# 第二章 贝叶斯决策理论

2010. 09. 27



## 基于最小风险的贝叶斯决策

- 定义
  - **状态空间Ω：** 由$c$个可能的状态（$c$类）组成

$$\Omega=\{\omega_1,\omega_2,\cdots,\omega_c\}$$

  - **决策空间$\mathcal{A}$：** 由所有可能采取的决策组成

$$\mathcal{A}=\{\alpha_1,\alpha_2,\cdots,\alpha_k\}$$

  - **损失函数** $\lambda(\alpha_i,\omega_j),\quad i=1,\cdots,k,\quad j=1,\cdots,c$ ：表示对真实状态为$\omega_j$的样本，采取决策 $\alpha_i$时所造成的损失。



## 基于最小风险的贝叶斯决策

| 损失 决策 \ 状态 | 自然状态 | | | | | |
|---|---|---|---|---|---|---|
| | $\omega_1$ | $\omega_2$ | … | $\omega_j$ | … | $\omega_c$ |
| $\alpha_1$ | $\lambda(\alpha_1,\omega_1)$ | $\lambda(\alpha_1,\omega_2)$ | … | $\lambda(\alpha_1,\omega_j)$ | … | $\lambda(\alpha_1,\omega_c)$ |
| $\alpha_2$ | $\lambda(\alpha_2,\omega_1)$ | $\lambda(\alpha_2,\omega_2)$ | … | $\lambda(\alpha_2,\omega_j)$ | … | $\lambda(\alpha_2,\omega_c)$ |
| ⋮ | ⋮ | ⋮ | … | ⋮ | ⋮ | ⋮ |
| $\alpha_i$ | $\lambda(\alpha_i,\omega_1)$ | $\lambda(\alpha_i,\omega_2)$ | … | $\lambda(\alpha_i,\omega_j)$ | … | $\lambda(\alpha_i,\omega_c)$ |
| ⋮ | ⋮ | ⋮ | … | ⋮ | ⋮ | ⋮ |
| $\alpha_a$ | $\lambda(\alpha_a,\omega_1)$ | $\lambda(\alpha_a,\omega_2)$ | … | $\lambda(\alpha_a,\omega_j)$ | … | $\lambda(\alpha_a,\omega_c)$ |

决策表



## 基于最小风险的贝叶斯决策

- **条件期望损失（条件风险）：** 对于特定的观察样本 $\mathbf{x}$（特征向量），决策 $\alpha_i$ 造成的损失对 $\mathbf{x}$ 实际所属类别的各种可能的平均：

$$R(\alpha_i\,|\,\mathbf{x}) = E\left[\lambda(\alpha_i,\omega_j)\right] = \sum_{j=1}^{c}\lambda(\alpha_i,\omega_j)P(\omega_j\,|\,\mathbf{x})$$

- **期望风险：** 对所有 $\mathbf{x}$ 取值所作的决策$\alpha(\mathbf{x})$所带来的平均风险，即条件风险对 $\mathbf{x}$ 的数学期望。

$$R(\alpha) = E[R(\alpha(\mathbf{x})\,|\,\mathbf{x})] = \int R(\alpha(\mathbf{x})\,|\,\mathbf{x})\,p(\mathbf{x})\,d\mathbf{x}$$



## 基于最小风险的贝叶斯决策

- 目标：最小化决策所带来损失的平均值—（平均）风险
- 决策规则

$$Decide\quad \alpha_k,\quad \text{if}\ \ R(\alpha_k\,|\,\mathbf{x}) = \min_{j=1,\cdots,a} R(\alpha_j\,|\,\mathbf{x})$$

- 最小风险决策的计算步骤
  - 计算后验概率 ➔ 计算每个决策的条件风险 ➔ 依据最小条件风险决策



## 基于最小风险的贝叶斯决策

- 两类别问题
  - 行为 $\alpha_1$： deciding $\omega_1$;
  - 行为 $\alpha_2$： deciding $\omega_2$;
  - 损失 $\lambda_{ij}=\lambda(\alpha_i,\omega_j),\quad i,j=1,\ 2.$
  - 条件风险

$$R(\alpha_1\,|\,\mathbf{x}) = \lambda_{11}P(\omega_1\,|\,\mathbf{x})+\lambda_{12}P(\omega_2\,|\,\mathbf{x})$$
$$R(\alpha_2\,|\,\mathbf{x}) = \lambda_{21}P(\omega_1\,|\,\mathbf{x})+\lambda_{22}P(\omega_2\,|\,\mathbf{x})$$

## 基于最小风险的贝叶斯决策

- 两类别问题
  - 决策规则

$$\text{Decide}\begin{cases}\omega_1, & \text{if } (\lambda_{21}-\lambda_{11})P(\omega_1\mid\mathbf{x})>(\lambda_{12}-\lambda_{22})P(\omega_2\mid\mathbf{x})\\ \omega_2, & \text{otherwise}\end{cases}$$

$$\text{Decide}\begin{cases}\omega_1, & \text{if } \dfrac{P(\mathbf{x}\mid\omega_1)}{P(\mathbf{x}\mid\omega_2)}>\dfrac{(\lambda_{12}-\lambda_{22})}{(\lambda_{21}-\lambda_{11})}\dfrac{P(\omega_2)}{P(\omega_1)}\\ \omega_2, & \text{otherwise}\end{cases}$$



## 基于最小风险的贝叶斯决策

- 例：两类细胞识别问题：正常类和异常类
  - 根据已有知识和经验，两类的先验概率：

    正常（$\omega_1$）： $P(\omega_1)=0.9$

    异常（$\omega_2$）： $P(\omega_2)=0.1$

    对某一样本观察值 $\boldsymbol{x}$，通过计算或查表得到：
    $p(\mathbf{x}|\omega_1)=0.2$，$p(\mathbf{x}|\omega_2)=0.4$

    $\lambda_{11}=0$，$\lambda_{12}=6$，$\lambda_{21}=1$，$\lambda_{22}=0$
  - 依据最小风险决策如何对细胞 $\boldsymbol{x}$ 进行分类？



## 基于最小风险的贝叶斯决策

- 解：

$$P(\omega_1\mid\mathbf{x})=\frac{0.9\times0.2}{0.9\times0.2+0.1\times0.4}=0.818,$$

$$P(\omega_2\mid\mathbf{x})=\frac{0.4\times0.1}{0.2\times0.9+0.4\times0.1}=0.182;$$

$$R(\alpha_1\mid\mathbf{x})=\sum_{j=1}^{2}\lambda_{1j}P(\omega_j\mid\mathbf{x})=\lambda_{12}P(\omega_2\mid\mathbf{x})=1.092,$$

$$R(\alpha_2\mid\mathbf{x})=\sum_{j=1}^{2}\lambda_{2j}P(\omega_j\mid\mathbf{x})=\lambda_{21}P(\omega_1\mid\mathbf{x})=0.818;$$

$$\because R(\alpha_1\mid\mathbf{x})>R(\alpha_2\mid\mathbf{x}),\quad \therefore\ \text{Decide}\ \ \alpha_2,\ \ \mathbf{x}\in\omega_2.$$



## 两种决策方法之间的关系

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

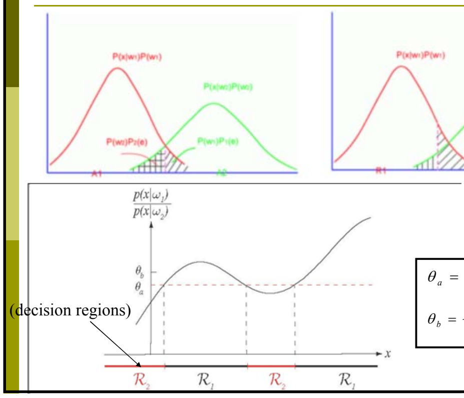
$$R(\alpha_i\mid\mathbf{x})=\sum_{j=1}^{c}\lambda(\alpha_i,\omega_j)P(\omega_j\mid\mathbf{x})$$

$$=\sum_{j=1,j\neq i}^{c}P(\omega_j\mid\mathbf{x})=1-P(\omega_i\mid\mathbf{x})$$

$$\min_{\alpha}R(\alpha_i\mid\mathbf{x})\Leftrightarrow\max_{\alpha}P(\omega_i\mid\mathbf{x})$$



## 两种决策方法之间的关系




$$\theta_a=P(\omega_2)/P(\omega_1)$$

$$\theta_b=\frac{P(\omega_2)(\lambda_{12}-\lambda_{22})}{P(\omega_1)(\lambda_{21}-\lambda_{11})}$$



## Neyman-Pearson 决策

- 两类错误率
  - 令 $R$ 是整个特征空间，$R_1$ 是类别$\omega_1$的决策域，$R_2$ 是类别$\omega_2$的决策域：$R_1+R_2=R$。

$$P(error)=\int_{R_1}P(\omega_2,\mathbf{x})d\mathbf{x}+\int_{R_2}P(\omega_1,\mathbf{x})d\mathbf{x}$$

$$=\int_{R_1}p(\mathbf{x}\mid\omega_2)P(\omega_2)d\mathbf{x}+\int_{R_2}p(\mathbf{x}\mid\omega_1)P(\omega_1)d\mathbf{x}$$

$$=P(\omega_2)\int_{R_1}p(\mathbf{x}\mid\omega_2)d\mathbf{x}+P(\omega_1)\int_{R_2}p(\mathbf{x}\mid\omega_1)d\mathbf{x}$$

$$=P(\omega_2)P_2(error)+P(\omega_1)P_1(error)$$

两类错误率

## Neyman-Pearson 决策

13

- 决策目标：在 $P_2(error)=\varepsilon_0$ 条件下，求 $P_1(error)$ 极小值。
  - 根据Lagrange乘子法

$$\gamma = P_1(error) + \lambda(P_2(error) - \varepsilon_0),$$

其中λ是Lagrange乘子，目标是求γ的极小值。

注意：$P_1(error) = \int_{R_2} p(\mathbf{x} \mid \omega_1) d\mathbf{x} = 1 - \int_{R_1} p(\mathbf{x} \mid \omega_1) d\mathbf{x};$

$P_2(error) = \int_{R_1} p(\mathbf{x} \mid \omega_2) d\mathbf{x} = 1 - \int_{R_2} p(\mathbf{x} \mid \omega_2) d\mathbf{x} = \varepsilon_0.$

## Neyman-Pearson 决策

14

- 决策目标：极小化γ

$$\gamma = (1 - \lambda\varepsilon_0) + \int_{R_1} [\lambda p(\mathbf{x}|\omega_2) - p(\mathbf{x}|\omega_1)] d\mathbf{x};$$

或者

$$\gamma = (1 - \varepsilon_0)\lambda + \int_{R_2} [p(\mathbf{x}|\omega_1) - \lambda p(\mathbf{x}|\omega_2)] d\mathbf{x};$$

$$\frac{\partial \gamma}{\partial \lambda} = 0 \quad \Rightarrow \quad \int_{R_1} p(\mathbf{x} \mid \omega_2) d\mathbf{x} = \varepsilon_0,$$

$$\frac{\partial \gamma}{\partial \mathbf{t}} = 0 \quad \Rightarrow \quad \lambda^* = \frac{p(\mathbf{t}^* \mid \omega_1)}{p(\mathbf{t}^* \mid \omega_2)}.$$

## Neyman-Pearson 决策

15

- 决策准则

$$\text{if } \lambda p(\mathbf{x}|\omega_2) \underset{>}{<} p(\mathbf{x}|\omega_1), \quad \text{then } \mathbf{x} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

or

$$l(\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} \underset{<}{>} \lambda, \quad \text{then } \mathbf{x} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

➔N-P决策规则归结为寻找阈值λ，使得

$$\int_{R_1} p(\mathbf{x} \mid \omega_2) d\mathbf{x} = \varepsilon_0.$$

## Neyman-Pearson 决策

16

- 例：一个两类问题中，模式均为二维正态分布，其均值矢量和协方差阵分别为：

$$\mu_1 = (-1, 0)^T, \ \mu_2 = (1, 0)^T, \ \Sigma_1 = \Sigma_2 = I.$$

设 $\varepsilon_0 = 0.09$，求Neyman-Pearson的决策阈值。

- 解：

$$p(\mathbf{x} \mid \omega_1) = \frac{1}{2\pi}\exp\left[-\frac{1}{2}(\mathbf{x}-\mu_1)^T(\mathbf{x}-\mu_1)\right] = \frac{1}{2\pi}\exp\left[-\frac{1}{2}\left((x_1+1)^2 + x_2^2\right)\right];$$

$$p(\mathbf{x} \mid \omega_2) = \frac{1}{2\pi}\exp\left[-\frac{1}{2}(\mathbf{x}-\mu_2)^T(\mathbf{x}-\mu_2)\right] = \frac{1}{2\pi}\exp\left[-\frac{1}{2}\left((x_1-1)^2 + x_2^2\right)\right];$$

$$\therefore \frac{p(\mathbf{x} \mid \omega_1)}{p(\mathbf{x} \mid \omega_2)} = \exp(-2x_1).$$

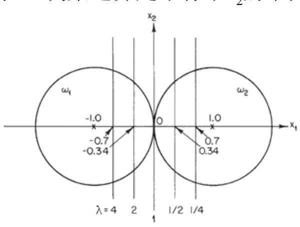
## Neyman-Pearson 决策

17

- 解：判决准则

$$\text{if } \exp(-2x_1) \underset{<}{>} \lambda, \quad \text{i. e. } x_1 \underset{>}{<} -\frac{1}{2}\ln\lambda, \quad \text{then } \mathbf{x} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases};$$

$\therefore$ 对于不同的$\lambda$，决策边界是平行于$x_2$的不同直线。（如图）




## Neyman-Pearson 决策

18

- 解：通过计算 $P_2(error)=\varepsilon_0$求解λ：

$$P_2(error) = \int_{R_1} p(\mathbf{x} \mid \omega_2) d\mathbf{x}$$

$$= \int_{-\infty}^{-\frac{1}{2}\ln\lambda} \int_{-\infty}^{\infty} \frac{1}{2\pi}\exp\left[-\frac{(x_1-1)^2 + x_2^2}{2}\right] dx_2 dx_1$$

$$= \int_{-\infty}^{-\frac{1}{2}\ln\lambda} \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{(x_1-1)^2}{2}\right] dx_1.$$

| λ | 4 | 2 | 1 | 1/2 | 1/4 |
|---|---|---|---|---|---|
| $\varepsilon_0$ | 0.046 | 0.089 | 0.0159 | 0.258 | 0.378 |

$\lambda$与$\varepsilon_0$的关系表

## 其他决策方法（自学）

- 最大最小决策
  - 基本思想：类先验概率未知，考查先验概率变化对错误率的影响，找出使最小风险贝叶斯决策的风险最大的先验概率，以这种最坏情况设计分类器。
- 序贯分类方法
  - 基本思想：除考虑分类造成的损失外，还考虑特征获取所造成的代价。先用一部分特征分类，然后逐步加入新特征以减少分类损失，同时衡量总的损失，以求得最优的效益。

# 分类器设计



## 分类器设计

- **判别函数**：是模式（或特征向量）$\mathbf{x}$ 的函数，用于表述决策规则
  - 对于 $c$ 类别问题，相应于每一类别定义一个函数，构成一组判别函数 $g_i(\mathbf{x}),\ i = 1,2,\ldots,c$，使得

$$g_i(\mathbf{x}) > g_j(\mathbf{x}) \quad \Rightarrow \quad \mathbf{x} \in \omega_i \quad j = 1,\cdots,c, \quad j \neq i;$$

  - 最小错误率Bayes决策的判别函数

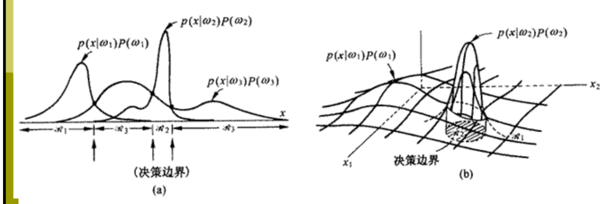
$$(1)\ \ g_i(\mathbf{x}) = P(\omega_i \mid \mathbf{x})$$
$$(2)\ \ g_i(\mathbf{x}) = p(\mathbf{x} \mid \omega_i)P(\omega_i)$$
$$(3)\ \ g_i(\mathbf{x}) = \ln p(\mathbf{x} \mid \omega_i) + \ln P(\omega_i)$$



## 分类器设计

- **决策面方程**：相邻的两个决策域在决策面上的判别函数值相等，即

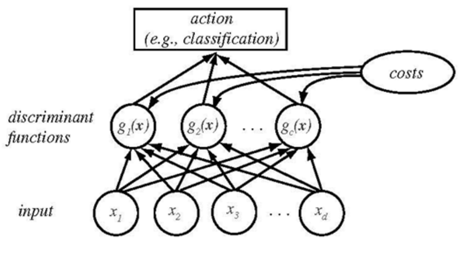
$$g_i(\mathbf{x}) = g_j(\mathbf{x}).$$






## 分类器设计

- **分类器**：计算 $c$ 个判别函数并选取与最大判别函数值相对应的类别的网络或机器。






## 分类器设计

- 两类别的最小错误率Bayes决策
  - 判决函数：

$$g(\mathbf{x}) = g_1(\mathbf{x}) - g_2(\mathbf{x}),$$

  - 相应的决策规则

$$\text{if } g(\mathbf{x}) \underset{<}{>} 0, \text{ then decide } \mathbf{x} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}.$$

$$\boxed{\begin{aligned} &(1)\ \ g(\mathbf{x}) = P(\omega_1 \mid \mathbf{x}) - P(\omega_2 \mid \mathbf{x}) \\ &(2)\ \ g(\mathbf{x}) = p(\mathbf{x} \mid \omega_1)P(\omega_1) - p(\mathbf{x} \mid \omega_2)P(\omega_2) \\ &(3)\ \ g(\mathbf{x}) = \ln \frac{p(\mathbf{x} \mid \omega_1)}{p(\mathbf{x} \mid \omega_2)} + \ln \frac{P(\omega_1)}{P(\omega_2)} \end{aligned}}$$

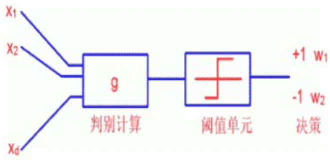
## 分类器设计

25

- 两类别的最小错误率Bayes决策
  - 决策面方程

$$g(\mathbf{x}) = 0.$$

  - 分类器

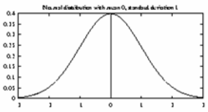



## 正态分布的<br>最小错误率贝叶斯决策

## 单变量的正态分布

27

A bell-shaped distribution defined by the probability density function



$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

If the random variable $X$ follows a normal distribution, then

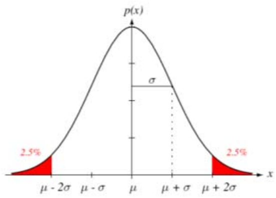
- The probability that $X$ will fall into the interval $(a,b)$ is given by $\int_a^b p(x)dx$
- Expected, or mean, value of $X$ is $E[X] = \int_{-\infty}^{\infty} xp(x)dx = \mu$
- Variance of $X$ is $Var(x) = E[(x-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx = \sigma^2$
- Standard deviation of X, $\sigma^2$, is $\sigma_x = \sigma$

## 单变量的正态分布

28

- $p(x)$ 完全由 μ 与 σ $^2$确定，常记作N(μ, σ$^2$)。
- 正态分布的熵(entropy)在所有的已知均值及方差的分布中最大。
- $p(x)$ 关于均值对称，最大值位于 $x=\mu$处，

$$p(\mu) = \frac{1}{\sqrt{2\pi}\sigma}.$$




## 多元正态分布

29

- 概率密度函数

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

其中：

$\mathbf{x} = [x_1, x_2, \cdots, x_d]^T$是$d$维列向量（$T$表示向量的转置）；

$\boldsymbol{\mu} = [\mu_1, \mu_2, \cdots, \mu_d]^T$是$d$维均值向量；

$\mathbf{\Sigma}$是$d \times d$协方差矩阵；

$\mathbf{\Sigma}^{-1}$是$\mathbf{\Sigma}$的逆矩阵，$|\mathbf{\Sigma}|$是$\mathbf{\Sigma}$的行列式。

## 多元正态分布

30

- 均值 μ 和协方差矩阵 Σ

$$\boldsymbol{\mu} = E[\mathbf{x}] = \int \mathbf{x} p(\mathbf{x}) d\mathbf{x};$$

$$\mathbf{\Sigma} = E[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T] = \int (\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T p(\mathbf{x}) d\mathbf{x};$$

$$\mu_i = E[x_i] = \int x_i p(x_i) dx_i$$

$$\sigma_{ij}^2 = E[(x_i-\mu_i)(x_j-\mu_j)^T] = \iint (x_i-\mu_i)(x_j-\mu_j)^T p(x_i,x_j) dx_i dx_j.$$

$$\Sigma = \begin{bmatrix} E[(x_1-u_1)^2] & E[(x_1-u_1)(x_2-u_2)] \\ E[(x_2-u_2)(x_1-u_1)] & E[(x_2-u_2)^2] \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_2^2 \end{bmatrix}.$$