模式识别

中国科学技术大学 汪增福

- 第一章 绪论
- 第二章 统计模式识别中的几何方法
- 第三章 统计模式识别中的概率方法
- ✓ 第四章 分类器的错误率
- 第五章 统计模式识别中的聚类方法
- 第六章 结构模式识别中的句法方法
- 第七章 总结

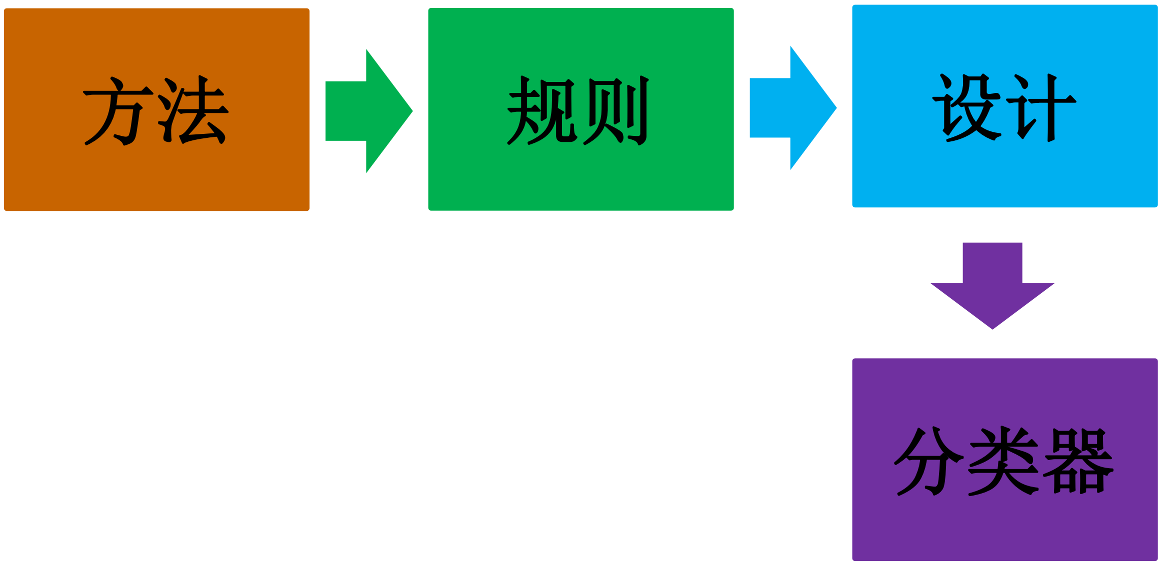
第四章 分类器的错误率

- 本章主要内容

  主要讨论各种分类器错误率的理论计算和实验估计问题。

  - 问题概述
  - 正态分布下的错误率
  - 各维统计独立情况下的错误率
  - 错误率界限的理论估计
  - 分类器错误率的实验估计

§ 4.1 问题概述



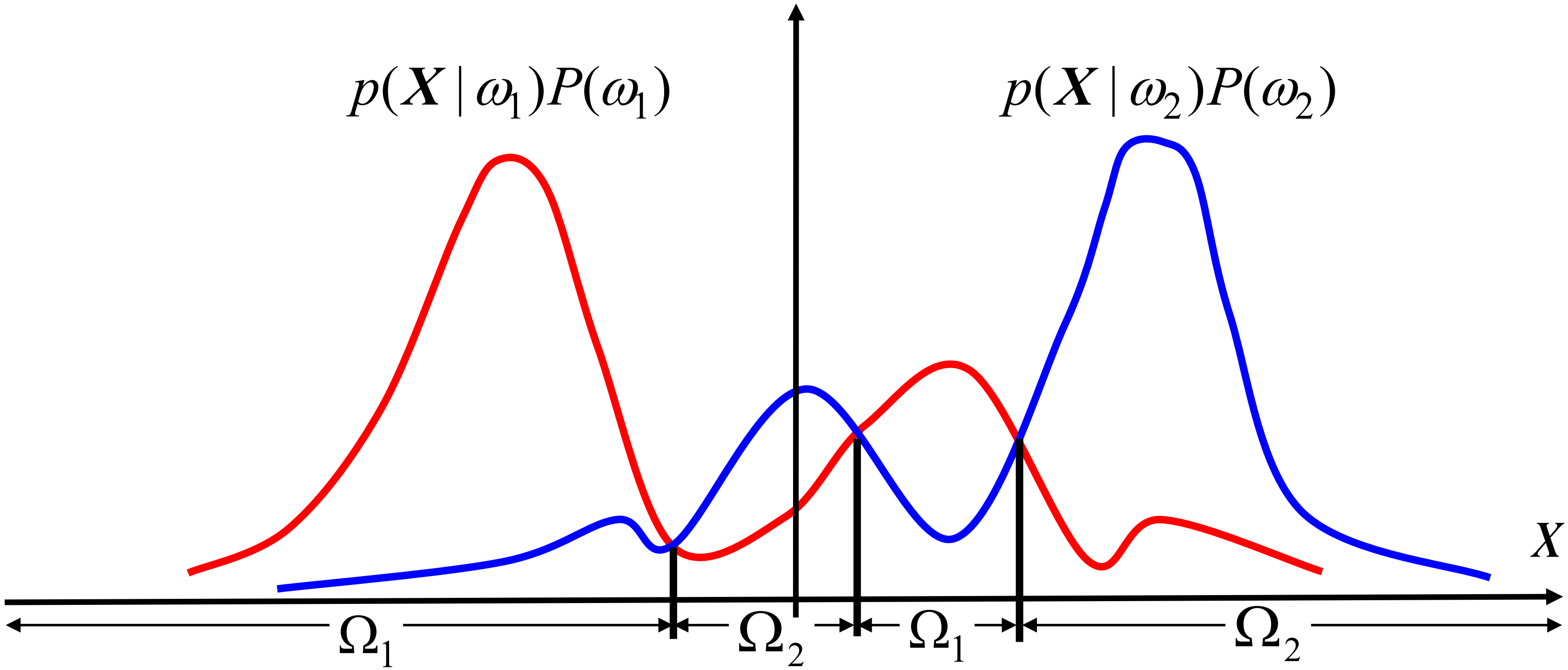
性能如何？

性能评估

- 错误率
  - 理论计算
  - 实验估计
- 实时性
- 计算代价

§ 4.1 问题概述

两类情况下分类器的分类错误率计算



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

$$P_1(e)=\int_{\Omega_2} p(\boldsymbol{X}\mid\omega_1)d\boldsymbol{X} \qquad P_2(e)=\int_{\Omega_1} p(\boldsymbol{X}\mid\omega_2)d\boldsymbol{X}$$

§ 4.2 正态分布下的错误率

两类情况下的分类错误率计算

$$\text{两类问题}\begin{cases}\omega_1\text{:} & N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \\ \omega_2\text{:} & N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)\end{cases}$$

$$p(\boldsymbol{X} \mid \omega_1) = \frac{1}{(2\pi)^{\frac{d}{2}}\left|\boldsymbol{\Sigma}_1\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_1)\right\}$$

$$p(\boldsymbol{X} \mid \omega_2) = \frac{1}{(2\pi)^{\frac{d}{2}}\left|\boldsymbol{\Sigma}_2\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_2^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_2)\right\}$$

若 $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$，则：

$$u_{12}(\boldsymbol{X}) = \ln\frac{p(\boldsymbol{X} \mid \omega_1)}{p(\boldsymbol{X} \mid \omega_2)} = -\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_1) + -\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_2)$$

$$= \boldsymbol{X}^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) - \frac{1}{2}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$$

§ 4.2 正态分布下的错误率

两类情况下的分类错误率计算

若采用最小风险判别规则，则 $\theta_{12} = \ln \dfrac{(L(\alpha_1 \mid \omega_2) - L(\alpha_2 \mid \omega_2))P(\omega_2)}{(L(\alpha_2 \mid \omega_1) - L(\alpha_1 \mid \omega_1))P(\omega_1)}$

判别规则为 $\begin{cases} u_{12}(\boldsymbol{X}) > \theta_{12} & \Rightarrow \quad \boldsymbol{X} \in \omega_1 \\ u_{12}(\boldsymbol{X}) < \theta_{12} & \Rightarrow \quad \boldsymbol{X} \in \omega_2 \end{cases}$

总的分类错误率为

$$\begin{aligned} \mathrm{P}(e) &= \mathrm{P}(\omega_1)\mathrm{P}_1(\mathrm{e}) + \mathrm{P}(\omega_2)\mathrm{P}_2(\mathrm{e}) \\ &= \mathrm{P}(\omega_1)P(u_{12}(\boldsymbol{X}) < \theta_{12} \mid \omega_1) + \mathrm{P}(\omega_2)P(u_{12}(\boldsymbol{X}) > \theta_{12} \mid \omega_2) \end{aligned}$$

$$u_{12}(\boldsymbol{X}) = \boldsymbol{X}^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

服从一维正态分布

§ 4.2 正态分布下的错误率

两类情况下的分类错误率计算

$X \in \omega_1$ 时，$u_{12}(X) = X^T \Sigma^{-1}(\mu_1 - \mu_2) - \frac{1}{2}(\mu_1 + \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2)$ 的均值和方差

$$\eta_1 = E(u_{12}(X)) = E(X^T \Sigma^{-1}(\mu_1 - \mu_2) - \frac{1}{2}(\mu_1 + \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2))$$

$$= \mu_1^T \Sigma^{-1}(\mu_1 - \mu_2) - \frac{1}{2}(\mu_1 + \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2)$$

$$= \frac{1}{2}(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2) = \frac{1}{2}\gamma_{ij}^2$$

$$\sigma_1^2 = E((u_{12}(X) - \eta_1)^2)$$

$$= E(X^T \Sigma^{-1}(\mu_1 - \mu_2) - \frac{1}{2}(\mu_1 + \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2) - \frac{1}{2}(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2))^2$$

$$= E((X - \mu_1)^T \Sigma^{-1}(\mu_1 - \mu_2))^2$$

$$= E(((X - \mu_1)^T \Sigma^{-1}(\mu_1 - \mu_2))^T (X - \mu_1)^T \Sigma^{-1}(\mu_1 - \mu_2))$$

$$= E((\mu_1 - \mu_2)^T \Sigma^{-1}(X - \mu_1)(X - \mu_1)^T \Sigma^{-1}(\mu_1 - \mu_2))$$

$$= (\mu_1 - \mu_2)^T \Sigma^{-1} E((X - \mu_1)(X - \mu_1)^T) \Sigma^{-1}(\mu_1 - \mu_2)$$

$$= (\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2) = \gamma_{ij}^2$$

$$\Sigma_1 = \Sigma$$

§ 4.2 正态分布下的错误率

两类情况下的分类错误率计算

$\boldsymbol{X} \in \omega_2$时，$u_{12}(\boldsymbol{X}) = \boldsymbol{X}^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ 的均值和方差

$$
\begin{aligned}
\eta_2 &= E(u_{12}(\boldsymbol{X})) = E(\boldsymbol{X}^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)) \\
&= \boldsymbol{\mu}_2{}^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\
&= -\frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = -\frac{1}{2}\gamma_{ij}^2
\end{aligned}
$$

$$
\begin{aligned}
\sigma_2^2 &= E((u_{12}(\boldsymbol{X}) - \eta_2)^2) \\
&= E(\boldsymbol{X}^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))^2 \\
&= E((\boldsymbol{X} - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))^2 \\
&= E(((\boldsymbol{X} - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))^T (\boldsymbol{X} - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)) \\
&= E((\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_2)(\boldsymbol{X} - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)) \\
&= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} \underbrace{E((\boldsymbol{X} - \boldsymbol{\mu}_2)(\boldsymbol{X} - \boldsymbol{\mu}_2)^T)}_{\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\
&= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \gamma_{ij}^2
\end{aligned}
$$

§ 4.2 正态分布下的错误率

两类情况下的分类错误率计算

$$\boldsymbol{X} \in \omega_1 \text{时，} \; u_{12}(\boldsymbol{X}) \sim N(\frac{1}{2}\gamma_{ij}^2, \gamma_{ij}^2)$$

$$\boldsymbol{X} \in \omega_2 \text{时，} \; u_{12}(\boldsymbol{X}) \sim N(-\frac{1}{2}\gamma_{ij}^2, \gamma_{ij}^2)$$

$$\mathrm{P}_1(\mathrm{e}) = P(u_{12}(\boldsymbol{X}) < \theta_{12} \mid \omega_1)$$

$$= \int_{-\infty}^{\theta_{12}} \frac{1}{\sqrt{2\pi}\gamma_{ij}} \exp\left\{-\frac{1}{2}\left(\frac{u_{12}(\boldsymbol{X}) - 1/2\gamma_{ij}^2}{\gamma_{ij}}\right)^2\right\} du_{12}(\boldsymbol{X})$$

$$= \int_{-\infty}^{\frac{\theta_{12}-1/2\gamma_{ij}^2}{\gamma_{ij}}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y^2\right\} dy$$

$$= \Phi\left(\frac{\theta_{12} - 1/2\gamma_{ij}^2}{\gamma_{ij}}\right)$$

$$\Phi(\xi) = \int_{-\infty}^{\xi} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y^2\right\} dy$$

§ 4.2 正态分布下的错误率

两类情况下的分类错误率计算

$$\begin{aligned}
\mathrm{P}_2(\mathrm{e}) &= P(u_{12}(\boldsymbol{X}) > \theta_{12} \mid \omega_2) \\
&= \int_{\theta_{12}}^{+\infty} \frac{1}{\sqrt{2\pi}\gamma_{ij}} \exp\left\{-\frac{1}{2}\left(\frac{u_{12}(\boldsymbol{X}) + 1/2\gamma_{ij}^2}{\gamma_{ij}}\right)^2\right\} du_{12}(\boldsymbol{X}) \\
&= \int_{\frac{\theta_{12}+1/2\gamma_{ij}^2}{\gamma_{ij}}}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y^2\right\} dy \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y^2\right\} dy - \int_{-\infty}^{\frac{\theta_{12}+1/2\gamma_{ij}^2}{\gamma_{ij}}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y^2\right\} dy \\
&= 1 - \Phi\left(\frac{\theta_{12} + 1/2\gamma_{ij}^2}{\gamma_{ij}}\right)
\end{aligned}$$

§ 4.2 正态分布下的错误率

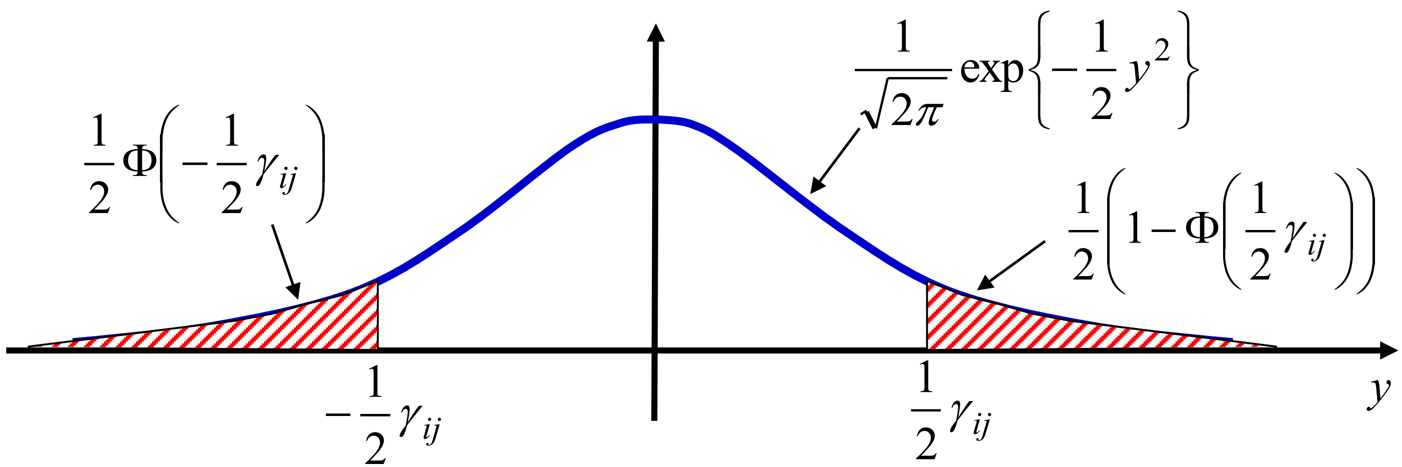
两类情况下的分类错误率计算

$$P(e) = \mathrm{P}(\omega_1)\mathrm{P}_1(\mathrm{e}) + \mathrm{P}(\omega_2)\mathrm{P}_2(\mathrm{e})$$

$$= \mathrm{P}(\omega_1)\Phi\left(\frac{\theta_{12} - 1/2\gamma_{ij}^2}{\gamma_{ij}}\right) + \mathrm{P}(\omega_2)\left(1 - \Phi\left(\frac{\theta_{12} + 1/2\gamma_{ij}^2}{\gamma_{ij}}\right)\right)$$

等先验概率, 0-1损失函数

$$P(e) = \frac{1}{2}\Phi\left(-\frac{1}{2}\gamma_{ij}\right) + \frac{1}{2}\left(1 - \Phi\left(\frac{1}{2}\gamma_{ij}\right)\right)$$



§ 4.3 各维统计独立情况下的错误率

两类情况下的分类错误率计算

$$p(\boldsymbol{X} \mid \omega_1) = \prod_{k=1}^{d} p(x_k \mid \omega_1)$$

$$p(\boldsymbol{X} \mid \omega_2) = \prod_{k=1}^{d} p(x_k \mid \omega_2)$$

$$u_{12}(\boldsymbol{X}) = \ln \frac{p(\boldsymbol{X} \mid \omega_1)}{p(\boldsymbol{X} \mid \omega_2)}$$

$$= \ln \frac{\prod_{k=1}^{d} p(x_k \mid \omega_1)}{\prod_{k=1}^{d} p(x_k \mid \omega_2)} = \ln \frac{p(x_1 \mid \omega_1) p(x_2 \mid \omega_1) \Lambda \, p(x_d \mid \omega_1)}{p(x_1 \mid \omega_2) p(x_2 \mid \omega_2) \Lambda \, p(x_d \mid \omega_2)}$$

$$= \sum_{k=1}^{d} \ln \frac{p(x_k \mid \omega_1)}{p(x_k \mid \omega_2)} = \sum_{k=1}^{d} u_{12}(x_k)$$

两个假设

- 维数d较大
- 各维之间统计独立

§ 4.3 各维统计独立情况下的错误率

两类情况下的分类错误率计算

最大似然判决 $\begin{cases} u_{12}(\boldsymbol{X}) > \theta_{12} & \Rightarrow \quad \boldsymbol{X} \in \omega_1 \\ u_{12}(\boldsymbol{X}) < \theta_{12} & \Rightarrow \quad \boldsymbol{X} \in \omega_2 \end{cases}$

各分量对数似然比 $u_{12}(x_k), k = 1,2,\ldots d$ 的均值和方差为

$$\eta_{i,k} = E(u_{12}(x_k \mid \omega_i)), \quad i = 1,2$$

$$\sigma_{i,k}^2 = E((u_{12}(x_k \mid \omega_i) - \eta_{i,k})^2), \quad i = 1,2$$

$u_{12}(\boldsymbol{X})$ 的均值和方差为

$$\eta_i = E(u_{12}(\boldsymbol{X} \mid \omega_i)) = E(\sum_{k=1}^{d} u_{12}(x_k \mid \omega_i)) = \sum_{k=1}^{d} \eta_{i,k}, \quad i = 1,2$$

$$\sigma_i^2 = E\left((u_{12}(\boldsymbol{X} \mid \omega_i) - \eta_i)^2\right) = E\left(\left(\sum_{k=1}^{d} u_{12}(x_k \mid \omega_i) - \sum_{k=1}^{d} \eta_{i,k}\right)^2\right) = \sum_{k=1}^{d} \sigma_{i,k}^2, \quad i = 1,2$$

§ 4.3 各维统计独立情况下的错误率

两类情况下的分类错误率计算

背景: 在实际中, 许多随机变量是由大量相互独立的偶然因素的综合影响所形成的. 每一个微小因素, 在总的影响中所起的作用很小, 但加起来却对总和有显著影响. 这种随机变量往往近似地服从正态分布.

中心极限定理:

独立同分布随机变量之和的极限分布是正态分布。

设随机变量$x_1, x_2, \ldots, x_n$相互独立且同分布, 并有有限的数学期望和方差, 则当n充分大时, 随机变量之和

$$\sum_{i=1}^{n} x_i$$

近似服从正态分布 $N(n\mu, n\sigma^2)$.

§ 4.3 各维统计独立情况下的错误率

两类情况下的分类错误率计算

根据中心极限定理，$u_{12}(\boldsymbol{X})$渐进服从正态分布 $N(\eta_i,\sigma_i^2),i=1,2$.

$$
\begin{aligned}
P_1(e) &= P(u_{12}(\boldsymbol{X})<\theta_{12}\mid\omega_1) \\
&= \int_{-\infty}^{\theta_{12}} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2}\left(\frac{u_{12}(\boldsymbol{X})-\eta_1}{\sigma_1}\right)^2\right\} du_{12}(\boldsymbol{X}) \\
&= \int_{-\infty}^{\frac{\theta_{12}-\eta_1}{\sigma_1}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y^2\right\} dy = \Phi\left(\frac{\theta_{12}-\eta_1}{\sigma_1}\right)
\end{aligned}
$$

$$
\begin{aligned}
P_2(e) &= P(u_{12}(\boldsymbol{X})>\theta_{12}\mid\omega_2) \\
&= \int_{\theta_{12}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{1}{2}\left(\frac{u_{12}(\boldsymbol{X})-\eta_2}{\sigma_2}\right)^2\right\} du_{12}(\boldsymbol{X}) \\
&= \int_{\frac{\theta_{12}-\eta_2}{\sigma_2}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y^2\right\} dy = 1-\Phi\left(\frac{\theta_{12}-\eta_2}{\sigma_2}\right)
\end{aligned}
$$

$$
\begin{aligned}
P(e) &= P(\omega_1)P_1(e)+P(\omega_2)P_2(e) \\
&= P(\omega_1)\Phi\left(\frac{\theta_{12}-\eta_1}{\sigma_1}\right)+P(\omega_2)\left(1-\Phi\left(\frac{\theta_{12}-\eta_2}{\sigma_2}\right)\right)
\end{aligned}
$$

§ 4.4 错误率界限的理论估计

Chernoff界限

两类问题 $\begin{cases} \omega_1: & p(\boldsymbol{X} \mid \omega_1) \\ \omega_2: & p(\boldsymbol{X} \mid \omega_2) \end{cases}$

负对数似然比

$$h(\boldsymbol{X}) = -u_{12}(\boldsymbol{X}) = -\ln \frac{p(\boldsymbol{X} \mid \omega_1)}{p(\boldsymbol{X} \mid \omega_2)} = -\ln p(\boldsymbol{X} \mid \omega_1) + \ln p(\boldsymbol{X} \mid \omega_2)$$

最大似然判决

$$\begin{cases} h(\boldsymbol{X}) < \theta_{12} & \Rightarrow \quad \boldsymbol{X} \in \omega_1 \\ h(\boldsymbol{X}) > \theta_{12} & \Rightarrow \quad \boldsymbol{X} \in \omega_2 \end{cases}$$

分类错误率 ➡ 上界

$$\mathrm{P}_1(\mathrm{e}) = P(u_{12}(\boldsymbol{X}) > \theta_{12} \mid \omega_1) = \int_{\theta_{12}}^{\infty} p(h \mid \omega_1) dh$$

$$\mathrm{P}_2(\mathrm{e}) = P(u_{12}(\boldsymbol{X}) < \theta_{12} \mid \omega_2) = \int_{-\infty}^{\theta_{12}} p(h \mid \omega_2) dh$$

§ 4.4 错误率界限的理论估计

Chernoff界限

$P_1(e)$ 的上界

引入

$$\varphi_i(s) = E_i(e^{sh}) = \int_{-\infty}^{\infty} e^{sh} p(h \mid \omega_i) dh, \quad i = 1,2$$

-矩母函数

$$\mu(s) = -\ln \varphi_1(s) = -\ln \int_{-\infty}^{\infty} e^{sh} p(h \mid \omega_1) dh$$

$$\begin{cases} \displaystyle\int_{-\infty}^{\infty} \left(\frac{e^{sh}}{\varphi_1(s)} \right) p(h \mid \omega_1) dh = 1 \\ \left(\dfrac{e^{sh}}{\varphi_1(s)} \right) p(h \mid \omega_1) \geq 0 \end{cases}$$

$$e^{-\mu(s)} = \int_{-\infty}^{\infty} e^{sh} p(h \mid \omega_1) dh$$

$$= \varphi_1(s)$$

概率密度，记为 $p(g = h \mid \omega_1) = \left(\dfrac{e^{sh}}{\varphi_1(s)} \right) p(h \mid \omega_1)$

§ 4.4 错误率界限的理论估计

Chernoff界限

g的均值和方差

$$E[g \mid \omega_1] = \int_{-\infty}^{\infty} g p(g \mid \omega_1) dg = \int_{-\infty}^{\infty} h\left(\frac{e^{sh}}{\varphi_1(s)}\right) p(h \mid \omega_1) dh$$

$$\frac{d\mu(\mathrm{s})}{ds} = \frac{d}{ds}(-\ln \varphi_1(\mathrm{s})) = -\frac{1}{\varphi_1(\mathrm{s})} \frac{d\varphi_1(\mathrm{s})}{ds}$$

$$= -\frac{1}{\varphi_1(\mathrm{s})} \frac{d}{ds} \int_{-\infty}^{\infty} e^{sh} p(h \mid \omega_1) dh = -\frac{1}{\varphi_1(\mathrm{s})} \int_{-\infty}^{\infty} h e^{sh} p(h \mid \omega_1) dh$$

$$= -\int_{-\infty}^{\infty} h\left(\frac{e^{sh}}{\varphi_1(\mathrm{s})}\right) p(h \mid \omega_1) dh = -E[g \mid \omega_1]$$

$$E[g \mid \omega_1] = -\frac{d\mu(s)}{ds} \qquad \sigma_g^2 = D[g \mid \omega_1] = -\frac{d^2\mu(s)}{ds^2}$$

§ 4.4 错误率界限的理论估计

Chernoff界限

$\mathrm{P}_1(\mathrm{e})$ 的上界

$$\mathrm{P}_1(\mathrm{e}) = \int_{\theta_{12}}^{\infty} p(h \mid \omega_1) dh$$

$$= \int_{\theta_{12}}^{\infty} \left(\frac{\varphi_1(\mathrm{s})}{e^{sh}} \right) \left(\frac{e^{sh}}{\varphi_1(\mathrm{s})} \right) p(h \mid \omega_1) dh$$

$$= \int_{\theta_{12}}^{\infty} \left(\frac{\varphi_1(\mathrm{s})}{e^{sh}} \right) p(g = h \mid \omega_1) dh$$

$$= \int_{\theta_{12}}^{\infty} e^{-\mu(s) - sh} p(g = h \mid \omega_1) dh$$

$$= e^{-\mu(s)} \int_{\theta_{12}}^{\infty} e^{-sh} p(g = h \mid \omega_1) dh$$ ⇨ 最小上界

§ 4.4 错误率界限的理论估计

Chernoff界限

$P_1(e)$ 的最小上界

$$P_1(e) = e^{-\mu(s)} \int_{\theta_{12}}^{\infty} e^{-sh} p(g = h \mid \omega_1) dh$$

$s \geq 0$

$h \geq \theta_{12} \longrightarrow -sh \leq -s\theta_{12} \longrightarrow e^{-sh} \leq e^{-s\theta_{12}}$

$$P_1(e) \leq e^{-\mu(s)-s\theta_{12}} \boxed{\int_{\theta_{12}}^{\infty} p(g = h \mid \omega_1) dh} \qquad s \geq 0$$

$$< \int_{-\infty}^{\infty} p(g = h \mid \omega_1) dh = 1$$

上界估计

$$P_1(e) \leq e^{-\mu(s)-s\theta_{12}} \qquad s \geq 0$$

当 θ_{12} 固定时，存在最佳的s使上界最小化。

§ 4.4 错误率界限的理论估计

Chernoff界限

$\mathrm{P}_2(\mathrm{e})$的最小上界

$$\mathrm{P}_2(\mathrm{e}) = \int_{\Omega_1} \mathrm{p}(\boldsymbol{X} \mid \omega_2)\mathrm{d}\boldsymbol{X}$$

$$\mathrm{P}_2(\mathrm{e}) = \int_{-\infty}^{\theta_{12}} p(h \mid \omega_2)dh$$

$$\int_{-\infty}^{\theta_{12}} p(h \mid \omega_2)dh = \int_{\Omega_1} p(\boldsymbol{X} \mid \omega_2)d\boldsymbol{X}$$

$$h(\boldsymbol{X}) = -\ln \frac{p(\boldsymbol{X} \mid \omega_1)}{p(\boldsymbol{X} \mid \omega_2)} = \ln \frac{p(\boldsymbol{X} \mid \omega_2)}{p(\boldsymbol{X} \mid \omega_1)}$$

$$e^{h(\boldsymbol{X})} = \frac{p(\boldsymbol{X} \mid \omega_2)}{p(\boldsymbol{X} \mid \omega_1)}$$

$$p(\boldsymbol{X} \mid \omega_2) = e^{h(\boldsymbol{X})} p(\boldsymbol{X} \mid \omega_1)$$

§ 4.4 错误率界限的理论估计

Chernoff界限

$\mathrm{P}_2(\mathrm{e})$的最小上界

$$\mathrm{P}_2(\mathrm{e}) = \int_{-\infty}^{\theta_{12}} p(h \mid \omega_2) dh$$

$$= \int_{\Omega_1} \mathrm{p}(\boldsymbol{X} \mid \omega_2) \mathrm{d}\boldsymbol{X} = \int_{\Omega_1} e^{h(\boldsymbol{X})} p(\boldsymbol{X} \mid \omega_1) \mathrm{d}\boldsymbol{X}$$

$$= \int_{-\infty}^{\theta_{12}} e^h p(h \mid \omega_1) dh = \int_{-\infty}^{\theta_{12}} e^h \left(\frac{\varphi_1(\mathrm{s})}{e^{sh}} \right) \left(\frac{e^{sh}}{\varphi_1(\mathrm{s})} \right) p(h \mid \omega_1) dh$$

$$= \int_{-\infty}^{\theta_{12}} \varphi_1(\mathrm{s}) e^{(1-s)h} p(g = h \mid \omega_1) dh = \varphi_1(\mathrm{s}) \int_{-\infty}^{\theta_{12}} e^{(1-s)h} p(g = h \mid \omega_1) dh$$

$$= e^{-\mu(s)} \int_{-\infty}^{\theta_{12}} e^{(1-s)h} p(g = h \mid \omega_1) dh$$

$$s \le 1 \longrightarrow (1-s)h \le (1-s)\theta_{12}$$

$$\longrightarrow e^{(1-s)h} \le e^{(1-s)\theta_{12}}$$

$$\mathrm{P}_2(\mathrm{e}) \le e^{-\mu(s)+(1-s)\theta_{12}} \int_{-\infty}^{\theta_{12}} p(g = h \mid \omega_1) dh$$

$$\mathrm{P}_2(\mathrm{e}) < e^{-\mu(s)+(1-s)\theta_{12}}$$

当 θ_{12}固定时, 存在最佳的s使上界最小化。

§ 4.4 错误率界限的理论估计

Chernoff界限

$P_1(e)$ 和$P_2(e)$的同时最小化

$$\begin{cases} P_1(e) \le e^{-\mu(s)-s\theta_{12}} & s \ge 0 \\ P_2(e) < e^{-\mu(s)+(1-s)\theta_{12}} & s \le 1 \end{cases}$$

s的选择：

$$\begin{cases} 0 \le s \le 1 \\ \dfrac{\partial}{\partial s}\left(e^{-\mu(s)-s\theta_{12}}\right) = 0 \\ \dfrac{\partial}{\partial s}\left(e^{-\mu(s)+(1-s)\theta_{12}}\right) = 0 \end{cases} \Rightarrow \quad \frac{d\mu(s)}{ds} = -\theta_{12} \qquad 0 \le s \le 1$$

$\Downarrow$ 其解s^*使上界最小化

Chernoff上界 (C上界)

§ 4.4 错误率界限的理论估计

Chernoff界限

P(e)的最小上界

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

$$= P(\omega_1)e^{-\mu(s)-s\theta_{12}}\int_{\theta_{12}}^{\infty} p(g=h\,|\,\omega_1)dh + P(\omega_2)e^{-\mu(s)+(1-s)\theta_{12}}\int_{-\infty}^{\theta_{12}} p(g=h\,|\,\omega_1)dh$$

设采用0-1损失函数，则 $\theta_{12} = \ln\dfrac{P(\omega_1)}{P(\omega_2)}$

$$e^{-\mu(s)-s\theta_{12}} = e^{-\mu(s)-s\ln\frac{P(\omega_1)}{P(\omega_2)}} = e^{-\mu(s)}\left(\frac{P(\omega_1)}{P(\omega_2)}\right)^{-s}$$

$$e^{-\mu(s)+(1-s)\theta_{12}} = e^{-\mu(s)+(1-s)\ln\frac{P(\omega_1)}{P(\omega_2)}} = e^{-\mu(s)}\left(\frac{P(\omega_1)}{P(\omega_2)}\right)^{1-s}$$

§ 4.4 错误率界限的理论估计

Chernoff界限

P(e)的最小上界

$$\mathrm{P(e)}=\mathrm{P}(\omega_1)e^{-\mu(s)-s\theta_{12}}\int_{\theta_{12}}^{\infty}p(g=h\,|\,\omega_1)dh$$

$$+\mathrm{P}(\omega_2)e^{-\mu(s)+(1-s)\theta_{12}}\int_{-\infty}^{\theta_{12}}p(g=h\,|\,\omega_1)dh$$

$$=\mathrm{P}(\omega_1)\left(\frac{\mathrm{P}(\omega_1)}{\mathrm{P}(\omega_2)}\right)^{-s}e^{-\mu(s)}\int_{\theta_{12}}^{\infty}p(g=h\,|\,\omega_1)dh$$

$$+\mathrm{P}(\omega_2)\left(\frac{\mathrm{P}(\omega_1)}{\mathrm{P}(\omega_2)}\right)^{1-s}e^{-\mu(s)}\int_{-\infty}^{\theta_{12}}p(g=h\,|\,\omega_1)dh$$

$$=\left(\mathrm{P}(\omega_1)\right)^{1-s}\left(\mathrm{P}(\omega_2)\right)^{s}e^{-\mu(s)}\int_{-\infty}^{\infty}p(g=h\,|\,\omega_1)dh$$

$$\mathrm{P(e)}\le\left(\mathrm{P}(\omega_1)\right)^{1-s}\left(\mathrm{P}(\omega_2)\right)^{s}e^{-\mu(s)},\qquad 0\le s\le 1$$

在***s****处达到最小

§ 4.4 错误率界限的理论估计

Bhattacharyya界限

两类问题 $\begin{cases} \omega_1\text{:} & p(\boldsymbol{X} \mid \omega_1) \\ \omega_2\text{:} & p(\boldsymbol{X} \mid \omega_2) \end{cases}$

最小错误率判决

$$\begin{cases} P(\omega_1 \mid \boldsymbol{X}) > P(\omega_2 \mid \boldsymbol{X}) \Rightarrow & \boldsymbol{X} \in \omega_1 \\ P(\omega_1 \mid \boldsymbol{X}) < P(\omega_2 \mid \boldsymbol{X}) \Rightarrow & \boldsymbol{X} \in \omega_2 \end{cases}$$

当接收样本为$\boldsymbol{X}$时

正确分类的概率为 $Max\{P(\omega_1 \mid \boldsymbol{X}), P(\omega_2 \mid \boldsymbol{X})\}$ ➡ a

错误分类的概率为 $Min\{P(\omega_1 \mid \boldsymbol{X}), P(\omega_2 \mid \boldsymbol{X})\}$ ➡ b

$$\left(P(e \mid \boldsymbol{X}) = 1 - Max\{P(\omega_1 \mid \boldsymbol{X}), P(\omega_2 \mid \boldsymbol{X})\}\right)$$

则由公式 $\sqrt{ab} \geq b$，当$a \geq b > 0$时，有：

$$P(e \mid \boldsymbol{X}) \leq \sqrt{Max\{P(\omega_1 \mid \boldsymbol{X}), P(\omega_2 \mid \boldsymbol{X})\} \cdot Min\{P(\omega_1 \mid \boldsymbol{X}), P(\omega_2 \mid \boldsymbol{X})\}}$$

$$= \sqrt{P(\omega_1 \mid \boldsymbol{X}) P(\omega_2 \mid \boldsymbol{X})}$$

§ 4.4 错误率界限的理论估计

Bhattacharyya界限

分类错误率 $\mathrm{P(e)} = \int_{\boldsymbol{E}_d} P(e \mid \boldsymbol{X}) p(\boldsymbol{X}) d\boldsymbol{X}$

$$\begin{aligned}
\mathrm{P(e)} &\le \int_{\boldsymbol{E}_d} \sqrt{P(\omega_1 \mid \boldsymbol{X}) P(\omega_2 \mid \boldsymbol{X})}\, p(\boldsymbol{X}) d\boldsymbol{X} \\
&= \int_{\boldsymbol{E}_d} \sqrt{\frac{P(\omega_1) p(\boldsymbol{X} \mid \omega_1)}{p(\boldsymbol{X})} \frac{P(\omega_2) p(\boldsymbol{X} \mid \omega_2)}{p(\boldsymbol{X})}}\, p(\boldsymbol{X}) d\boldsymbol{X} \\
&= \sqrt{P(\omega_1) P(\omega_2} \int_{\boldsymbol{E}_d} \sqrt{p(\boldsymbol{X} \mid \omega_1) p(\boldsymbol{X} \mid \omega_2)}\, d\boldsymbol{X}
\end{aligned}$$

定义 $J_B = -\ln \int_{\boldsymbol{E}_d} \sqrt{p(\boldsymbol{X} \mid \omega_1) p(\boldsymbol{X} \mid \omega_2)}\, d\boldsymbol{X}$ **-Bhattacharyya系数**

$\mathrm{P(e)} \le \sqrt{\mathrm{P}(\omega_1)\mathrm{P}(\omega_2)} \exp(-J_B)$ **-B上界**

§ 4.4 错误率界限的理论估计

B上界和C上界之间的关系

$$J_B = -\ln\int_{E_d}\sqrt{p(\boldsymbol{X}\mid\omega_1)p(\boldsymbol{X}\mid\omega_2)}\,d\boldsymbol{X} = -\ln\int_{E_d}\sqrt{\frac{p(\boldsymbol{X}\mid\omega_2)}{p(\boldsymbol{X}\mid\omega_1)}}\,p(\boldsymbol{X}\mid\omega_1)\,d\boldsymbol{X}$$

$$\frac{p(\boldsymbol{X}\mid\omega_2)}{p(\boldsymbol{X}\mid\omega_1)} = e^{h(\boldsymbol{X})} \quad\Rightarrow\quad \sqrt{\frac{p(\boldsymbol{X}\mid\omega_2)}{p(\boldsymbol{X}\mid\omega_1)}} = e^{\frac{1}{2}h(\boldsymbol{X})}$$

$$J_B = -\ln\int_{E_d} e^{\frac{1}{2}h(\boldsymbol{X})}\,p(\boldsymbol{X}\mid\omega_1)\,d\boldsymbol{X} = -\ln\left(\varphi\left(\frac{1}{2}\right)\right) = \mu\left(\frac{1}{2}\right)$$

$$\mathrm{P(e)} \le \left(\mathrm{P}(\omega_1)\right)^{1-s}\left(\mathrm{P}(\omega_2)\right)^{s} e^{-\mu(s)}$$ **-C上界**

$$= \sqrt{\mathrm{P}(\omega_1)\mathrm{P}(\omega_2)}\,e^{-\mu\left(\frac{1}{2}\right)}$$

$$= \sqrt{\mathrm{P}(\omega_1)\mathrm{P}(\omega_2)}\,\exp(-J_B)$$ **-B上界**

§ 4.5 分类器错误率的实验估计

样本
- 训练样本 ➡ 设计分类器
- 检验样本 ➡ 确定分类器的错误率

- 检验样本集应该独立于训练样本集
- 检验样本彼此间应该统计独立

检验样本的个数 n

误判样本的个数 k

检验样本的采集方式
- 随机抽取
- 选择抽取

§ 4.5 分类器错误率的实验估计

■ 随机抽取

✓ 检验样本的个数 ***n***

✓ 误判样本的个数 ***k***

误判概率服从二项分布 $P_n(k)=\binom{n}{k}P^k(e)(1-P(e))^{n-k}$

最大似然估计

$$\frac{\partial}{\partial P(e)}\ln P_n(k)=\frac{1}{P_n(k)}\frac{\partial P_n(k)}{\partial P(e)}$$

$$\frac{\partial P_n(k)}{\partial P(e)}=P_n(k)\frac{\partial}{\partial P(e)}\ln P_n(k)$$

$$=P_n(k)\frac{\partial}{\partial P(e)}(k\ln P(e)+(n-k)\ln(1-P(e)))$$

$$=P_n(k)\left(\frac{k}{P(e)}-\frac{n-k}{1-P(e)}\right)=0 \quad \Rightarrow \quad \hat{P}(e)=\frac{k}{n}$$

§ 4.5 分类器错误率的实验估计

■ 随机抽取

$$\hat{P}(e)=\frac{k}{n}$$

估计的均值和方差

已知：$E(k)=nP(e)$

$$\sigma_k^2=D(k)=nP(e)(1-P(e))$$

$$E(\hat{P}(e))=E\left(\frac{k}{n}\right)=\frac{E(k)}{n}=P(e)$$ **-无偏估计**

$$\sigma_{\hat{P}(e)}^2=D(\hat{P}(e))=D\left(\frac{k}{n}\right)=\frac{\sigma_k^2}{n^2}=\frac{1}{n}P(e)(1-P(e))$$

§ 4.5 分类器错误率的实验估计

■ 选择抽取（两类情况）

✓ 检验样本的个数 ***n*：** $n_1 = P(\omega_1)n \quad n_2 = P(\omega_2)n$

✓ 误判样本的个数 ***k*:** $k_1 \quad k_2$

误判概率

$$P_n(k_1, k_2) = P_{n_1}(k_1)P_{n_2}(k_2)$$

$$= \binom{n_1}{k_1} P_1^{k_1}(e)(1 - P_1(e))^{n_1 - k_1} \binom{n_2}{k_2} P_2^{k_2}(e)(1 - P_2(e))^{n_2 - k_2}$$

最大似然估计 $\hat{\mathrm{P}}_1(\mathrm{e}) = \dfrac{k_1}{n_1} \qquad \hat{\mathrm{P}}_2(\mathrm{e}) = \dfrac{k_2}{n_2}$

$$\hat{\mathrm{P}}(\mathrm{e}) = \mathrm{P}(\omega_1)\hat{\mathrm{P}}_1(\mathrm{e}) + \mathrm{P}(\omega_2)\hat{\mathrm{P}}_2(\mathrm{e}) = \mathrm{P}(\omega_1)\frac{k_1}{n_1} + \mathrm{P}(\omega_2)\frac{k_2}{n_2}$$

$$E(\hat{P}(e)) = P(\omega_1)E(\hat{P}_1(e)) + P(\omega_2)E(\hat{P}_2(e)) = P(\omega_1)\frac{E(k_1)}{n_1} + P(\omega_2)\frac{E(k_2)}{n_2}$$

$$= P(\omega_1)\frac{n_1 P_1(e)}{n_1} + P(\omega_2)\frac{n_2 P_2(e)}{n_2} = P(\omega_1)P_1(e) + P(\omega_2)P_2(e) = P(e)$$

§ 4.5 分类器错误率的实验估计

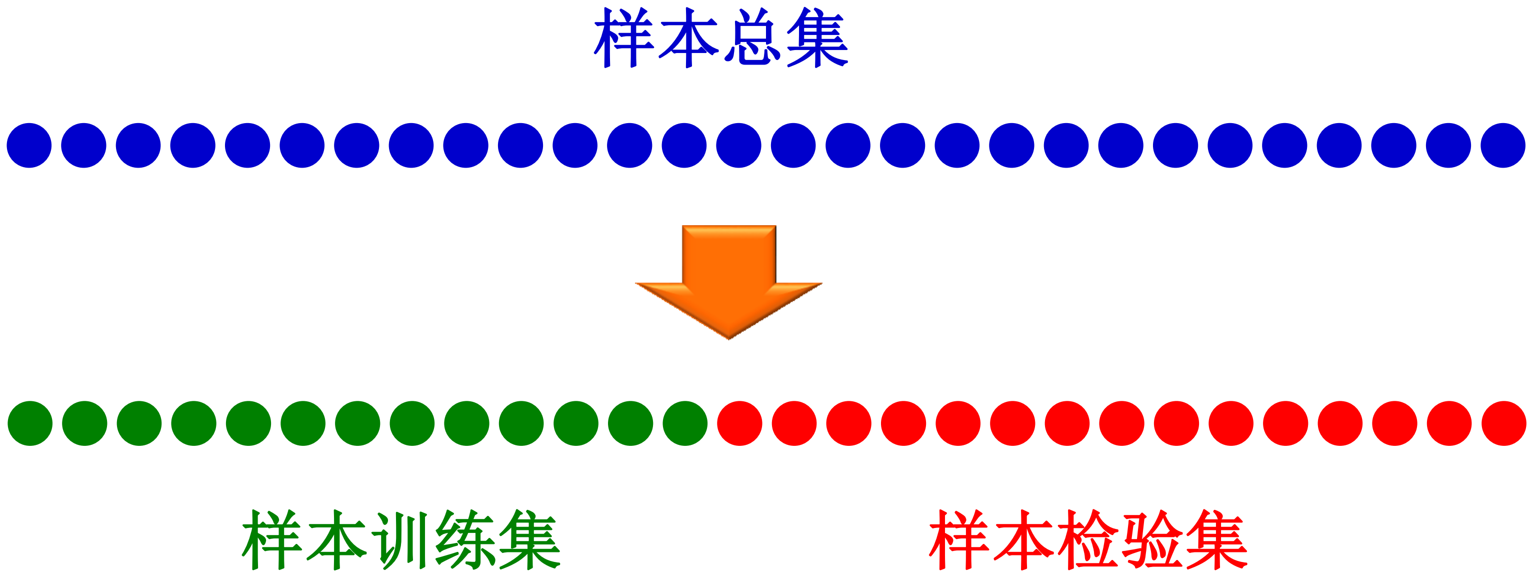
- 有限样本情况下分类器错误率的实验估计
  - ✓ 充分利用已知的有限个样本得到性能好的分类器；
  - ✓ 充分利用已知的有限个样本给出可靠的错误率估计。

样本的分组方式：
- 两分法
- 留一法
- 分组交替法
- 重复使用法

§ 4.5 分类器错误率的实验估计

■ 有限样本情况下分类器错误率的实验估计

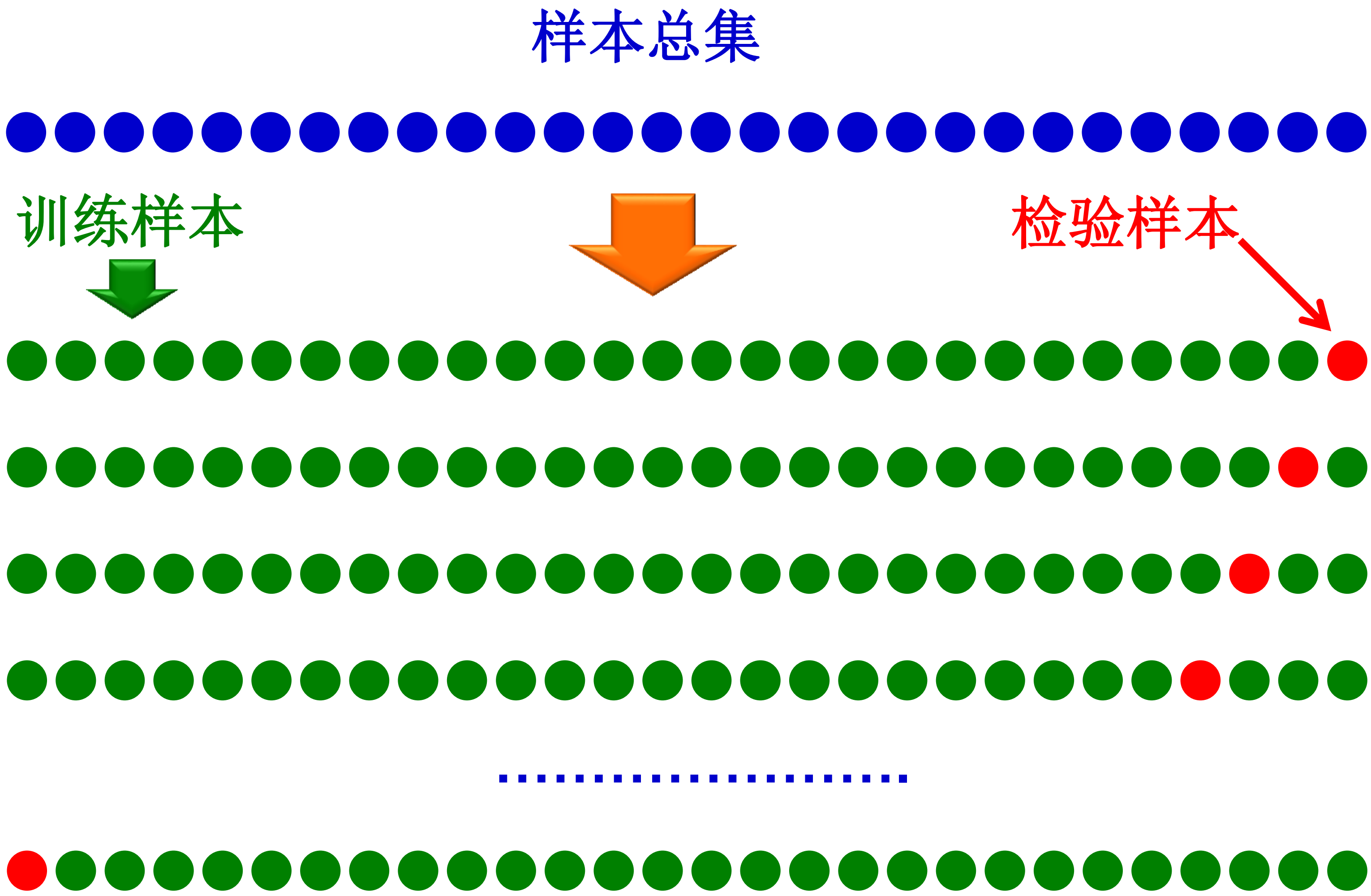
两分法



§ 4.5 分类器错误率的实验估计

■ 有限样本情况下分类器错误率的实验估计

留一法

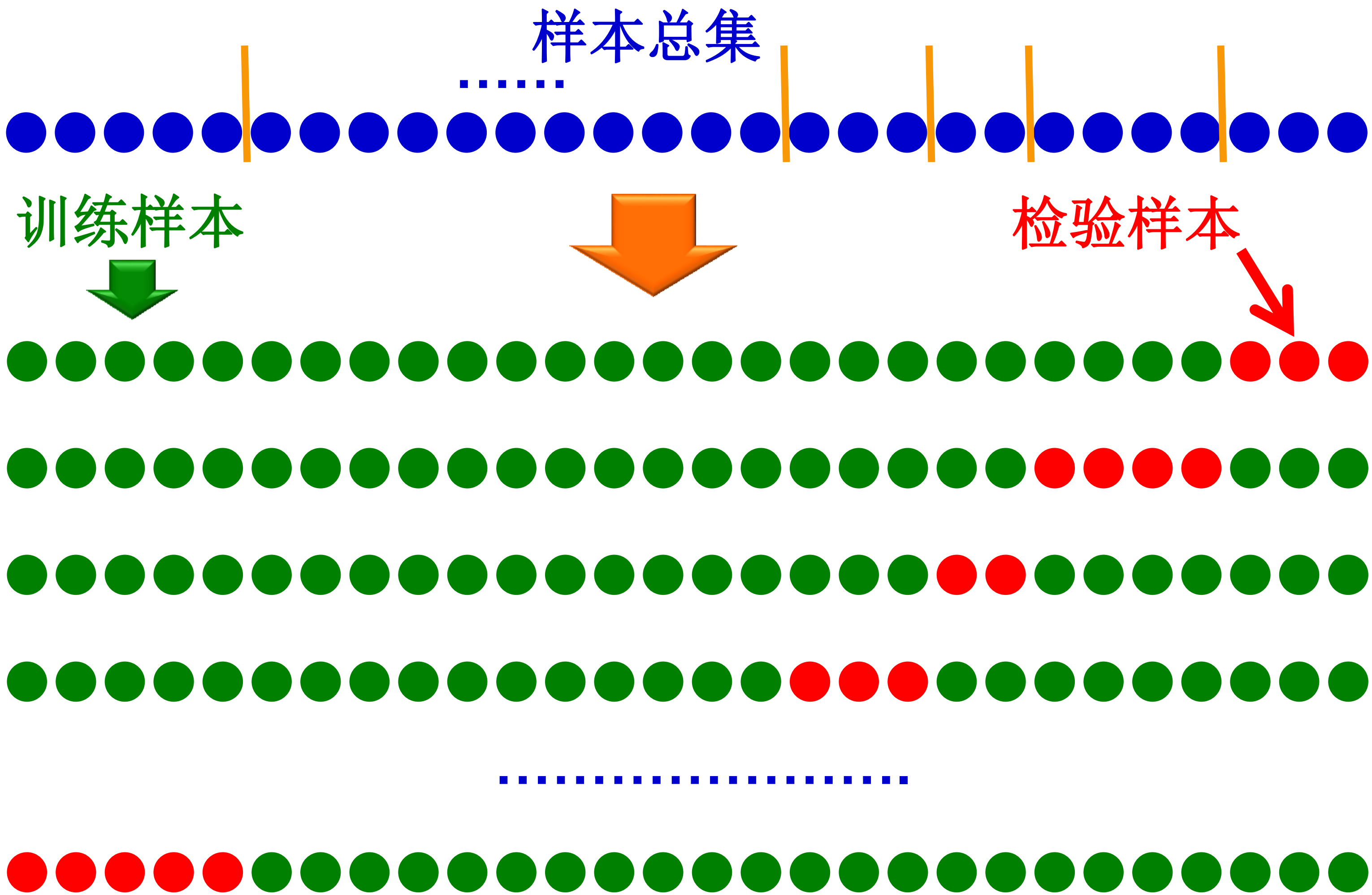


§ 4.5 分类器错误率的实验估计

■ 有限样本情况下分类器错误率的实验估计

$$\hat{P}(e)=\sum_{i=1}^{m}\frac{n_i}{n}\frac{k_i}{n_i}=\frac{\sum_{i=1}^{m}k_i}{n}$$

分组交替法



§ 4.5 分类器错误率的实验估计

■ 有限样本情况下分类器错误率的实验估计

分组交替法

分组数

第 i 个分组中被误分的样本数

$$\hat{P}(e)=\sum_{i=1}^{m}\frac{n_i}{n}\frac{k_i}{n_i}=\frac{\sum_{i=1}^{m}k_i}{n}$$

第i个分组包含的样本数

§ 小结

- 正态分布下的分类器错误率
- 高维空间中各维统计独立情况下的分类器错误率
- 分类器错误率界限的理论估计
- 分类器错误率的实验估计

若干图片材料取自网络，特此致谢。

谢谢聆听！

中国科学技术大学