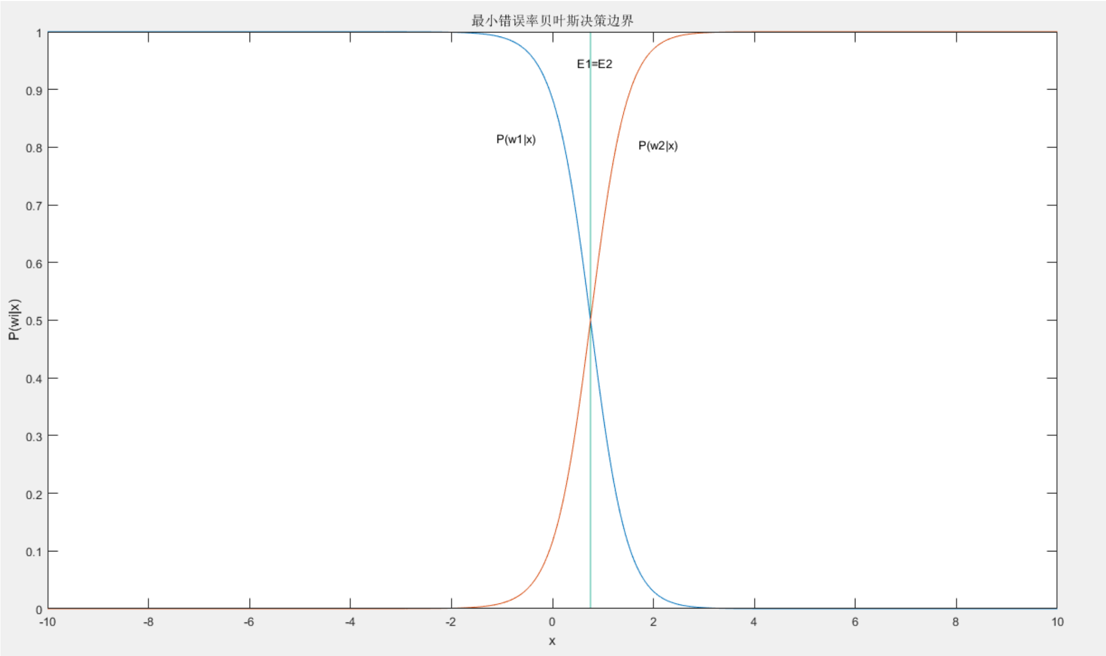

图 2.2.2 最小错误率贝叶斯决策边界

2.3 最小风险贝叶斯决策

由式(1-6)、式(1-7)可求出当 $R(\alpha_1|\mathbf{x})$=$R(\alpha_2|\mathbf{x})$ 时的 x 取值。决策边界方程为：

$$\frac{0.9\times 6}{\sqrt{2\pi}\sigma_1}\exp(-\frac{(x-\mu_1)^2}{2\sigma_1{}^2})=\frac{0.1}{\sqrt{2\pi}\sigma_2}\exp(-\frac{(x-\mu_2)^2}{2\sigma_2{}^2})$$

做出最小风险贝叶斯决策的决策边界如图 2.3 所示。

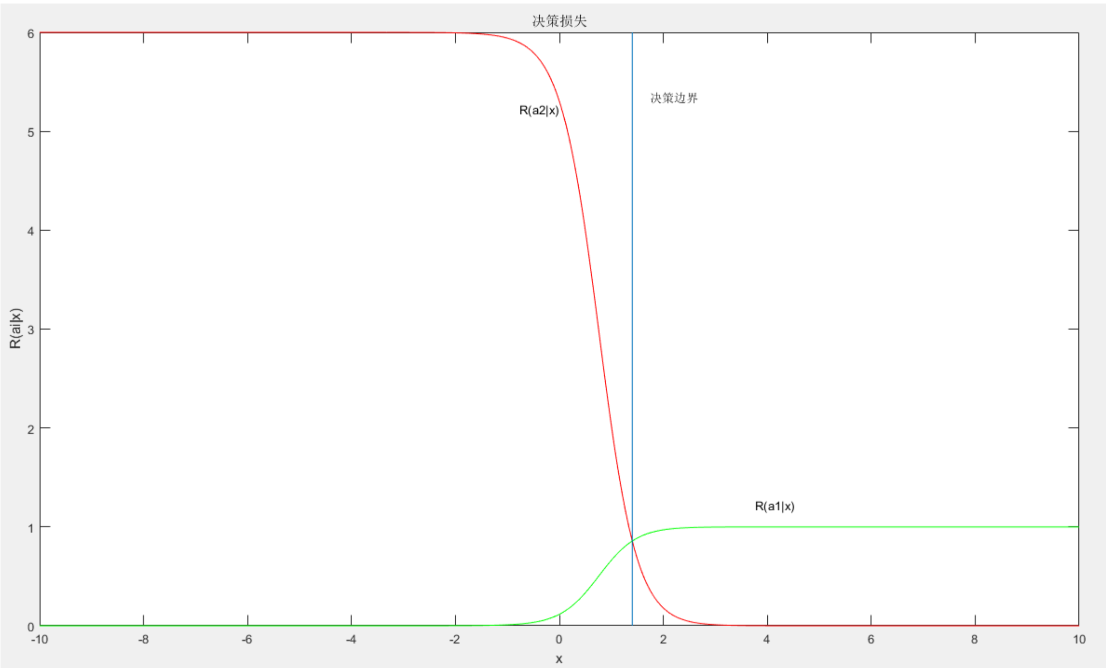

图 2.3 最小风险贝叶斯决策边界

对比图 2.2.2 与图 2.3 可以发现，最小风险贝叶斯决策的决策边界比最小错误率的决策边界的 x 值要更大。这是因为引入损失函数后，将ω_1类的样本错分为ω_2类时，要付出更大的代价。因此，即使ω_2类的后验概率要略大于ω_1类的后验概率，也会将样本判定为ω_1类。

3. 扩大类别数和特征维度

3.1 类条件概率

扩大类别数目及特征维度。为便于可视化，构造样本集，类别数为 3 类、样本数为 600 例（每类 200 例样本）、样本特征维度为 2 维。样本集中样本分布如图 3.1.1 所示。

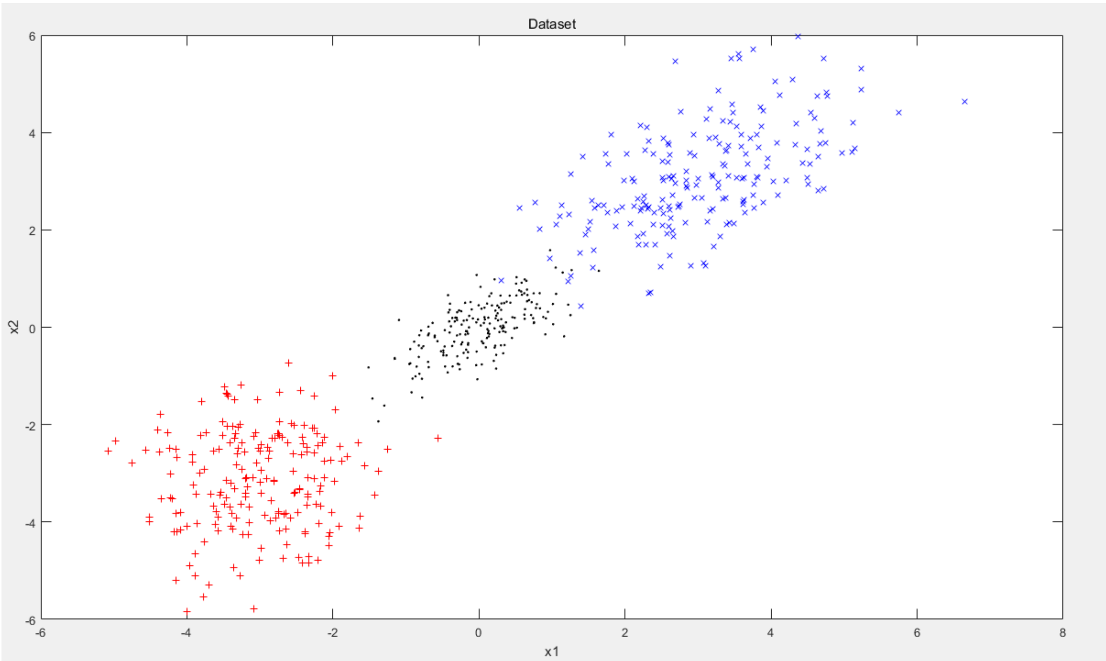

图 3.1.1 样本分布

考虑多元正态分布的数学模型。其概率密度函数为：

$$p(\mathbf{x})=\frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \tag{3-1}$$

式中：$\boldsymbol{\mu}=\left[\mu_1,\mu_2,...\mu_d\right]$是 d 维均指向量。$\mathbf{\Sigma}$是$d\times d$维协方差矩阵。$\mathbf{\Sigma}^{-1}$是$\mathbf{\Sigma}$的逆矩阵，$|\mathbf{\Sigma}|$是$\mathbf{\Sigma}$的行列式。

根据数学推导，有：

$$\boldsymbol{\mu}=E\left[\mathbf{x}\right] \tag{3-2}$$

$$\mathbf{\Sigma}=E\left[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\right] \tag{3-3}$$

根据式(3-2)和式(3-3)，可以求出三类样本各自的$\boldsymbol{\mu}$、$\boldsymbol{\Sigma}$：并做出各自的类条件概率如图 3.1.2 所示。

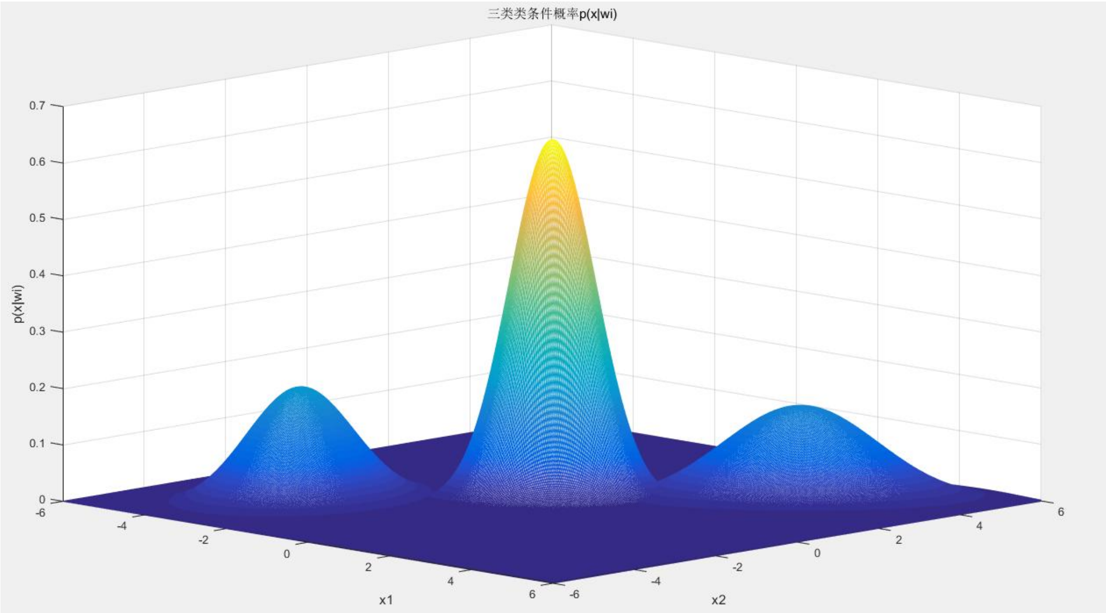

图 3.1.2 三类类条件概率

其中：

$$\boldsymbol{\mu}_1=\begin{bmatrix}3.0227\\3.0736\end{bmatrix},\boldsymbol{\mu}_2=\begin{bmatrix}0.0173\\-0.0065\end{bmatrix},\boldsymbol{\mu}_3=\begin{bmatrix}-3.0304\\-3.1449\end{bmatrix}$$

$$\boldsymbol{\Sigma}_1=\begin{bmatrix}1.1718&0.6689\\0.6689&1.1392\end{bmatrix},\boldsymbol{\Sigma}_2=\begin{bmatrix}0.3435&0.2264\\0.2264&0.3283\end{bmatrix},\boldsymbol{\Sigma}_3=\begin{bmatrix}0.6088&0.0569\\0.0569&1.0046\end{bmatrix}$$

且有第 i 类类条件概率为：

$$p(\mathbf{x}\mid\omega_i)=\frac{1}{(2\pi)^{3/2}\left|\boldsymbol{\Sigma}_i\right|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right\},i=1,2,3$$

3.2 后验概率

假设三类的先验概率为：

$$p(\omega_1)=0.35,p(\omega_2)=0.15,\ \ p(\omega_3)=0.5$$

则由式(1-5)可求出第 i 类的后验概率为：

$$P(\omega_i\mid x)=\frac{p(x\mid\omega_i)P(\omega_i)}{\sum_{j=1}^{3}p(x\mid\omega_j)P(\omega_j)},i=1,2,3$$

做出各类后验概率如图 3.2.1 所示。

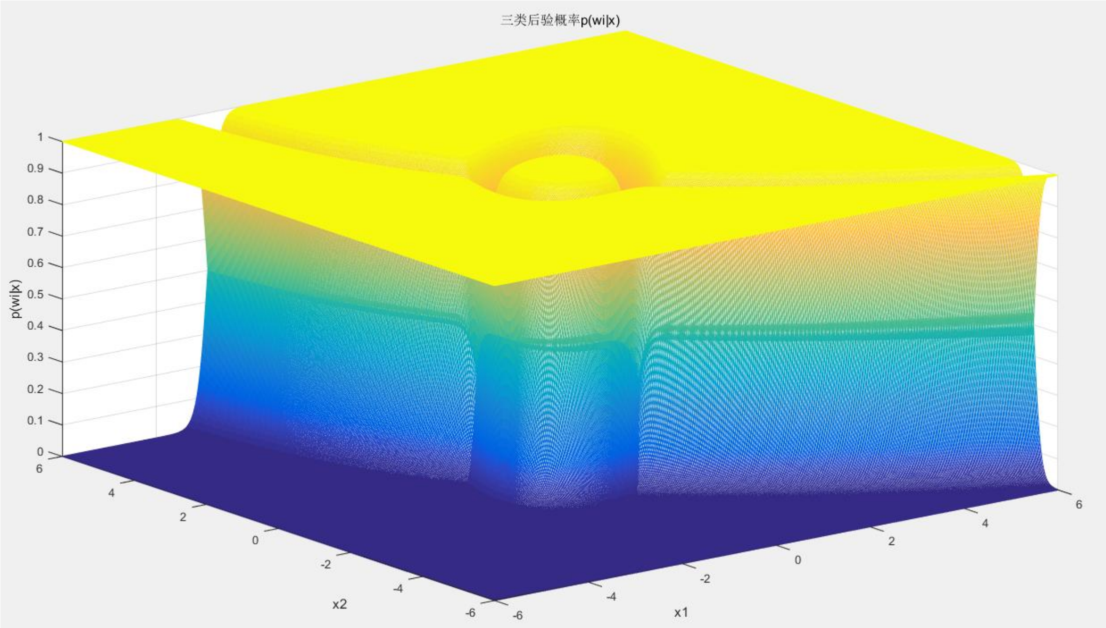

图 3.2.1 三类后验概率

3.3 最小风险贝叶斯决策

设损失函数 $\lambda(\alpha_i,\omega_j)$ 的决策表如表 3.3.1 所示。

表 3.3.1 损失函数 $\lambda(\alpha_i,\omega_j)$ 的决策表

决策	自然状态		
	ω_1	ω_2	ω_3
α_1	0	2	18
α_2	3	0	3
α_3	8	6	0

则，对于某个样本 **x**，对它采取决策 α_1，α_2，α_3 的期望损失为：

$$\begin{cases} R(\alpha_1|\mathbf{x})=\lambda(\alpha_1,\omega_2)P(\omega_2|\mathbf{x})+\lambda(\alpha_1,\omega_3)P(\omega_3|\mathbf{x})=6P(\omega_2|\mathbf{x})+2P(\omega_3|\mathbf{x}) \\ R(\alpha_2|\mathbf{x})=\lambda(\alpha_2,\omega_1)P(\omega_1|\mathbf{x})+\lambda(\alpha_2,\omega_3)P(\omega_3|\mathbf{x})=3P(\omega_1|\mathbf{x})+3P(\omega_3|\mathbf{x}) \\ R(\alpha_3|\mathbf{x})=\lambda(\alpha_3,\omega_1)P(\omega_1|\mathbf{x})+\lambda(\alpha_3,\omega_2)P(\omega_2|\mathbf{x})=2P(\omega_1|\mathbf{x})+6P(\omega_2|\mathbf{x}) \end{cases} \quad (3\text{-}4)$$

根据式(3-4)做出决策损失图，如图 3.3.1 所示。

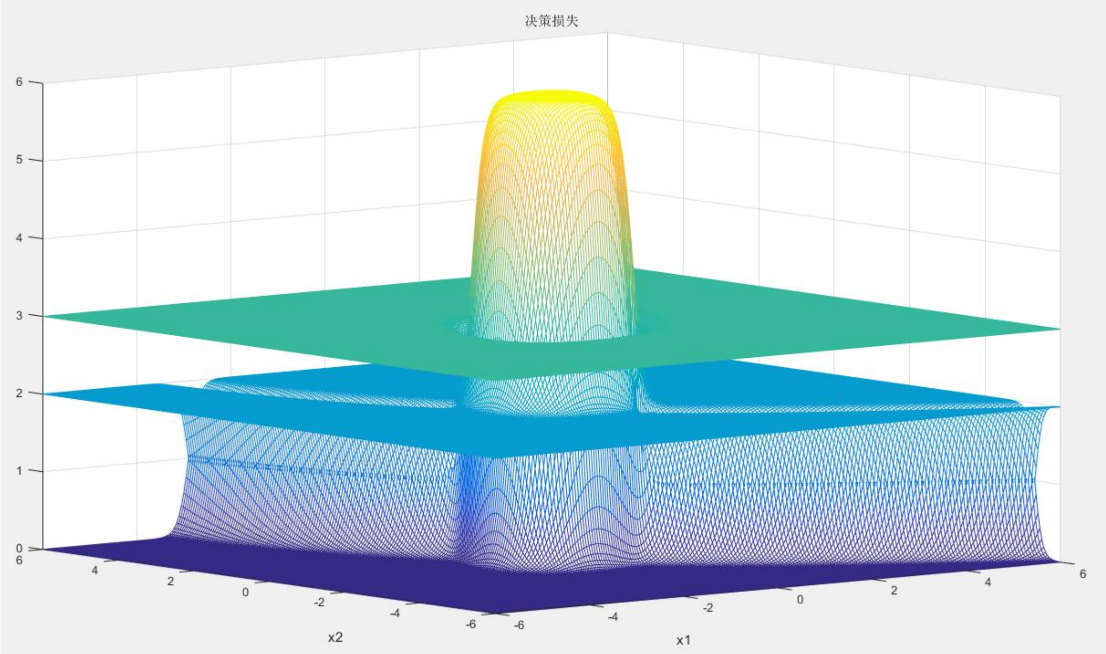

图 3.3.1 决策损失

通过比较三种决策的期望损失，可以得出分类决策：

$$\alpha = \arg\min_{i=1,2,3} R(\alpha_i \mid \mathbf{x}) \tag{3-5}$$

3.4 最小错误率决策边界与最小风险决策边界

由图 3.2.2 以及最小错误率决策：

$$\alpha = \arg\max_{i=1,2,3} p(\omega_i \mid \mathbf{x})$$

可以得出，最小错误率决策的决策边界方程为：

$$\begin{cases} p(x \mid \omega_1)p(\omega_1) = p(x \mid \omega_2)p(\omega_2) \ldots\ldots\ldots\ldots\text{第1, 2类决策边界} \\ p(x \mid \omega_1)p(\omega_1) = p(x \mid \omega_3)p(\omega_3) \ldots\ldots\ldots\ldots\text{第1, 3类决策边界} \\ p(x \mid \omega_2)p(\omega_2) = p(x \mid \omega_3)p(\omega_3) \ldots\ldots\ldots\ldots\text{第2, 3类决策边界} \end{cases}$$

做出决策边界曲线，即是图 3.2.2 中各类后验概率相交的曲线，如图 3.4.1 所示。

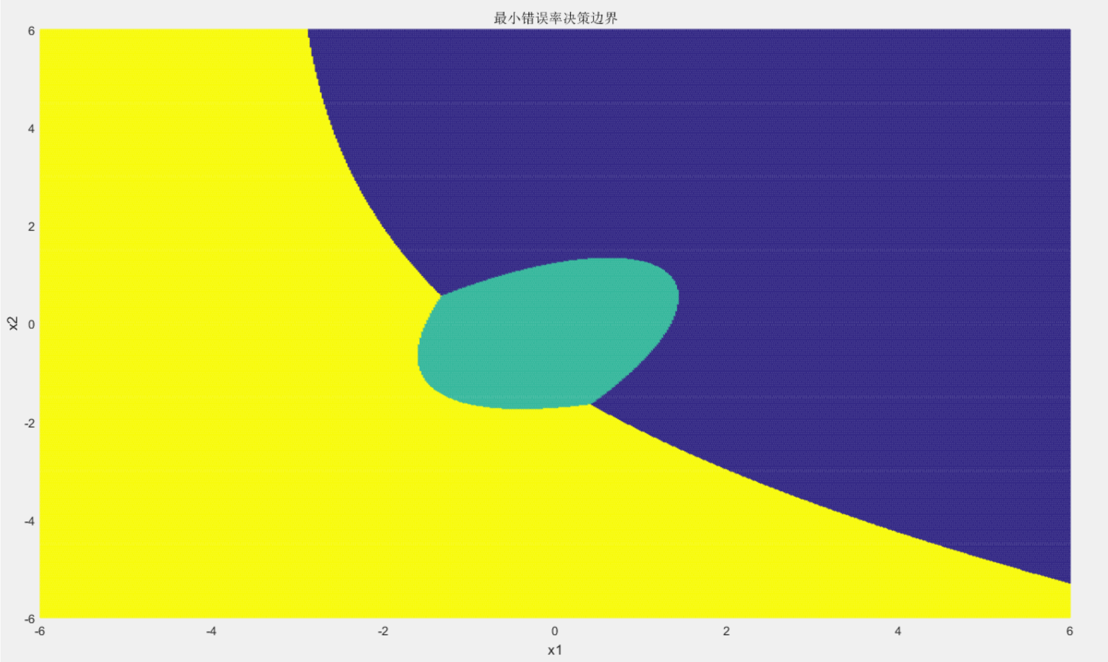

图 3.4.1 最小错误率决策边界

由式(3-4)和式(3-5)可以得出最小风险贝叶斯决策的决策边界方程为:

$$\begin{cases} 6P(\omega_2\,|\,\mathbf{x})+2P(\omega_3\,|\,\mathbf{x})=3P(\omega_1\,|\,\mathbf{x})+3P(\omega_3\,|\,\mathbf{x}) \text{..............................第1, 2类决策边界} \\ 6P(\omega_2\,|\,\mathbf{x})+2P(\omega_3\,|\,\mathbf{x})=2P(\omega_1\,|\,\mathbf{x})+6P(\omega_2\,|\,\mathbf{x}) \text{..............................第1, 3类决策边界} \\ 3P(\omega_1\,|\,\mathbf{x})+3P(\omega_3\,|\,\mathbf{x})=2P(\omega_1\,|\,\mathbf{x})+6P(\omega_2\,|\,\mathbf{x}) \text{..............................第2, 3类决策边界} \end{cases}$$

做出决策边界曲线，如图 3.4.3 所示。

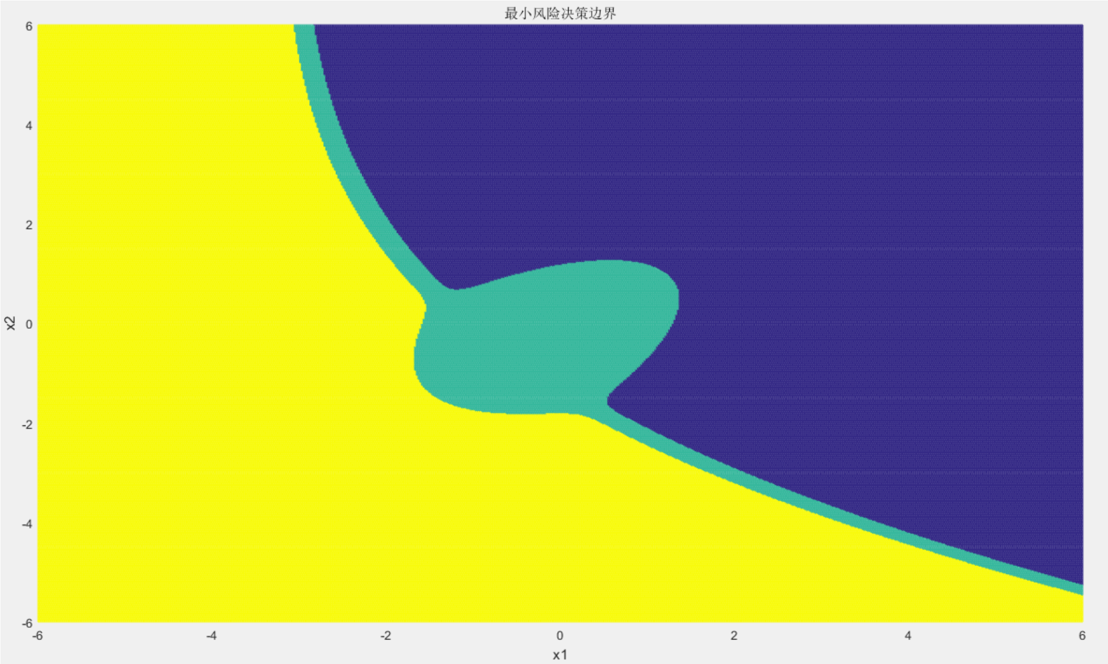

图 3.4.3 最小风险决策边界

可以看出，加入损失函数 $\lambda(\alpha_i,\omega_j)$ 之后，利用最小风险贝叶斯决策做出的决策边界略有变化。

4. 结论分析

由以上实验可以看出，一维与多维、一类与多类的贝叶斯分类器在本质原理上并没有太大差别。当维度很高时，可能设计贝叶斯分类器会变得极为麻烦。

相比于线性分类器，贝叶斯分类器在处理多类问题的时候会显得更加有用。但是当样本的概率模型不确定或不是很清楚的时候，应当谨慎使用或避免使用贝叶斯分类器。

附录：MATLAB 程序代码

```
clc;clear;n=200;
X1 = load('data1.txt');
X2 = load('data2.txt');
X3 = load('data3.txt');
figure(1);plot(X1(:,1),X1(:,2),'bx',X2(:,1),X2(:,2),'k.',X3(:,1),X3(:,2),'r+');
xlabel('x1');ylabel('x2');title('Dataset');
X1=X1';X2=X2';X3=X3';
[mu1,sigma1]=get_mu_sigma(X1,n);
[mu2,sigma2]=get_mu_sigma(X2,n);
[mu3,sigma3]=get_mu_sigma(X3,n);
p1=0.35;p2=0.15;p3=0.5;
[x,y]=meshgrid(-6:0.02:6,-6:0.02:6);
P1 = pxy(x,y,mu1,sigma1);
P2 = pxy(x,y,mu2,sigma2);
P3 = pxy(x,y,mu3,sigma3);
figure(2);mesh(x,y,P1);hold on;mesh(x,y,P2);hold on;mesh(x,y,P3);
xlabel('x1');ylabel('x2');zlabel('p(x|wi)');
title('三类类条件概率p(x|wi)');
P1=P1*p1;P2=P2*p2;P3=P3*p3;Pall=P1+P2+P3;
P1=P1./Pall; P2=P2./Pall; P3=P3./Pall;
figure(3);mesh(x,y,P1);hold on;mesh(x,y,P2);hold on;mesh(x,y,P3);hold on;
xlabel('x1');ylabel('x2');zlabel('p(wi|x)');title('三类后验概率p(wi|x)');
region = cregion(x, P1, P2, P3, '1');
figure(4);mesh(x, y, region);
xlabel('x1');ylabel('x2');title('最小错误率决策边界');
r1 = 2*P2 + 8*P3; r2 = 3*P1 + 3*P3; r3 = 6*P2 + 18*P1;
figure(5);mesh(x, y, r1);hold on;mesh(x, y, r2);mesh(x, y, r3);
xlabel('x1');ylabel('x2');title('决策损失');
region = cregion(x, r1, r2, r3, '2');
figure(6);mesh(x, y, region);
title('最小风险决策边界');xlabel('x1');ylabel('x2');
```

```
function [mu,sigma]=get_mu_sigma(X,n)
mu=0;
for i=1:n
    mu = mu + X(:,i);
end
mu = mu ./ n;
sigma = [0 0;0 0];
for i=1:n
   sigma = sigma + (X(:,i)-mu)*(X(:,i)-mu)' ;
end
sigma = sigma ./ n;
end

function Z = pxy(X,Y,mu,sigma)
DX=sigma(1,1);       dx=sqrt(DX);
DY=sigma(2,2);       dy=sqrt(DY);
COV=sigma(1,2);      r=COV/(dx*dy);
part1=1/(2*pi*dx*dy*sqrt(1-r^2));
p1=-1/(2*(1-r^2));
px=(X-mu(1)).^2./DX;
py=(Y-mu(2)).^2./DY;
pxy=2*r.*(X-mu(1)).*(Y-mu(2))./(dx*dy);
Z=part1*exp(p1*(px-pxy+py));
end

function region = cregion(x, p1, p2, p3, op)
assert(nargin == 5);
region = zeros(size(p1));
for i = 1 : size(x, 1)
    for j = 1 : size(x, 2)
        if(strcmp(op, '1'))
            [~, loc] = max([p1(i, j), p2(i, j), p3(i, j)]);
        elseif(strcmp(op, '2'))
            [~, loc] = min([p1(i, j), p2(i, j), p3(i, j)]);
        end
        region(i, j) = loc - 1;
    end
end
end
```