# 第3章 统计模式识别中的概率方法

## 3.1 用概率方法描述分类问题

略

## 3.2 几个相关的概念

贝叶斯公式： $P(\omega_j|X)=\frac{p(X|\omega_j)P(\omega_j)}{P(X)}$

## 3.3 最小错误概率判决准则

选择相交的那个

## 3.4 最小风险判决规则

$$R(\alpha_i|X)=\min_{j=1,\cdots,A}\{R(\alpha_j|X)\}\Rightarrow X\in\omega_i$$

其中R为损失函数

## 3.5 贝叶斯统计判决规则的似然比表现形式

定义类别的似然比 $l_{1,2}(X)=\frac{p(X|\omega_1)}{p(X|\omega_2)}$ ，判决阈值为 $\theta_{1,2}=\frac{P(\omega_2)}{P(\omega_1)}$ ，则判定简化为 $l_{1,2}$ 和 $\theta_{1,2}$ 的比较。

最小风险判决规则： $\theta_{1,2}=\frac{(L(\alpha_1|\omega_2)-L(\alpha_2|\omega_2))P(\omega_2)}{(L(\alpha_2|\omega_1)-L(\alpha_1|\omega_1))P(\omega_1)}$

## 3.6 拒绝判决

可能产生拒绝判决的前提条件： $P(\omega_i|X)<\left(1-\frac{\lambda_R}{\lambda_F}\right),\ \forall x\in[1,n]\Rightarrow X\in\alpha_{N+1}$

## 3.7 贝叶斯分类器的一般结构

取 $g_i(X)=P(\omega_i|X)$ ，也就是 $g_i(X)=\frac{p(X|\omega_i)P(\omega_i)}{\sum\limits_{j=1}^{N}p(X|\omega_j)P(\omega_j)}$ ，或在最小风险判决规则下取 $g_i(X)=-R(\alpha_i|X)$

## 3.8 Neyman-Peason判决规则

在保持 $\epsilon_1$ 一定的情况下使得 $\epsilon_2$ 最小。构造 $J=\epsilon_1+\lambda(\epsilon_2-\alpha)=(1-\lambda\alpha)+\int_{\Omega_1}(\lambda p(x|\omega_2)-p(x|\omega_1))\mathrm{d}x$

其实就是 $\frac{p(x|\omega_1)}{p(x|\omega_2)}>\lambda\Rightarrow x\in\omega_1$

## 3.9 最小最大判决规则

太复杂了，餐间刻本

## 3.10 基于分段线性化的分类器设计

同上

# 3.11 正态分布下的分类器设计

多元正态分布：$p(X)=\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(X-\mu)^T\Sigma^{-1}(X-\mu)\}$

1. $\Sigma_i=\sigma^2I$ 的情况
   超平面：$W^T(X-X_0)$，其中 $W=\mu_i-\mu_j$， $X_0=\frac{1}{2}(\mu_i+\mu_j)-\frac{\ln\frac{P(\omega_i)}{P(\omega_j)}}{(\mu_i-\mu_j)^T\sigma^{-1}(\mu_i-\mu_j)}(\mu_i-\mu_j)$
2. $\Sigma_i=\Sigma$ 非对角阵
   超平面：$W^T(X-X_0)$，其中 $W=\Sigma^{-1}(\mu_i-\mu_j)$，$X_0$ 一样
3. $\Sigma_i$ 任意
   超二次曲面

# 3.12 有监督情况下类条件概率密度的参数估计

## 极大似然估计

也就是最大似然估计，和数理统计学的一样

需要注意的是，$\frac{\partial|\Sigma|}{\partial\Sigma}=|\Sigma|(\Sigma^T)^{-1}$

## 贝叶斯估计和贝叶斯学习

假定 $\lambda(\hat{\theta}|\theta)$ 为真实参数为 $\theta$ 、给出 $\hat{\theta}$ 估计时承担的风险。以下考虑 $\lambda(\hat{\theta}|\theta)=(\hat{\theta}-\theta)^2$，得到如下定理：

- **定理3.1** $\theta$ 的贝叶斯估计量 $\hat{\theta}$ 是在给定样本$X$ 条件下 $\theta$ 的数学期望。或：$\hat{\theta}=\int_\Theta\theta p(\theta|X)\mathrm{d}\theta$

贝叶斯学习：和贝叶斯分布的前三个步骤类似，最后从后验概率密度直接得到总体概率密度。

# 3.13 非监督情况下类条件概率密度的估计

分类未知，最大似然方法适用于待定参数是确定未知量，贝叶斯学习方法适用于待定参数是随机的未知量。

条件概率密度参数估计：P131 公式3.100，要求 $P(\omega_i)$ 为已知量。

该算法可用于迭代正态分布。

# 3.14 类条件概率密度的非参数估计

## Parzen窗函数

对每个点应用窗函数。

## $k_n$-近邻估计法

扩张直到包含这些样本。

## 正交级数逼近法

用正交函数系的基函数 $\{\phi_i(X),i=1,2,\cdots,m\}$ 表示 $\hat{p}_n(X)$，其需要做的是确定这些函数的线性加权系数。

若选择归一化正交函数则有 $c_i=\frac{1}{n}\sum_{k=1}^{n}u(X_k)\phi_i(X_k)$