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
人工智能系 、智能感知与机器人研究所
陈东岳

Bayesian Decision

贝叶斯决策论基础

CHAPTER ONE

概念

Introduction

一、概念

第一个任务——分类

- 把一个未知的对象 – **模式(Pattern)** – 分给一个正确的类别.
- 该任务被称为分类 **(Classification)**
- 分类依据：基于概率论的解释

概率(Probability) 是分类的基本依据

- 例子：

这是一只猫，因为“ 它是猫的**概率**最大！”

一、概念

▶相关概念与变量描述

- 类别: $\omega_i, i = 1,2, \dots, M$
- 观测样本的特征向量: $\boldsymbol{x} = [x_1, x_2, \dots, x_l]^T$
- 先验(A-priori)概率: $P(\omega_i), i = 1,2, \dots, M$
- 后验(A-posteriori)概率: $P(\omega_i|\boldsymbol{x}), i = 1,2, \dots, M$
- 类条件(Class-conditional)概率: $p(\boldsymbol{x}|\omega_i), i = 1,2, \dots, M$
- 特征概率: $p(\boldsymbol{x}) = \sum_{i=1}^{M} p(\boldsymbol{x}|\omega_i)p(\omega_i)$

全概率公式

一、概念

示例——学生群体

已知

- 类别：$\omega_1 = $ 男性, $\omega_2 = $ 女性
- 特征：$\boldsymbol{x} = $ 身高
- 先验概率： $P(\omega_i) = 0.5, i = 1,2$
- 类条件概率密度函数：$p(\boldsymbol{x}|\omega_i), i = 1,2$

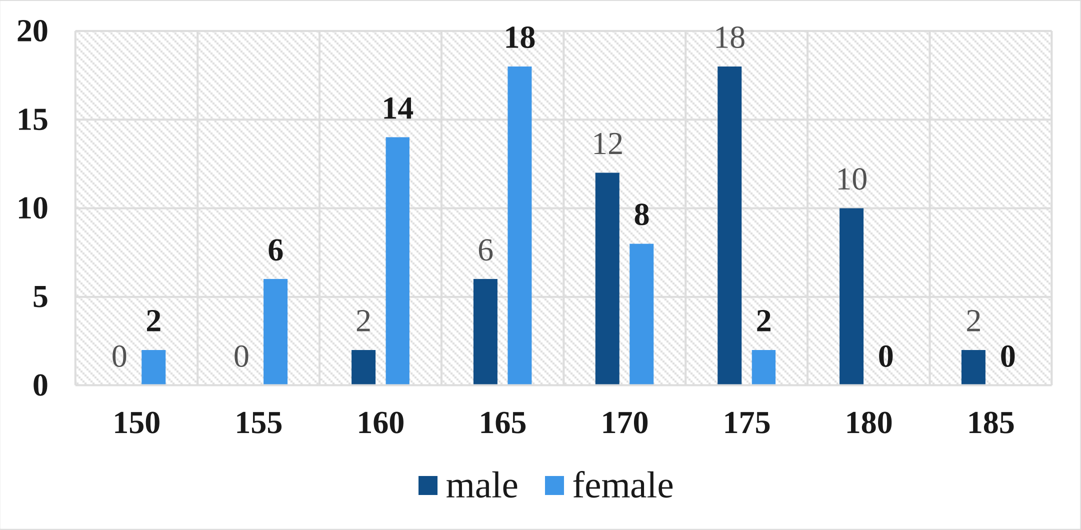
未知

- 特征概率密度函数： $p(\boldsymbol{x})$
- 后验概率: $P(\omega_i|\boldsymbol{x}), i = 1,2$

一、概念

▶ 示例——类条件概率密度函数

- 解释：ω_i类同学身高为$\boldsymbol{x}$的概率$p(\boldsymbol{x}|\omega_i)$
- 取值：$p(\boldsymbol{x}|\omega_i)$可以通过身高分布的直方图近似估计；



一、概念

示例——估计类条件概率

$$p(\boldsymbol{x}|\omega_i) \approx \frac{\omega_i\text{类中身高为}x\text{的人数}}{\omega_i\text{类总人数}}$$

	150	155	160	165	170	175	180	185	总数
男性人数	0	0	2	**6**	12	18	10	2	50
女性人数	2	6	14	**18**	8	2	0	0	50
$p(x\|\omega_1=\text{男})$	0	0	0.04	**0.12**	0.24	0.36	0.20	0.04	1
$p(x\|\omega_2=\text{女})$	0.04	0.12	0.28	**0.36**	0.16	0.04	0	0	1

$$p(165\,|\,\text{男}) \approx \frac{6}{50} = 0.12; \quad p(165\,|\,\text{女}) \approx \frac{18}{50} = 0.36$$

CHAPTER TWO

分类准则

Classification Rules

二、分类准则

▶ 概念辨析

- **分类准则**
  - ➢ —对于一个具体的样本 x ，该如何分类?
  - ➢ 最大后验概率分类准则
  - ➢ 最小错误概率分类准则
  - ➢ 最小风险分类准则
- **分类器设计准则**
  - ➢ 如何设计一个好的分类器$h(x)$?
  - ➢ 最小平均错误概率准则
  - ➢ 最小平均风险准则

二、分类准则

▶ 最大后验概率分类准则

将特征观测值为$\boldsymbol{x}$的样本分给$\omega_i, i = 1,2, \dots, M$中能够使得后验概率$P(\omega_i|\boldsymbol{x})$最大的一类$\omega^*$。

$$\boldsymbol{x} \rightarrow \omega^*: \underset{\omega_i}{\operatorname{argmax}}[P(\omega_i|\boldsymbol{x})]$$

我们最需要了解的**后验概率**
恰恰是
我们不知道的概率☹

二、分类准则

▶ 贝叶斯定理

贝叶斯公式为**类条件概率**与**后验概率**搭建了桥梁

贝叶斯公式

Bayes Formula

$$p(\boldsymbol{x}|\omega_i) \rightarrow P(\omega_i|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\omega_i)P(\omega_i)}{p(\boldsymbol{x})} \rightarrow P(\omega_i|\boldsymbol{x})$$

类条件概率 后验概率

$$p(\boldsymbol{x}) = \sum_{i=1}^{M} p(\boldsymbol{x}|\omega_i)P(\omega_i)$$

全概率公式

三、分类准则

最大后验概率分类准则 (for二分类问题)

- **分类准则**

$$\boldsymbol{x} \rightarrow \omega_1 \;\; if \; P(\omega_1|\boldsymbol{x}) > P(\omega_2|\boldsymbol{x})$$
$$\boldsymbol{x} \rightarrow \omega_2 \;\; if \; P(\omega_2|\boldsymbol{x}) > P(\omega_1|\boldsymbol{x})$$

根据贝叶斯公式:

$$p(\boldsymbol{x}|\omega_1)P(\omega_1) \; (>?<) \; p(\boldsymbol{x}|\omega_2)P(\omega_2)$$

当 $P(\omega_1) = P(\omega_2)$ ：

$$p(\boldsymbol{x}|\omega_1) \; (>?<) \; p(\boldsymbol{x}|\omega_2)$$

最大后验概率准则可以通过比较类条件概率来执行

三、分类准则

最大后验概率分类准则 (for多类问题)

- 当类别数量 $M > 2$时
- 如果：$p(\omega_i|\boldsymbol{x}) > p(\omega_j|\boldsymbol{x}), \forall j \neq i$
  - 将样本$\boldsymbol{x}$分给后验概率最大的类别ω_i
- 如果：$P(\omega_i) = P(\omega_j), \forall i, j,$
  - 只需：$p(\boldsymbol{x}|\omega_i) > p(x|\omega_j), \forall j \neq i$,
  - 将样本$\boldsymbol{x}$分给类条件概率最大的类别ω_i

三、分类准则

▶最小错误概率分类准则

- 将样本$\boldsymbol{x}$分给ω_i类的错误概率定义为：

$$P_e(\omega_i|\boldsymbol{x}) = 1 - P(\omega_i|\boldsymbol{x})$$

- 分类准则

$$\boldsymbol{x} \rightarrow \omega_1 \quad if \; P_e(\omega_1|\boldsymbol{x}) < P_e(\omega_2|\boldsymbol{x})$$

$$\boldsymbol{x} \rightarrow \omega_2 \quad if \; P_e(\omega_2|\boldsymbol{x}) < P_e(\omega_1|\boldsymbol{x})$$

最小错误概率分类准则等价于**最大后验概率分类原则**

三、分类准则

最小风险分类准则

- 欢迎来到真实世界，任意分类行为都必然承担风险；
- 以类别数量$M = 2$为例，可以定义风险矩阵如下：

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

- 其中λ_{ij}表示将ω_i类样本分给ω_j类的“风险值”。
  - 不同的决策可能有不同的风险值；
  - 即使决策是对的，对应的风险值也可能不为零；

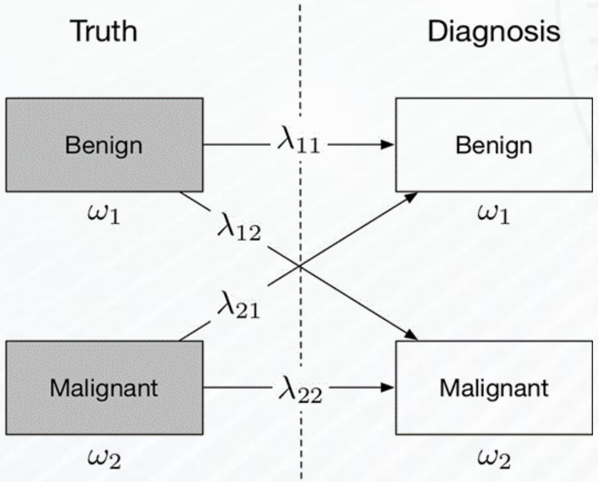
三、分类准则

▶ 最小风险分类准则

- **示例：肿瘤诊断**

良性ω_1 vs 恶性ω_2

- ➢ 恶性误判为良性的风险： λ_{21}
- ➢ 良性误判为恶性的风险： λ_{12}
- ➢ 恶性判为恶性的风险： λ_{22}
- ➢ 良性判为良性的风险： λ_{11}



$$\lambda_{21} > \lambda_{12} > \lambda_{22} > \lambda_{11}$$

三、分类准则

最小风险分类准则

- 将样本 $\boldsymbol{x}$ 分给 ω_1 类的风险为 ℓ_1

$$\ell_1 = \lambda_{11}P(\omega_1|\boldsymbol{x}) + \lambda_{21}P(\omega_2|\boldsymbol{x})$$

- 样本 $\boldsymbol{x}$ 分给 ω_2 类的风险为 ℓ_2

$$\ell_2 = \lambda_{12}P(\omega_1|\boldsymbol{x}) + \lambda_{22}P(\omega_2|\boldsymbol{x})$$

- 样本 $\boldsymbol{x}$ 分给决策风险最小的类别

$$\boldsymbol{x} \to \omega_1 \ \text{if}\ \ell_1 < \ell_2;\ \boldsymbol{x} \to \omega_2 \ \text{if}\ \ell_2 < \ell_1$$

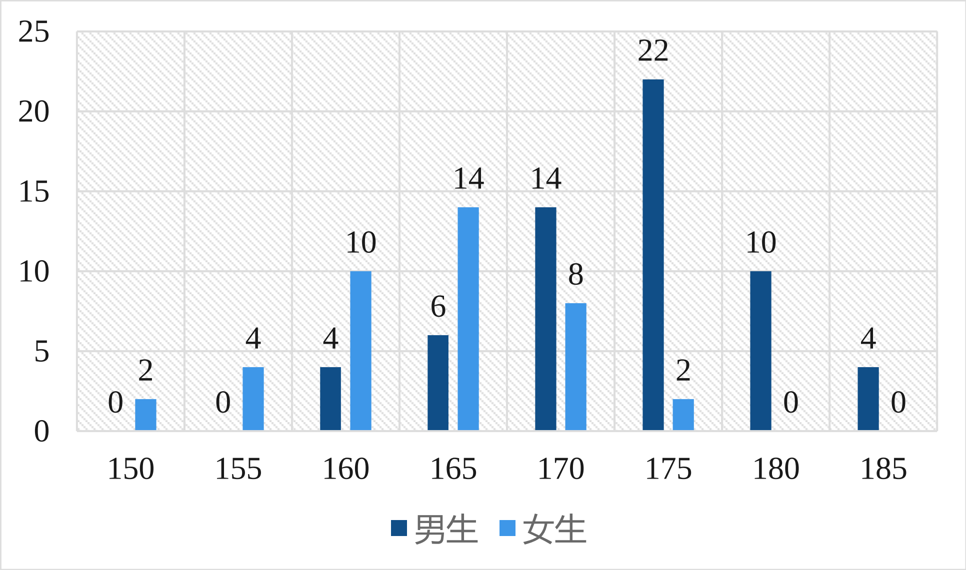
- 当$\lambda_{12} = \lambda_{21} > 0$，且$\lambda_{11} = \lambda_{22} = 0$时：

最小错误概率分类准则等价于**最小风险分类准则**

例题

例题——根据身高预测学生性别

- 东北大学人工智能专业2019级学生共100人
- 其中男生60人，女生40人，具体身高分布如下图所示：



- 若某学生身高为165公分，试用三种分类准则预测该生性别。

例题——根据身高预测学生性别

最大后验概率分类准则

- 设男生为ω_1类，女生为ω_2类，根据身高数据，则有：

$$p(165|\omega_1) \approx \frac{6}{60} = 0.10; \quad p(165|\omega_2) \approx \frac{14}{40} = 0.35$$

- 根据全概率公式：$p(165) = 0.10 \times 0.6 + 0.35 \times 0.4 = 0.20$
- 根据贝叶斯公式，有：

$$P(\omega_1|165) = \frac{p(165|\omega_1)P(\omega_1)}{p(165)} = \frac{0.10 \times 0.6}{0.20} = 0.3$$

$$P(\omega_2|165) = \frac{p(165|\omega_2)P(\omega_2)}{p(165)} = \frac{0.35 \times 0.4}{0.20} = 0.7$$

- 由于 $P(\omega_2|165) > P(\omega_1|165)$ ，根据最大后验概率分类准则。
- 预测该学生为女生。

例题——根据身高预测学生性别

▶ 最小错误概率分类准则

- 后验概率估计结果为$P(\omega_1|165)=0.3$； $P(\omega_2|165)=0.7$
- 将该生预测为男生的错误概率为：

$$P_e(\omega_1|165)=1-P(\omega_1|165)=0.7$$

- 将该生预测为女生的错误概率为：

$$P_e(\omega_2|165)=1-P(\omega_2|165)=0.3$$

- 由于 $P_e(\omega_2|165)<P_e(\omega_1|165)$ ，根据最小错误概率分类准则。
- 预测该学生为女生。

例题——根据身高预测学生性别

最小风险分类准则

- 设预测该学生性别的风险矩阵为：

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

- 预测该生为男生的风险为：
  $\ell_1 = \lambda_{11}P(\omega_1|165) + \lambda_{21}P(\omega_2|165) = 1 \times 0.7 = 0.7$
- 预测该生为女生的风险为：
  $\ell_2 = \lambda_{12}P(\omega_1|165) + \lambda_{22}P(\omega_2|165) = 1 \times 0.3 = 0.3$
- 由于$\ell_2 < \ell_1$，根据最小风险分类准则。
- 预测该生为女生。

总 结

- 贝叶斯决策论
  - 概念
    - 类别
    - 特征
    - 先验概率
    - 后验概率
    - 类条件概率
    - 特征概率
    - 错误概率
    - 决策风险
      - 风险值
      - 风险矩阵
  - 分类准则
    - 公式
      - 贝叶斯公式
      - 全概率公式
    - 准则
      - 最大后验概率分类准则
      - 最小错误概率分类准则
      - 最小风险分类准则

M L & P R

THANK YOU

感谢聆听

MADE BY DONGYUE CHEN