Machine Learning and Pattern Recognition

东北大学“智能+X”新工科课程系列

机器学习与模式识别

东北大学 信息科学与工程学院
智能感知与机器人研究所
陈东岳

Bayesian Classifier

贝叶斯分类器

CHAPTER ONE

分类器基础

Foundation of Classifier

一、分类器基础

分类器数学模型

- **从系统角度看，分类器是一个输入输出系统：**
  - 输入为样本的特征$\boldsymbol{x}$
  - 输出为样本的类别ω
- **从数学角度看，分类器是一个函数：**
  - 用于拟合样本特征与类别之间的映射关系
- **分类器的数学描述：**

样本特征向量

$$y = h(\boldsymbol{x})$$

类别预测结果

分类器映射函数

一、分类器基础

分类准则的函数形式

- **最大后验概率分类准则**

$$h_a(\boldsymbol{x}) = \underset{\omega_i}{\operatorname{argmax}}\, P(\omega_i|\boldsymbol{x})$$

- **最小错误概率分类准则**

$$h_e(\boldsymbol{x}) = \underset{\omega_i}{\operatorname{argmin}}\, P_e(\omega_i|\boldsymbol{x})$$

- **最小风险分类准则**

$$h_r(\boldsymbol{x}) = \underset{\omega_i}{\operatorname{argmin}}\, \ell_i = \min_{\omega_i} \sum_{j=1}^{M} \lambda_{ji} P(\omega_j|\boldsymbol{x})$$

分类准则直接给出
分类器映射函数
没有体现设计的过程

一、分类器基础

分类器设计

- **使用分类器**
  - 已知：分类器$h(\cdot)$和输入样本$\boldsymbol{x}$
  - 求：分类器的输出$y = h(\boldsymbol{x})$
- **设计分类器**
  - 已知：训练样本$X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N\}$
    真实输出$Y = \{y_1, y_2, \dots, y_N\}$
  - 求：一个最优的分类器$h^*(\cdot)$
- **分类器设计的本质是一个优化问题**

$$h^* = \underset{h}{\operatorname{argmin}}\, \mathcal{L}(h; X, Y)$$

分类器设计准则不直接定义分类器而是提出设计分类器的方法

一、分类器基础

分类器设计的优化问题

- **分类器设计的本质是一个优化问题**

$$h^* = \underset{h}{\operatorname{argmin}}\, \mathcal{L}(h; X, Y)$$

- $\mathcal{L}(h;X,Y)$是优化问题的目标函数
- X,Y是目标函数的参数
- 分类器函数$h(\cdot)$是优化问题的自变量
- $h(\cdot)$的数学形式是一个函数，很难直接被优化
- 优化过程不改变$h(\cdot)$的函数形式，只对其参数进行优化
- $h(\cdot)$的函数形式定义“假设空间”，参数值定义具体的“假设”

CHAPTER TWO

最小平均错误概率分类器

Minimum Average Error Probability Classifier

二、最小平均错误概率分类器

▶ 分类器模型

- **平均错误概率$P_E(h)$**

$$P_E(h) = \mathbb{E}_{\boldsymbol{x}}[P_e(h(\boldsymbol{x})|\boldsymbol{x})] = \int P_e(h(\boldsymbol{x})|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}$$

- $P_e(h(\boldsymbol{x})|\boldsymbol{x})$是分类器$h(\boldsymbol{x})$将样本$\boldsymbol{x}$错误分类的概率

- **最小平均错误概率分类器**

$$h_E(\boldsymbol{x}) = \underset{h}{\operatorname{argmin}}\, P_E(h) = \underset{h}{\operatorname{argmin}}\left[\int P_e(h(\boldsymbol{x})|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}\right]$$

二、最小平均错误概率分类器

分类器模型

- **优化求解思路**

  - 如果存在一个$h(\boldsymbol{x})$，对于每个具体的$\boldsymbol{x}$都能最小化$P_e(h(\boldsymbol{x})|\boldsymbol{x})p(\boldsymbol{x})$
  - 则$h(\boldsymbol{x})$就能够最小化$P_E(h)$，即有$h_E(\boldsymbol{x}) = h(\boldsymbol{x})$
  - 显然，最小错误概率分类器$h_e(\boldsymbol{x})$能够最小化$P_e(h(\boldsymbol{x})|\boldsymbol{x})p(\boldsymbol{x})$

$$h_E(\boldsymbol{x}) = \min_{\omega_i}[P_e(\omega_i|\boldsymbol{x})p(\boldsymbol{x})] \Leftrightarrow \min_{\omega_i}[P_e(\omega_i|\boldsymbol{x})] = h_e(\boldsymbol{x})$$

  - 根据贝叶斯定理： $\underset{\omega_i}{\operatorname{argmin}}[P_e(\omega_i|\boldsymbol{x})p(\boldsymbol{x})] = \underset{\omega_i}{\operatorname{argmin}}[P_e(\boldsymbol{x}|\omega_i)p(\omega_i)]$

- **结论**

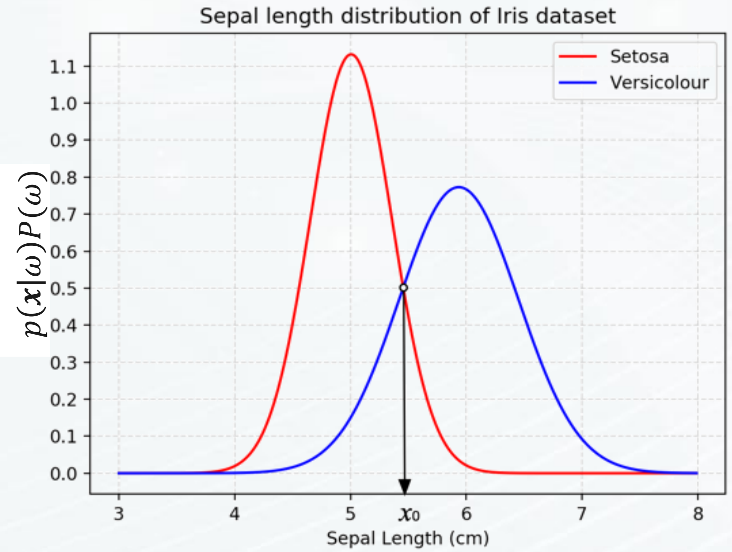
  - 最小错误概率分类器$h_e(\boldsymbol{x})$就是最小平均错误概率分类器$h_E(\boldsymbol{x})$

- **疑问： $h_e(x)$的具体数学形式是什么？**

二、最小平均错误概率分类器

▶ 例子

- **Iris数据库中的两类鸢尾花Setosa(ω_1)和Versicolour (ω_2)**
  - 萼片长度为特征x；设两类各自均服从正态分布$\mathcal{N}(\mu_1,\delta_1^2)$和$\mathcal{N}(\mu_2,\delta_2^2)$；
  - $P(\omega_1)=P(\omega_2)=0.5$



最小错误概率分类器

当$x=x_0$时

$p(x|\omega_1)P(\omega_1)=p(x|\omega_2)P(\omega_2)$

当$x<x_0$时：$p_e(\omega_1|x)<p_e(\omega_2|x)$

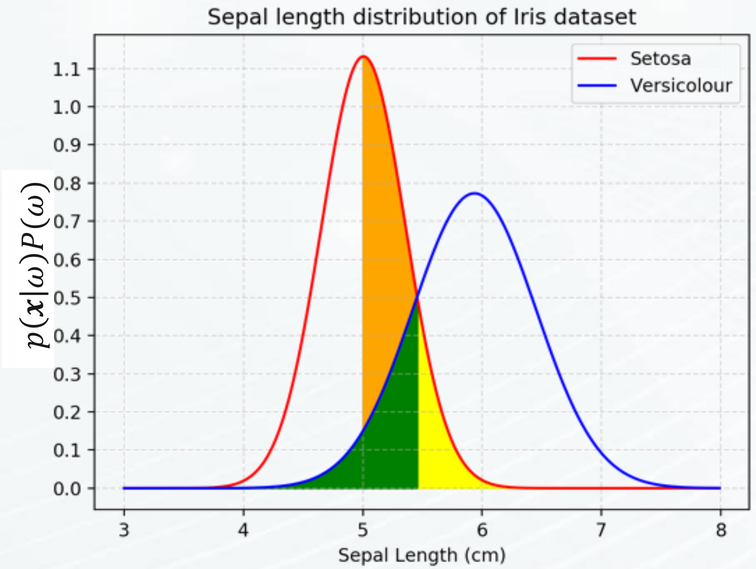
当$x>x_0$时：$p_e(\omega_2|x)<p_e(\omega_1|x)$

$$h_e(x)=\begin{cases}\omega_1, & x<x_0\\ \omega_2, & x\geq x_0\end{cases}$$

二、最小平均错误概率分类器

例子

- Iris数据库中的两类鸢尾花Setosa(ω_1)和Versicolour (ω_2)
  - 绿色面积：$\int_{-\infty}^{x_0} p(\omega_2|x)P(x)dx = \int_{-\infty}^{x_0} p(x|\omega_2)P(\omega_2)dx$
  - 黄色面积：$\int_{x_0}^{\infty} p(\omega_1|x)P(x)dx = \int_{x_0}^{\infty} p(x|\omega_1)P(\omega_1)dx$



最小平均错误概率分类器

当分类阈值 $T = x_0$时

绿色面积为ω_2被错分为ω_1的概率；

黄色面积为ω_1 被错分为ω_2的概率；

两者之和为$h_E(x)$的平均错误概率

当$T < x_0$， 橙色面积为增加的错误概率，

因此$T = x_0$时平均错误概率最小

CHAPTER THREE

最小平均风险分类器

Minimum Average Risk Classifier

三、最小平均风险分类器

分类器模型

- **平均风险$R(h)$**

$$R(h)=\sum_{j=1}^{M}\sum_{i=1}^{M}\lambda_{ij}P_{ij}(h)$$

- λ_{ij}是将ω_i类样本分给ω_j类的风险

- P_{ij} (h)是分类器$h(\cdot)$将ω_i类样本分给ω_j类的概率

$$P_{ij}(h)=\int_{h(\boldsymbol{x})=\omega_j}P(\omega_i|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}$$

- **分类器函数**

$$h_R(\boldsymbol{x})=\underset{h}{\operatorname{argmin}}\,R(h)$$

三、最小平均风险分类器

分类器模型

$$\boxed{\mathcal{L}_j = \sum_{i=1}^{M} \lambda_{ij} P_{ij}(h)\ ;\quad P_{ij}(h) = \int_{h(\boldsymbol{x})=\omega_j} P(\omega_i|\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}}$$

- **求解思路**

  - 根据$h(\boldsymbol{x})$的类别预测结果，将论域空间分为M个区域$\Omega_j, j=1,\ldots,M$
  - $h(\boldsymbol{x})=\omega_j, \forall \boldsymbol{x}\in\Omega_j$，区域$\Omega_j$对应的平均风险记为$\mathcal{L}_j$

$$R(h) = \sum_{j=1}^{M} \mathcal{L}_j = \sum_{j=1}^{M} \left[\int_{\Omega_j} \left(\sum_{i=1}^{M} \lambda_{ij} P(\omega_i|\boldsymbol{x}) \right) p(\boldsymbol{x}) d\boldsymbol{x} \right] = \sum_{j=1}^{M} \left[\int_{\Omega_j} \ell_j(\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x} \right]$$

  - $\ell_j(\boldsymbol{x})$是将样本$\boldsymbol{x}$分为第j类的决策风险；
  - 假设在分类器$h(\boldsymbol{x})$所划出的区域Ω_j上，$\ell_j(\boldsymbol{x})<\ell_i(\boldsymbol{x}), \forall i\neq j$
  - 则该分类器$h(\boldsymbol{x})$能够最小化$R(h)$，即有$h_R(x)=h(x)$
  - 显然，基于最小风险分类准则的分类器$h_r(x)$能够满足这一要求

$$h_R(\boldsymbol{x}) = \min_{\omega_j} [\ell_j(\boldsymbol{x}) p(\boldsymbol{x})] \Leftrightarrow \min_{\omega_j} [\ell_j(\boldsymbol{x})] = h_r(\boldsymbol{x})$$

三、最小平均风险分类器

分类器模型

- **以二分类问题为例**

  - $\ell_1(\boldsymbol{x}) = \lambda_{11}P(\omega_1|\boldsymbol{x}) + \lambda_{21}P(\omega_2|\boldsymbol{x})$
  - $\ell_2(\boldsymbol{x}) = \lambda_{12}P(\omega_1|\boldsymbol{x}) + \lambda_{22}P(\omega_2|\boldsymbol{x})$
  - 则在分割区域Ω_1和Ω_2的分界点$x = x_r$处有$\ell_1(\boldsymbol{x}) = \ell_2(\boldsymbol{x})$;

$$\lambda_{11}P(\omega_1|\boldsymbol{x}) + \lambda_{21}P(\omega_2|\boldsymbol{x}) = \lambda_{12}P(\omega_1|\boldsymbol{x}) + \lambda_{22}P(\omega_2|\boldsymbol{x})$$
$$(\lambda_{21} - \lambda_{22})p(\boldsymbol{x}|\omega_2)P(\omega_2) = (\lambda_{12} - \lambda_{11})p(\boldsymbol{x}|\omega_1)P(\omega_1)$$

  - 在Ω_1区域，$\ell_1(\boldsymbol{x}) < \ell_2(\boldsymbol{x})$

$$(\lambda_{21} - \lambda_{22})p(\boldsymbol{x}|\omega_2)P(\omega_2) < (\lambda_{12} - \lambda_{11})p(\boldsymbol{x}|\omega_1)P(\omega_1)$$

  - 在Ω_2区域，$\ell_1(\boldsymbol{x}) > \ell_2(\boldsymbol{x})$

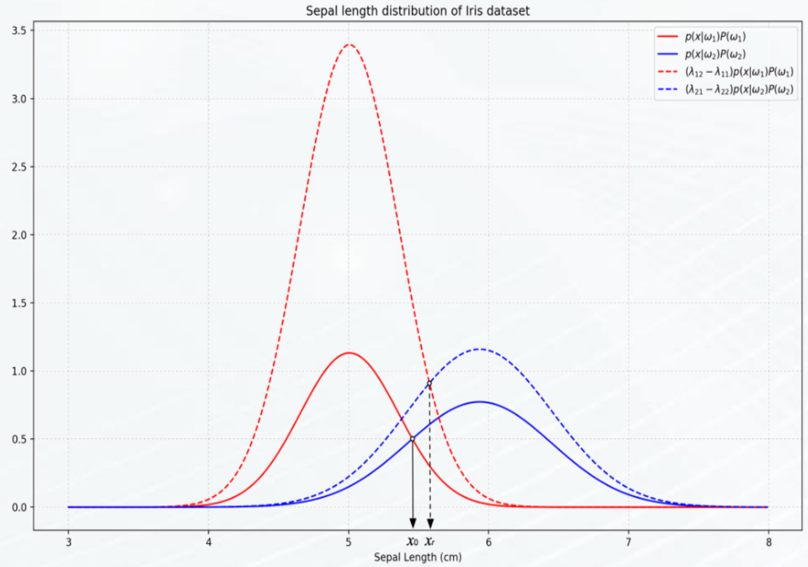
$$(\lambda_{21} - \lambda_{22})p(\boldsymbol{x}|\omega_2)P(\omega_2) > (\lambda_{12} - \lambda_{11})p(\boldsymbol{x}|\omega_1)P(\omega_1)$$

三、最小平均风险分类器

例子

- **Iris数据库中的两类鸢尾花Setosa(ω_1)和Versicolour (ω_2)**

  - 假设λ_{12}=3，λ_{21}=1.5，λ_{11}=λ_{22}=0



把ω_1类误判为ω_2类风险系数为 $\lambda_{12}=3$;

把ω_2类误判为ω_1类风险系数为 $\lambda_{21}=1.5$;

相比于$h_E(x)$，$h_R(x)$的分类阈值T从x_0移动到了x_r;
$h_R(x)$更倾向于将x判为ω_1类。

例题

例题——两种鸢尾花分类

- 某花卉工厂在某岛上采集了两类鸢尾花，其中：
  - ω_1类600朵，ω_2类400朵，花萼长度分别服从正态分布：
    - $P(x|\omega_1) = \mathcal{N}(\mu_1, \sigma_1^2)$, $\mu_1 = 5\text{ cm}, \sigma_1 = 1\text{ cm}$;
    - $P(x|\omega_2) = \mathcal{N}(\mu_2, \sigma_2^2)$, $\mu_2 = 7\text{ cm}, \sigma_2 = 2\text{ cm}$;

- 花卉工厂将鸢尾花分类卖给岛上的花店，如果出现分类错误：
  - ω_1类被错分为ω_2类，每朵罚款3元，
  - ω_2类被错分为ω_1类，每朵罚款1元。

- 请根据最小平均错误概率准则和最小平均风险准则分别设计两个分类器，并告知花卉工厂工人相应的分类阈值。

例题——两种鸢尾花分类

最小平均错误概率分类器

- 根据两类各自的类条件概率密度函数，可知：

$$P(x|\omega_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right], i = 1 \text{ or } 2$$

- 根据最小平均错误概率分类器$h_E(x)$的分类阈值x_0，为以下方程的解，其中$P(\omega_1) = 0.6, P(\omega_2) = 0.4$

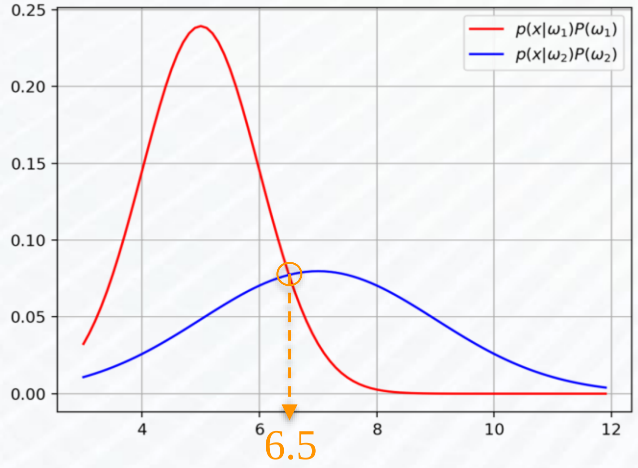
$$p(x|\omega_1)P(\omega_1) = p(x|\omega_2)P(\omega_2)$$

$$\Rightarrow \frac{0.6}{\sqrt{2\pi}} \exp\left[-\frac{(x-5)^2}{2\times 1}\right] = \frac{0.4}{2\sqrt{2\pi}} \exp\left[-\frac{(x-7)^2}{2\times 4}\right]$$

$$\Rightarrow \ln 3 - \frac{(x-5)^2}{2} = -\frac{(x-7)^2}{8}$$

$$\Rightarrow x^2 - \frac{26}{3}x + \frac{51-8\ln 3}{3} = 0$$

$$\Rightarrow x_0 = \frac{13}{3} + \frac{2}{3} \times \sqrt{4 + 6\ln 3} \approx 6.5 \text{ cm}$$



例题——两种鸢尾花分类

最小平均风险分类器

- 根据题意可以设决策风险矩阵为:

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} = \begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix}$$

- 最小平均风险分类器阈值x_r为以下方程的解

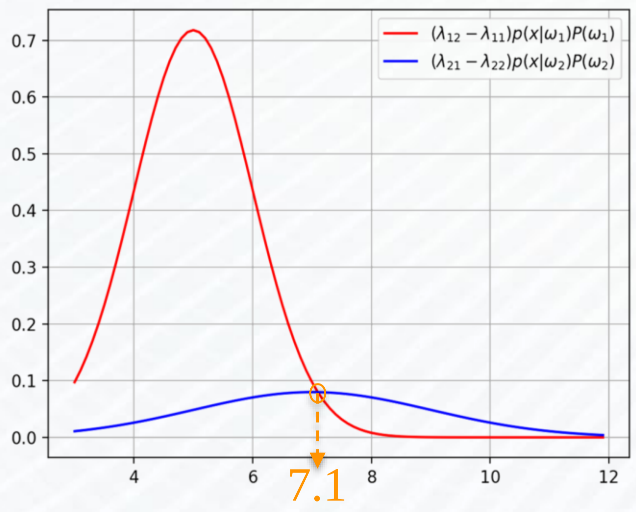
$$(\lambda_{12} - \lambda_{11})p(\boldsymbol{x}|\omega_1)P(\omega_1) = (\lambda_{21} - \lambda_{22})p(\boldsymbol{x}|\omega_2)P(\omega_2)$$

$$\Rightarrow \frac{3\times 0.6}{\sqrt{2\pi}} \exp\left[-\frac{(x-5)^2}{2\times 1}\right] = \frac{1\times 0.4}{2\sqrt{2\pi}} \exp\left[-\frac{(x-7)^2}{2\times 4}\right]$$

$$\Rightarrow \ln 9 - \frac{(x-5)^2}{2} = -\frac{(x-7)^2}{8}$$

$$\Rightarrow x^2 - \frac{26}{3}x + \frac{51-8\ln 9}{3} = 0$$

$$\Rightarrow x_0 = \frac{13}{3} + \frac{2}{3} \times \sqrt{4 + 6\ln 9} \approx 7.1 \text{ cm}$$



总 结


贝叶斯分类器
分类器基础
数学模型
函数形式
设计思想
分类器设计的优化问题
典型分类器
最小平均错误概率分类器
平均错误概率
与最小错误概率分类准则的等价性
正态分布假设下的分类阈值
最小平均风险分类器
平均风险
与最小风险分类准则的等价性
正态分布假设下的分类阈值