



贝叶斯决策

南开大学网络空间安全学院 计算机学院

模式识别问题

- ❖ 已知：模式特征矢量（输入变量）集合

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$$

- ❖ 求解：与模式特征矢量相匹配的特性值（输出变量）（集合）

$$\mathbf{Y} = \{y_i \mid y_i = f(\mathbf{x}_i), y_i \in [1, C], i \in [1, n]\}$$

- ❖ 前提：分类模型“学会”如何将输入变量映射为输出变量

南开大学网络空间安全学院 计算机学院

2

贝叶斯规则

❖ 令 A 和 B 表示两个事件，它们的联合概率表示为

$$P(A, B) = P(A|B)P(B)$$

或

$$P(A, B) = P(B|A)P(A)$$

由上述方程可得贝叶斯规则如下：

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

贝叶斯规则

❖ 表示形式

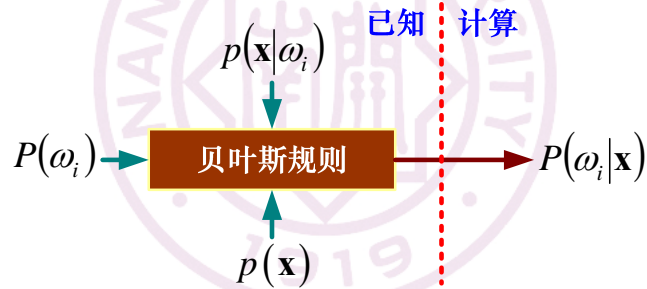
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \Rightarrow P(\omega_i | \mathbf{x}) = \frac{p(\mathbf{x} | \omega_i)P(\omega_i)}{p(\mathbf{x})}$$

其中

- ◆ $P(\omega_i)$: 模式类别 ω_i 的先验概率 (观测到模式特征 \mathbf{x} 之前)
- ◆ $p(\mathbf{x} | \omega_i)$: 模式特征 \mathbf{x} 的似然函数 (模式特征 \mathbf{x} 来自于模式类别 ω_i 的可能性)
- ◆ $p(\mathbf{x})$: 证据因子 (比例因子)
- ◆ $P(\omega_i | \mathbf{x})$: 模式类别 ω_i 的后验概率 (观测到模式特征 \mathbf{x} 之后)

贝叶斯规则

- ❖ 类别 ω_i 的两个概率
 - ◆ 先验概率 $P(\omega_i)$: 观测到模式之前
 - ◆ 后验概率 $P(\omega_i|\mathbf{x})$: 观测到模式之后
- ❖ 贝叶斯规则将二者联系起来



贝叶斯规则

- ❖ 如果 C 个模式类别是互斥和完备的, 那么有

$$\sum_{i=1}^C P(\omega_i|\mathbf{x}) = 1$$

- ❖ 贝叶斯规则的另一个表示形式

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

- ❖ 典型决策方法:

$$\omega^* = \arg \max_{\omega_i} P(\omega_i|\mathbf{x})$$

贝叶斯规则

❖ 问题：已知模式类别的先验概率，那么，在未观测到模式 \mathbf{x} 之前，模式 \mathbf{x} 类别的预测结果是什么？

◆ 例：已知一筐水果中包含70个苹果和30个鸭梨，随机拿出一个水果，它最可能是苹果还是鸭梨？

❖ 最优决策规则：

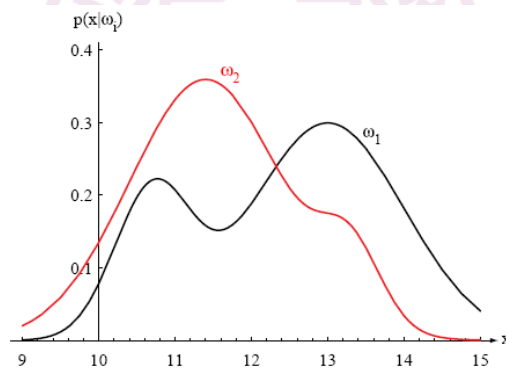
$$\text{Decide } \mathbf{x} \in \omega_i \text{ if } P(\omega_i) > P(\omega_j), \forall j \neq i$$

- ◆ 最优决策意味着决策错误率最小的决策；
- ◆ 在未观测到模式之前，具有最大先验概率的决策就是最优决策；
- ◆ 如果先验概率保持不变，那么多次实验将得到相同的决策结果。

贝叶斯规则

❖ 问题：如果获得更多有关模式的信息之后，是否能够获得更好的决策？

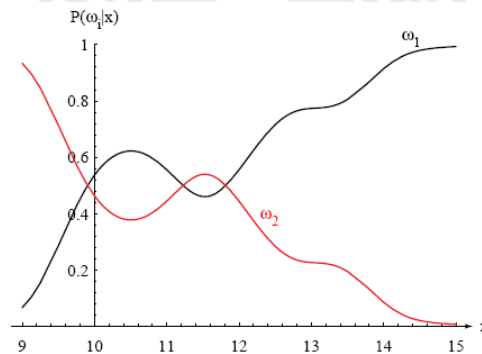
- ◆ 如果已知模式的似然函数 $p(\mathbf{x}|\omega_i)$ ，则可以通过贝叶斯规则获得最大后验概率决策。
- ◆ 模式特征是模式分类决策的基础。



贝叶斯规则

❖ 根据贝叶斯规则，有

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



南开大学网络空间安全学院 计算机学院

9

似然率测试规则

❖ 观测到模式后的最优决策就是**最大后验概率决策**，即

$$\text{Decide } \mathbf{x} \in \omega_i \text{ if } P(\omega_i | \mathbf{x}) > P(\omega_j | \mathbf{x}), \forall j \neq i$$

❖ 根据贝叶斯规则，上述决策等价于

$$\text{Decide } \mathbf{x} \in \omega_i \text{ if } \frac{p(\mathbf{x} | \omega_i) P(\omega_i)}{p(\mathbf{x})} > \frac{p(\mathbf{x} | \omega_j) P(\omega_j)}{p(\mathbf{x})}, \forall j \neq i$$

因为 $p(\mathbf{x})$ 与决策无关，故整理上式，可得

$$\text{Decide } \mathbf{x} \in \omega_i \text{ if } \Lambda(\mathbf{x}) = \frac{p(\mathbf{x} | \omega_i)}{p(\mathbf{x} | \omega_j)} > \frac{P(\omega_j)}{P(\omega_i)}, \forall j \neq i$$

其中： $\Lambda(\mathbf{x})$ 称为似然率。

南开大学网络空间安全学院 计算机学院

10

似然率测试规则

- ❖ 已知：模式的似然函数为

$$p(x|\omega_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-4)^2}, p(x|\omega_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-10)^2}$$

- ❖ 求解：似然率测试规则(LRT)

- ❖ 答：假设两个模式类别的先验概率相同，将模式似然函数和先验概率代入到似然率测试规则中，有

$$\text{Decide } x \in \omega_1 \text{ if } \Lambda(x) = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-4)^2}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-10)^2}} > \frac{1}{2}$$

化简表达式，有

似然率测试规则

$$\text{Decide } x \in \omega_1 \text{ if } \Lambda(x) = \frac{e^{-\frac{1}{2}(x-4)^2}}{e^{-\frac{1}{2}(x-10)^2}} > 1$$

不等式两边取对数，有

$$\text{Decide } x \in \omega_1 \text{ if } (x-4)^2 - (x-10)^2 < 0$$

化简上式，得到似然率测试规则

$$\text{Decide } x \in \omega_1 \text{ if } x < 7$$

- ❖ 思考：当先验概率不同时，似然率测试规则会有何变化？

错误概率

- ❖ 问题：如何度量决策规则的性能？
- ❖ 答案：选择决策错误概率来度量决策规则的性能

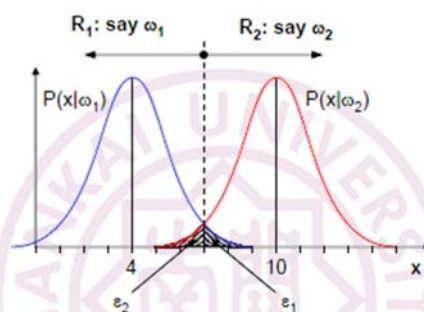
$$\Pr(\text{error}) = \sum_{i=1}^C \Pr(\text{error} | \omega_i) P(\omega_i)$$

$$\Pr(\text{error} | \omega_i) = \Pr(\text{decide } \omega_j | \omega_i) = \int_{R_j} p(\mathbf{x} | \omega_i) d\mathbf{x}$$

- ◆ 例：两类情形

$$\Pr(\text{error}) = \Pr(\omega_1) \underbrace{\int_{R_2} p(\mathbf{x} | \omega_1) d\mathbf{x}}_{\varepsilon_1} + \Pr(\omega_2) \underbrace{\int_{R_1} p(\mathbf{x} | \omega_2) d\mathbf{x}}_{\varepsilon_2}$$

错误概率

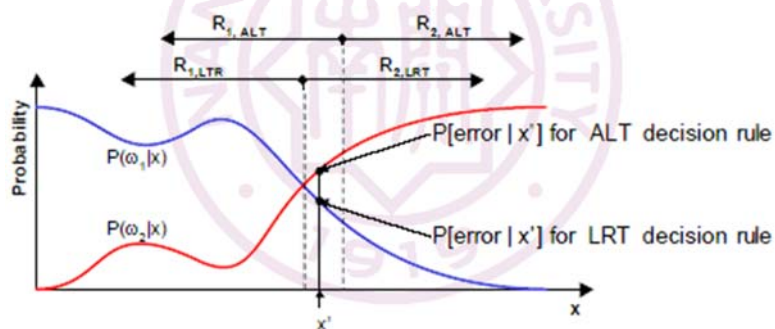


- ❖ 似然率测试规则的性能：将错误概率表示为后验概率的形式

$$\Pr(\text{error}) = \int_{-\infty}^{+\infty} \Pr(\text{error} | \mathbf{x}) p(\mathbf{x}) d\mathbf{x}$$

错误概率

- ❖ 最小错误概率要求在每个 x 取值上 $\Pr(\text{error}|x)$ 均最小, 得到
 - ◆ 积分值最小
 - ◆ 决策错误概率最小
 - ◆ 决策规则最优
- ❖ $\Pr(\text{error}|x')$ 等价于 $x' \in \omega_i$ 时的 $\Pr(\omega_j|x'), \forall j \neq i$



南开大学网络空间安全学院 计算机学院

15

错误概率

- ❖ 结论: 对于任何给定问题, 可以通过似然率测试决策规则得到**最小的决策错误概率**。这个决策错误概率称为**贝叶斯决策错误率**, 它是从所有决策规则中可以得到的**最佳决策结果**。

南开大学网络空间安全学院 计算机学院

16

贝叶斯风险

- 有些情形下，决策错误的代价是不同的。

◆ 代价函数 C_{ij} ：决策 $\text{decide } \omega_i \mid \mathbf{x} \in \omega_j$ 的代价

- 两类问题的贝叶斯风险

$$\begin{aligned} \text{Risk} &= \mathbb{E}[C] \\ &= \sum_{i=1}^2 \sum_{j=1}^2 C_{ij} \Pr(\text{decide } \omega_i \mid \mathbf{x} \in \omega_j) \\ &= \sum_{i=1}^2 \sum_{j=1}^2 C_{ij} p(\mathbf{x} \in R_i \mid \omega_j) \Pr(\omega_j) \end{aligned}$$

最小化贝叶斯风险决策规则

- 因为 $p(\mathbf{x} \in R_i \mid \omega_j) = \int_{R_i} p(\mathbf{x} \mid \omega_j) d\mathbf{x}$ ，故贝叶斯风险可表示为

$$\begin{aligned} \text{Risk} &= \int_{R_1} [C_{11} \Pr(\omega_1) p(\mathbf{x} \mid \omega_1) + C_{12} \Pr(\omega_2) p(\mathbf{x} \mid \omega_2)] d\mathbf{x} + \\ &\quad \int_{R_2} [C_{21} \Pr(\omega_1) p(\mathbf{x} \mid \omega_1) + C_{22} \Pr(\omega_2) p(\mathbf{x} \mid \omega_2)] d\mathbf{x} \end{aligned}$$

- 对每个似然函数，有

$$\int_{R_1} p(\mathbf{x} \mid \omega_i) d\mathbf{x} + \int_{R_2} p(\mathbf{x} \mid \omega_i) d\mathbf{x} = \int_{R_1 \cup R_2} p(\mathbf{x} \mid \omega_i) d\mathbf{x} = 1$$

最小化贝叶斯风险决策规则

❖ 将上式代入贝叶斯风险公式，整理后可得

$$\begin{aligned}
 Risk = & \underbrace{C_{11} \Pr(\omega_1) \int_{R_1} p(\mathbf{x}|\omega_1) d\mathbf{x}}_{\text{green}} + \underbrace{C_{12} \Pr(\omega_2) \int_{R_1} p(\mathbf{x}|\omega_2) d\mathbf{x}}_{\text{blue}} \\
 & + \underbrace{C_{21} \Pr(\omega_1) \int_{R_2} p(\mathbf{x}|\omega_1) d\mathbf{x}}_{\text{red}} + \underbrace{C_{22} \Pr(\omega_2) \int_{R_2} p(\mathbf{x}|\omega_2) d\mathbf{x}}_{\text{orange}} \\
 & + \underbrace{C_{21} \Pr(\omega_1) \int_{R_1} p(\mathbf{x}|\omega_1) d\mathbf{x}}_{\text{red}} + \underbrace{C_{22} \Pr(\omega_2) \int_{R_1} p(\mathbf{x}|\omega_2) d\mathbf{x}}_{\text{orange}} \\
 & - \underbrace{C_{21} \Pr(\omega_1) \int_{R_1} p(\mathbf{x}|\omega_1) d\mathbf{x}}_{\text{green}} - \underbrace{C_{22} \Pr(\omega_2) \int_{R_1} p(\mathbf{x}|\omega_2) d\mathbf{x}}_{\text{blue}}
 \end{aligned}$$

最小化贝叶斯风险决策规则

❖ 消除在区域 R_2 上的所有积分项，可得到

$$\begin{aligned}
 Risk = & \underbrace{C_{21} \Pr(\omega_1)}_{\text{red}} + \underbrace{C_{22} \Pr(\omega_2)}_{\text{orange}} \\
 & + \underbrace{(C_{12} - C_{22}) \Pr(\omega_2) \int_{R_1} p(\mathbf{x}|\omega_2) d\mathbf{x}}_{\substack{\text{blue} \\ >0}} \\
 & - \underbrace{(C_{21} - C_{11}) \Pr(\omega_1) \int_{R_1} p(\mathbf{x}|\omega_1) d\mathbf{x}}_{\substack{\text{green} \\ >0}}
 \end{aligned}$$

◆ 前两项与区域 R_1 无关，对于求解最小值操作可以视为常数。

最小化贝叶斯风险决策规则

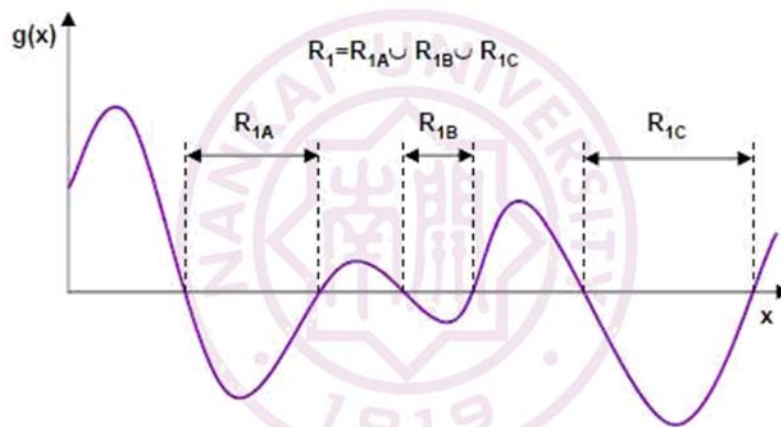
❖ 最小化贝叶斯风险的决策区域 R_1 为

$$R_1 = \arg \min_R \left\{ \int_R \left[(C_{12} - C_{22}) \Pr(\omega_2) p(\mathbf{x}|\omega_2) - (C_{21} - C_{11}) \Pr(\omega_1) p(\mathbf{x}|\omega_1) \right] d\mathbf{x} \right\}$$

$$= \arg \min_R \left\{ \int_R g(\mathbf{x}) d\mathbf{x} \right\}$$

最小化贝叶斯风险决策规则

❖ 最小化积分值的区域 R_1 等价于 $g(\mathbf{x}) < 0$ 的区域



最小化贝叶斯风险决策规则

❖ $g(\mathbf{x}) < 0$ 等价于

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

整理得到

$$\text{Decide } \mathbf{x} \in \omega_1 \text{ if } \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{(C_{12} - C_{22})}{(C_{21} - C_{11})} \cdot \frac{\Pr(\omega_2)}{\Pr(\omega_1)}$$

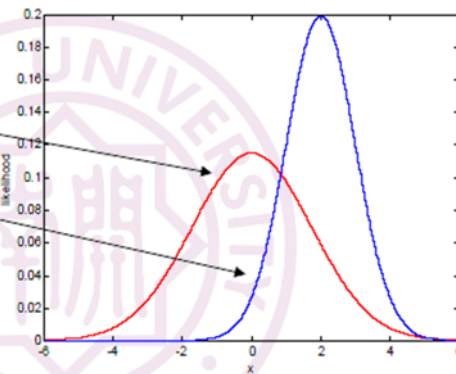
❖ 最小化贝叶斯风险决策规则等价于似然率测试规则。

贝叶斯风险

❖ 问题：考虑如下似然函数定义的两类分类问题

$$p(x|\omega_1) = \frac{1}{\sqrt{2\pi}\sqrt{3}} e^{-\frac{1}{2} \frac{x^2}{3}}$$

$$p(x|\omega_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-2)^2}$$



❖ 假设： $\Pr(\omega_1) = \Pr(\omega_2) = 0.5$, $C_{11} = C_{22} = 0$, $C_{12} = 1$, $C_{21} = \sqrt{3}$

❖ 求解：最小化决策错误概率的决策规则

贝叶斯风险

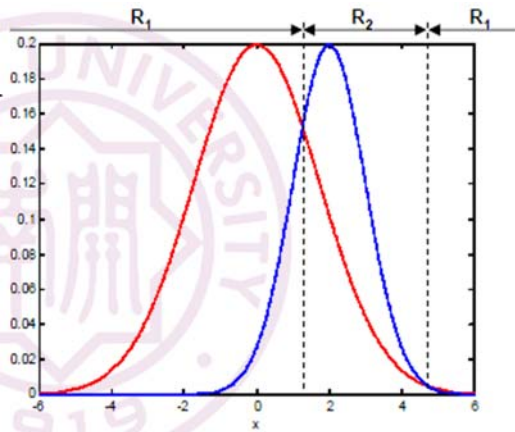
❖ 答案

$$\Lambda(x) = \frac{\frac{1}{\sqrt{2\pi}\sqrt{3}} e^{-\frac{1}{2} \frac{x^2}{3}}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-2)^2}} \stackrel{\omega_1}{>} \frac{1}{\sqrt{3}}$$

$$\frac{e^{-\frac{1}{2} \frac{x^2}{3}}}{e^{-\frac{1}{2}(x-2)^2}} \stackrel{\omega_1}{>} 1$$

$$-\frac{1}{2} \cdot \frac{x^2}{3} + \frac{1}{2}(x-2)^2 \stackrel{\omega_1}{>} 0$$

$$2x^2 - 12x + 12 \stackrel{\omega_1}{>} 0 \Rightarrow x = 4.73, 1.27$$



南开大学网络空间安全学院 计算机学院

25

似然率测试规则的变形

❖ 贝叶斯判据：最小化贝叶斯风险的似然率测试规则

$$\text{Decide } \mathbf{x} \in \omega_1 \text{ if } \Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{(C_{12} - C_{22}) \Pr(\omega_2)}{(C_{21} - C_{11}) \Pr(\omega_1)}$$

❖ 最大后验概率(MAP)判据：使用对称和零壹代价函数时，贝叶斯判据的特殊形式

$$C_{ij} = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \Rightarrow \Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{\Pr(\omega_2)}{\Pr(\omega_1)} \Leftrightarrow \frac{\Pr(\omega_1|\mathbf{x})}{\Pr(\omega_2|\mathbf{x})} > 1$$

南开大学网络空间安全学院 计算机学院

26

似然率测试规则的变形

- ❖ **最大似然(ML)判据**: 具有相同先验概率且使用零壹代价函数时, 最小化贝叶斯风险的似然率测试规则。

$$\left. \begin{aligned} C_{ij} &= \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \\ \Pr(\omega_i) &= \frac{1}{C} \quad \forall i \end{aligned} \right\} \Rightarrow \Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} \underset{\omega_2}{\overset{\omega_1}{>}} 1$$

- ❖ **Neyman-Pearson判据**: 在固定某个类别错误概率的情形下, 最小化其余类别错误概率的决策规则。
- ❖ **最小最大判据**: 最小化最大贝叶斯风险的决策规则。

多类问题的最小错误率决策规则

- ❖ 将错误决策概率表示为正确决策概率的形式

$$\Pr(\text{error}) = 1 - \Pr(\text{correct})$$

- ❖ 正确决策概率

$$\Pr(\text{correct}) = \sum_{i=1}^C \Pr(\omega_i) \int_{R_i} p(\mathbf{x}|\omega_i) d\mathbf{x}$$

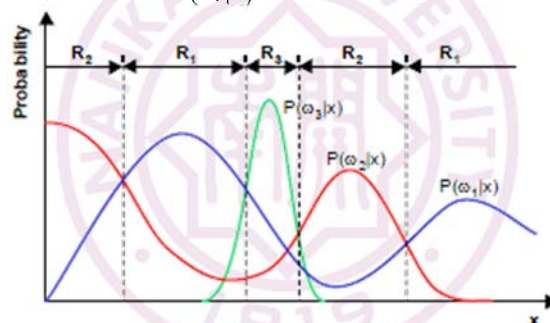
- ❖ 最小化错误决策概率等价于最大化正确决策概率

$$\begin{aligned} \Pr(\text{correct}) &= \sum_{i=1}^C \Pr(\omega_i) \int_{R_i} p(\mathbf{x}|\omega_i) d\mathbf{x} = \sum_{i=1}^C \int_{R_i} p(\mathbf{x}|\omega_i) \Pr(\omega_i) d\mathbf{x} \\ &= \sum_{i=1}^C \underbrace{\int_{R_i} \Pr(\omega_i|\mathbf{x}) p(\mathbf{x}) d\mathbf{x}}_{\mathfrak{S}_i} \end{aligned}$$

多类问题的最小错误率决策规则

❖ 最大化正确决策概率

- ◆ 需要最大化求和式中每个积分项 \mathfrak{I}_i
- ◆ 通过选择能够最大化 $\Pr(\omega_i | \mathbf{x})$ 的类别 ω_i 来最大化每个积分项 \mathfrak{I}_i
- ◆ 定义区域 R_i 为最大化 $\Pr(\omega_i | \mathbf{x})$ 的区域



❖ 结论：最小错误概率规则等价于最大后验概率判据。

南开大学网络空间安全学院 计算机学院

29

多类问题的最小贝叶斯风险决策规则

❖ 决策模式特征 \mathbf{x} 为模式类别 ω_i 的条件风险为

$$\text{Risk}(\alpha(\mathbf{x}) \rightarrow \alpha_i) = \text{Risk}(\alpha_i | \mathbf{x}) = \sum_{j=1}^C C_{ij} \Pr(\omega_j | \mathbf{x})$$

其中：

- ◆ α_i 表示选择模式类别 ω_i 的决策
- ◆ $\alpha(\mathbf{x})$ 表示对模式特征的决策规则，它完成从模式特征 \mathbf{x} 到模式类别 ω_i 的映射： $\alpha(\mathbf{x}) \rightarrow (\alpha_1, \alpha_2, \dots, \alpha_C)$

❖ 决策规则 $\alpha(\mathbf{x})$ 的贝叶斯风险

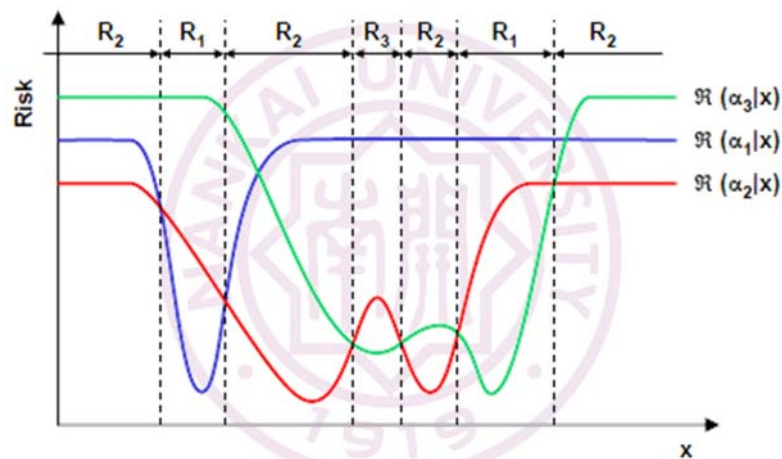
$$\text{Risk}(\alpha(\mathbf{x})) = \int \text{Risk}(\alpha(\mathbf{x}) | \mathbf{x}) p(\mathbf{x}) d\mathbf{x}$$

- ❖ 最小化贝叶斯风险需要最小化条件风险 $\text{Risk}(\alpha(\mathbf{x}) | \mathbf{x})$
- ❖ 最小化条件风险等价于满足最小化 $\text{Risk}(\alpha | \mathbf{x})$ 的模式类别 ω_i

南开大学网络空间安全学院 计算机学院

30

多类问题的最小贝叶斯风险决策规则

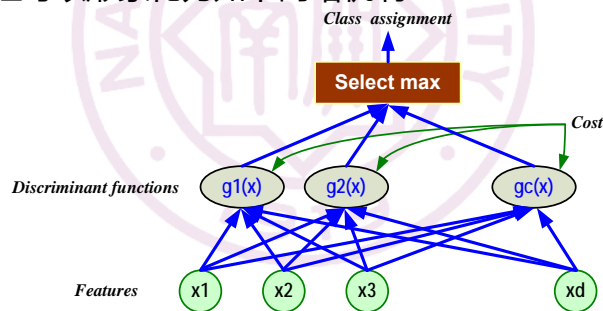


南开大学网络空间安全学院 计算机学院

31

判别式函数

- ❖ 决策规则拥有相同结构：在特征空间的每个点 \mathbf{x} 上选择最大化或最小化某个量 $g_i(\mathbf{x})$ 的类别 ω_i 。
- ❖ 这种结构可以形式化地表示为
 - ◆ 一组判别式函数： $g_i(\mathbf{x}), i = 1, 2, \dots, C$
 - ◆ 一个决策规则： decide $\mathbf{x} \in \omega_i$ if $g_i(\mathbf{x}) > g_j(\mathbf{x}) \quad \forall j \neq i$
- ❖ 决策过程可以形象化为如下网络机制



南开大学网络空间安全学院 计算机学院

32

判别式函数

❖ 以判别式函数形式表示的三个基本决策规则

Criterion	Discriminant Function
Bayes	$g_i(\mathbf{x}) = \mathcal{R}(\alpha_i \mathbf{x})$
MAP	$g_i(\mathbf{x}) = \Pr(\omega_i \mathbf{x})$
ML	$g_i(\mathbf{x}) = p(\mathbf{x} \omega_i)$

人物介绍：托马斯-贝叶斯



Thomas Bayes
1701–1761

Thomas Bayes was born in Tunbridge Wells and was a clergyman as well as an amateur scientist and a mathematician. He studied logic and theology at Edinburgh University and was elected Fellow of the Royal Society in 1742. During the 18th century, issues regarding probability arose in connection with

gambling and with the new concept of insurance. One particularly important problem concerned so-called inverse probability. A solution was proposed by Thomas Bayes in his paper 'Essay towards solving a problem in the doctrine of chances', which was published in 1764, some three years after his death, in the *Philosophical Transactions of the Royal Society*. In fact, Bayes only formulated his theory for the case of a uniform prior, and it was Pierre-Simon Laplace who independently rediscovered the theory in general form and who demonstrated its broad applicability.