机器学习

Machine learning

# 第二章 贝叶斯学习

## Bayesian Learning

授课人：周晓飞
zhouxiaofei@iie.ac.cn
2022-9-2

课件放映 PDF-〉视图-〉全屏模式

# 第二章 贝叶斯学习

2.1 概述

2.2 贝叶斯决策论

2.3 贝叶斯分类器

2.4 贝叶斯学习与参数估计问题

# 第二章 贝叶斯学习

2.1 概述
2.2 贝叶斯决策论
2.3 贝叶斯分类器
2.4 贝叶斯学习与参数估计问题

# 概述

## 例子：天气预报

### 依赖先验的决策

某地全年 365 天，晴朗 265 天，非晴朗 100 天。A=晴，~A=非晴；

$P(A)=265/365=0.726$，$P(\sim A)=100/365=0.274$；判断明天天气如何？ 计算 P（A），P（~A）。

# 概述

## 例子：天气预报

### 依赖先验的决策

某地全年 365 天，晴朗 265 天，非晴朗 100 天。A=晴，~A=非晴；

$P(A)=265/365=0.726$，$P(\sim A)=100/365=0.274$；判断明天天气如何？ 计算 P（A）， P（~A）。

只有一种预测可能：$P(A)>P(\sim A)$ 即 P（晴天）> P（非晴天）

# 概述

## 例子：天气预报

### 依赖先验的决策

某地全年 365 天，晴朗 265 天，非晴朗 100 天。A=晴，~A=非晴；

$P(A)=265/365=0.726$，$P(\sim A)=100/365=0.274$；判断明天天气如何？ 计算 P（A），P（~A）。

只有一种预测可能：$P(A)>P(\sim A)$ 即 P（晴天）> P（非晴天）

### 增加类别条件概率信息

若增加可观测信息：晴朗（非晴朗）天气前一天特征（是否有晚霞）统计。B=晚霞，~B=无晚霞；

$P(B/A)=0.7$，$P(\sim B/A)=0.3$，$P(B/\sim A)=0.1$，$P(\sim B/\sim A)=0.9$；

今天有晚霞，判断明天天气如何？ 计算 P（A/B），P（~A/B）

当然，还可能的问题：今天没有晚霞，判断明天天气如何？ 计算 P（A/~B），P（~A/~B）

# 概述

## 例子：天气预报

### Bayesian 决策

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

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

$p(A,B)=p(B|A)p(A)=0.5082$；$p(\sim A,B)=p(B|\sim A)p(\sim A)=0.0274$；$p(B)=p(B|A)p(A)+p(B|\sim A)p(\sim A)=0.5356$

$$P(A|B)=\frac{p(A,B)}{p(B)}=0.9488\text{，}\quad P(\sim A|B)=\frac{p(\sim A,B)}{p(B)}=0.0512$$

结果：$P(A/B)>P(\sim A/B)$ 即 P（晴天/晚霞）> P（非晴天/晚霞）

# 第二章 贝叶斯学习


2.1 概述

**2.2 贝叶斯决策论**

2.3 贝叶斯分类器

2.4 贝叶斯学习与参数估计问题

## 基础知识

### 1．概率基础

事件 A 的概率： $0 \le P(A) \le 1$

条件概率： $P(A|B)=\dfrac{P(AB)}{P(B)}$， $P(B|A)=\dfrac{P(AB)}{P(A)}$

乘法定理： $P(AB)=P(A|B)P(B)=P(B|A)P(A)$

全概率公式： $A_1 \cup A_2 \cup \cdots A_n = \Omega$ 且 $A_i \cap A_j = \varphi$ ，则 $P(A)=\sum_{i=1}^{n} P(A|B_i)P(B_i)$

（重要的）Bayes 公式： $P(A_i \mid B)=\dfrac{P(B \mid A_i)P(A_i)}{\sum_{j=1}^{n} P(B \mid A_j)P(A_j)}$,

$$P(A_i \mid B) \propto P(B \mid A_i)P(A_i)$$

# 贝叶斯决策论

## 基础知识

### 2．Bayes 决策

- 基于观察特征、类别的贝叶斯公式

$$P(\omega_i|x) = \frac{p(x|\omega_i)p(\omega_i)}{p(x)} \qquad \left(posterior = \frac{likelihood * prior}{evidence}\right)$$

$$= \frac{p(x|\omega_i)p(\omega_i)}{\sum_j p(x|\omega_j)p(\omega_j)}$$

$$P(\omega_i|x) \propto p(x|\omega_i)p(\omega_i) \qquad (posterior \propto likelihood * prior)$$

# 贝叶斯决策论

## 基础知识

- 贝叶斯决策

$$\text{Decide}\begin{cases}\omega_1 & if\ p(\omega_1|x)>p(\omega_2|x)\\ \omega_2 & otherwise\end{cases} \Rightarrow \begin{cases}\omega_1 & if\ p(x|\omega_1)p(\omega_1)>p(x|\omega_2)p(\omega_2)\\ \omega_2 & otherwise\end{cases}$$

**or**

$$\begin{cases}\omega_1 & if\ \dfrac{p(x|\omega_1)}{p(x|\omega_2)}>\dfrac{p(\omega_2)}{p(\omega_1)}\\ \omega_2 & otherwise\end{cases}$$

# 贝叶斯决策论

## 基础知识

- 类别相似性函数

$$g_i(x)=p(\omega_i|x)=\frac{p(x|\omega_i)p(\omega_i)}{\sum_{j=1}^{c}p(x|\omega_j)p(\omega_j)}$$

$$g_i(x)=p(x|\omega_i)p(\omega_i)$$

$$g_i(x)=\ln p(x|\omega_i)+\ln p(\omega_i)$$

# 贝叶斯决策论

## 基础知识

- **决策函数** $g(x)=g_1(x)-g_2(x)$ ? 〉0

$$g(x)=p(\omega_1\mid x)-p(\omega_2\mid x)$$ ? 〉0

$$g(x)=\ln\frac{p(x|\omega_1)}{p(x|\omega_2)}+\ln\frac{p(\omega_1)}{p(\omega_2)}$$ ? 〉0

# 参考文献


1. 周志华，机器学习，清华大学出版社，2016.
2. Duda, R.O. et al. Pattern classification. 2nd, 2003.
2. 边肇祺，张学工等编著，模式识别(第二版)，清华大学，1999。