第二章 贝叶斯决策理论

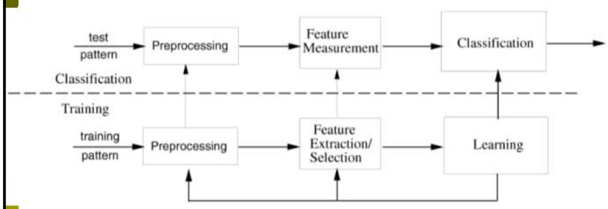
2009.09.20

本章主要内容

- 引言
- 基于最小错误率的Bayes决策
- 基于最小风险的Bayes决策
- 基于判别函数的分类器设计
- 正态分布的最小错误率Bayes决策
- 讨论

引言

引言



引言

- 客观现象或事物的发生和发展，按照“可预见性”可分两类情况：确定性和随机性；
  - 随机性事物的结果无法预知，但具有统计规律；
  - 随机性事物的特征观察值是随机变量。
- 特征的观察值总含有某种误差，其具有一定的随机性；而且同类的不同对象的某个特征的值通常也是按某种规律散布的。

➔ 模式类别和判决结果的随机性；

➔ 用概率统计的理论和方法来解决识别问题是合理的。

引言

- 统计模式识别的要点：将模式的特征向量考虑为符合某种统计规律（概率密度/分布函数）的随机向量；而任一样本是取自总体中的一个个体。
- 需要解决三个问题：
  - **判别问题**：已知若干总体分布，当给出一个个体样本时，要确定这个样本属于哪个总体？
  - **训练问题**：已知一些个体样本，分别属于某些总体，要确定这些总体的分布规律（或参数）。
  - **误判率问题**：研究运用上述模型所造成的误判率的计算。

概念和名词约定

- **样本sample**：待研究对象的个体，包括性质已知或未知的个体。
- **类别class**：将所研究的样本性质离散化成有限的类别，同一类的样本在该性质上不可区分。
  - 类别用ω_i (i=1，2，…，c，共c类)表示；
- **已知样本**：类别情况已知的样本。
- **未知样本**：类别情况未知的样本。
- **样本集**：若干样本的集合，分已知样本集和未知样本集。

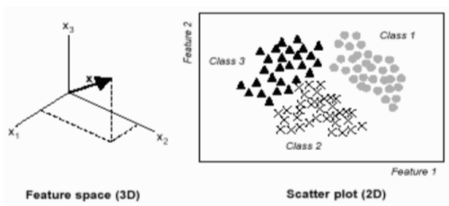
概念和名词约定

- **特征features**
  - 样本的任何可区分且可观测的属性；
  - 包括定量特征和定性特征，通常最后转化为定量特征；
  - **特征向量feature vectors**：样本的所有特征组成的d维向量，是样本在数学上的表达，因此也称为样本；

$$\mathbf{x} = [x_1, x_2, \cdots, x_d]^T.$$

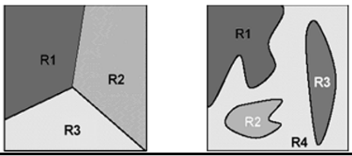
概念和名词约定

- **特征空间feature space**：**d**维特征向量的所有可能取值范围构成的**d**维特征空间。
  - 每个样本（特征向量）是该空间中的一个点，每类别则是该空间中的一个区域。



概念和名词约定

- **分类器classifier**：能将每个样本分到某个类别中（或拒绝）的计算机算法
  - 是从特征空间到决策空间的映射；
  - 分类器将特征空间划分为若干区域（决策域，**decision region**）；
  - 不同类别区域之间的边界称作分类/决策边界、分类/决策面（**decision boundary**）。



贝叶斯决策理论

贝叶斯定理

Reverend Thomas Bayes

1702-1761

$$P(x|y) = \frac{P(y|x)P(x)}{\sum_{x \in \mathcal{X}} P(y|x)P(x)}$$

Bayes set out his theory of probability in *Essay towards solving a problem in the doctrine of chances* published in the *Philosophical Transactions of the Royal Society of London* in 1764. The paper was sent to the Royal Society by Richard Price, a friend of Bayes', who wrote:-

I now send you an essay which I have found among the papers of our deceased friend Mr Bayes, and which, in my opinion, has great merit... In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times.

贝叶斯决策理论概述

- 例1：医生根据病人血液中白细胞的浓度来判断病人是否患有血液病。
  - 一个人的白细胞浓度是3100，医生应该做出怎样的判断？
  - 根据医学知识和以往的经验，医生知道：
    - 一般人群中，患病的人数比例为0.5%；
    - 患病的人白细胞的浓度服从均值2000，方差1000的正态分布；未患病的人白细胞的浓度服从均值7000，方差3000的正态分布。

贝叶斯决策理论概述

- 数学表示
  - 用Ω表示“类别”随机变量，类别ω_1和ω_2分别表示“患病” 和“未患病”；
  - 用 x 表示“白细胞浓度值”随机变量；
  - 决策空间Θ={ω_1，ω_2}。

贝叶斯决策理论概述

- 先验概率

$$P(\Omega=\omega_1)=0.5\%$$
$$P(\Omega=\omega_2)=99.5\%$$
$$P(\Omega=\omega_1)+P(\Omega=\omega_2)=1 \quad \text{(排他性、穷举性)}$$

  - 仅依据先验信息的判决规则

$$\text{Decide}\ \begin{cases}\omega_1, & \text{if } P(\omega_1)>P(\omega_2)\\ \omega_2, & \text{otherwise}\end{cases}$$

  - 判决的误差概率

$$P(error)=\min\{P(\omega_1),\ P(\omega_1)\}$$

贝叶斯决策理论概述

- 类条件概率
  - x 的分布取决于类别状态（患病或未患病），用类条件概率密度函数来表示：

$$p(x\mid\omega_1)\sim N(2000,1000);$$
$$p(x\mid\omega_2)\sim N(7000,3000);$$

    - $p(x\mid\omega_i)$ ：类别ω_i中，在一个连续的函数空间中观测到 x 的可能性；
    - $p(x\mid\omega_1)$ 和 $p(x\mid\omega_2)$ 间的区别表示了血液病人和非血液病人之间白细胞浓度值的区别。

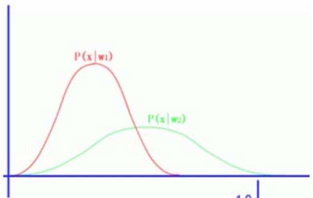
贝叶斯决策理论概述

- 后验概率
  - 问题：已知先验概率 $P(\omega_i)$ 和类条件概率密度函数 $p(x\mid\omega_i)$，$i=1,2$；对于一个样本 x=3100，判定 $x\in\omega_1$ 或 $x\in\omega_2$?
  - 即，计算在观测样本 x 下，其类别状态是ω_i（i=1,2）的概率：$P(\omega_i\mid x)$ 。
  - 后验概率即一个具体事物属于某种类别的概率。
    - 一个样本只可能属于两个类别之一，即有约束

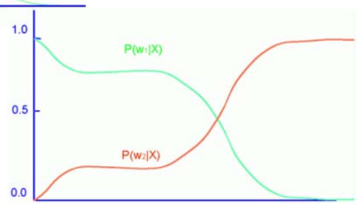
$$P(\omega_1\mid x)+P(\omega_2\mid x)=1$$

    - 区别 $P(\omega_i\mid x)$ 和 $P(\omega_i)$ 。

贝叶斯决策理论概述



$$P(\omega_i\mid x)=\frac{P(\omega_i,x)}{p(x)}=\frac{P(\omega_i)p(x\mid\omega_i)}{\sum_j P(\omega_j)p(x\mid\omega_j)}$$



基于最小错误率的贝叶斯决策

- 目标：最小化决策的**平均错误率** $P(error)$
  - 特征向量观测值的可能取值范围内的分类错误率的均值。

$$P(error) = \int_{-\infty}^{+\infty} P(error, x)dx$$
$$= \int_{-\infty}^{+\infty} P(error \mid x)p(x)dx$$
$$= E(P(error \mid x)).$$

  - 平均错误率是条件错误率的数学期望。

基于最小错误率的贝叶斯决策

- 条件错误率

$$P(error \mid x) = \begin{cases} P(\omega_2 \mid x), & \text{if assign } \ x \in \omega_1 \\ P(\omega_1 \mid x), & \text{if assign } \ x \in \omega_2 \end{cases}$$

$$\because P(error \mid x) \ge 0, \ p(x) \ge 0, \ \forall x$$
$$\therefore \min P(error \mid x), \ \text{for all } x \ \Rightarrow \ \min P(error)$$
$$\Rightarrow P(error \mid x) = \min\{P(\omega_1 \mid x), P(\omega_2 \mid x)\}, \ \forall x.$$

基于最小错误率的贝叶斯决策

- 判决规则及其等价形式
  - 最大化后验概率准则

$$\text{assign } x \in \begin{cases} \omega_1, & \text{if } P(\omega_1 \mid x) > P(\omega_2 \mid x) \\ \omega_2, & \text{otherwise} \end{cases}$$

$$\text{assign } x \in \begin{cases} \omega_1, & \text{if } p(x \mid \omega_1)P(\omega_1) > p(x \mid \omega_2)P(\omega_2) \\ \omega_2, & \text{otherwise} \end{cases}$$

基于最小错误率的贝叶斯决策

- 判决规则及其等价形式
  - 最大化似然比准则

$$\text{assign } x \in \begin{cases} \omega_1, & \text{if } l(x) = \dfrac{p(x \mid \omega_1)}{p(x \mid \omega_2)} > \dfrac{P(\omega_2)}{P(\omega_1)} \\ \omega_2, & \text{otherwise} \end{cases}$$

  - 对似然比取负对数

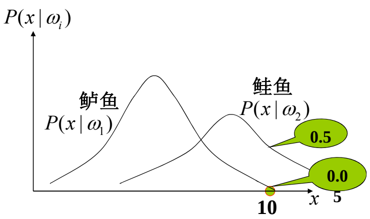
$$\text{assign } x \in \begin{cases} \omega_1, & \text{if } h(x) = -\ln p(x \mid \omega_1) + \ln p(x \mid \omega_2) > \ln \dfrac{P(\omega_2)}{P(\omega_1)}) \\ \omega_2, & \text{otherwise} \end{cases}$$

基于最小错误率的贝叶斯决策

- 例：两类鱼的自动分类问题，鲈鱼（ω_1）和鲑鱼（ω_2），用鱼长度的观察值（x）为特征。
  - 根据统计结果：

$P(\omega_1) = 1/3;$

$P(\omega_2) = 2/3;$



  - 如何将一条长为10的鱼分类？

基于最小错误率的贝叶斯决策

- 解法一：

$$P(\omega_1 \mid x=10) = \frac{p(x=10 \mid \omega_1)P(\omega_1)}{p(x=10)}$$
$$= \frac{p(x=10 \mid \omega_1)P(\omega_1)}{p(x=10 \mid \omega_1)P(\omega_1) + p(x=10 \mid \omega_2)P(\omega_2)}$$
$$= \frac{0.05 \times 1/3}{0.05 \times 1/3 + 0.50 \times 2/3} = 0.048;$$

$P(\omega_2 \mid x=10) = 1 - P(\omega_1 \mid x=10) = 0.952;$

$\Rightarrow P(\omega_1 \mid x=10) < P(\omega_2 \mid x=10);$

$\Rightarrow x = 10 \in \omega_2$， 即是鲑鱼。

基于最小错误率的贝叶斯决策

25

- 解法二（用似然比）：

$$l_{12}(x=10)=\frac{p(x=10\,|\,\omega_1)}{p(x=10\,|\,\omega_2)}=\frac{0.05}{0.50}=0.1;$$

$$\text{判决阀值}\ \theta_{12}=\frac{P(\omega_2)}{P(\omega_1)}=\frac{2/3}{1/3}=2;$$

$\because\ \ l_{12}(x=10)<\theta_{12}$ ；

$\therefore\ \ x=10\in\omega_2$ ，即是鲑鱼。

基于最小错误率的贝叶斯决策

26

- 再看决策的错误率

$$\begin{aligned}P(error)&=\int_{R_1}P(\omega_2\,|\,x)p(x)dx+\int_{R_2}P(\omega_1\,|\,x)p(x)dx\\&=\int_{R_1}P(x\,|\,\omega_2)P(\omega_2)dx+\int_{R_2}P(x\,|\,\omega_1)P(\omega_1)dx\\&=P(\omega_2)\int_{R_1}P(x\,|\,\omega_2)dx+P(\omega_1)\int_{R_2}P(x\,|\,\omega_1)dx\end{aligned}$$

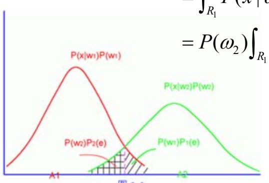


图 2.3

错误率为图中两个划线部分之和，
对应的错误率区域面积为最小。

基于最小错误率的贝叶斯决策

27

- 推广
  - 允许使用多于一个的特征，即用特征向量；
  - 允许多于两种的类别状态；
  - 允许除了判定类别以外的其他行为（如拒绝）；
  - 可以引入比误差概率更一般的损失函数。

$(1)\ \text{assign}\ x\in\omega_i\ \ \text{if}\ P(\omega_i\,|\,x)=\max_{j=1,\cdots,c}P(\omega_j\,|\,x)$

$(2)\ \text{assign}\ x\in\omega_i\ \ \text{if}\ p(x\,|\,\omega_i)P(\omega_i)=\max_{j=1,\cdots,c}p(x\,|\,\omega_j)P(\omega_j);$

$$P(error)=1-P(correct)=1-\sum_{j=1}^{c}\int_{R_j}p(x\,|\,\omega_j)P(\omega_j)dx$$