朴素贝叶斯

Naive Bayes
learning note For reading translation

我真的不懂忧郁

Delft University of Technology

TU Delft

朴素贝叶斯

Naive Bayes
learning note For reading translation

by

我真的不懂忧郁

Student Name	Student Number
First Surname	1234567

Instructor:	I. Surname
Teaching Assistant:	I. Surname
Project Duration:	Month, Year - Month, Year
Faculty:	Faculty of Aerospace Engineering, Delft

TU Delft

Preface

A preface...

我真的不懂忧郁

Delft, June 2024

Summary

A summary...

目录

Nomenclature

If a nomenclature is required, a simple template can be found below for convenience. Feel free to use, adapt or completely remove.

Abbreviations

Abbreviation	Definition
ISA	International Standard Atmosphere
...	

Symbols

Symbol	Definition	Unit
V	Velocity	[m/s]
...		
ρ	Density	[kg/m^3]
...		

0.1. 分类问题

朴素贝叶斯决策面对的基本问题是，假设我们有一个描述样本的分类属性的集合

$$X = \{X_1, X_2, \cdots, X_m\} \tag{1}$$

现在有一个样本 $C = \{X_1 = c_1, X_2 = c_2, \cdots, X_m = c_m\}$，我们要根据分类属性把样本这个归类到下面的集合

$$Y = \{Y_1, Y_2, \cdots, Y_n\} \tag{2}$$

例如要考虑一个西瓜是好是坏，一般通过敲击西瓜的声音来判断，声音清脆的瓜一般没熟透，不适合当前吃，声音沉闷的瓜则普遍比较适合当前食用。但是也不尽然，有点声音清脆的瓜很好吃属于好瓜，有的声音沉闷的瓜被虫蛀了属于坏瓜，所以我们只能用概率来描述。我们将这个问题描述为如下条件概率

$$\begin{aligned} &P(\text{好瓜}|\text{声音沉闷}) \\ &P(\text{坏瓜}|\text{声音清脆}) \end{aligned} \tag{3}$$

尽管我们无法能肯定声音沉闷一定是好瓜，声音清脆一定是坏瓜，但是我们能确定的是，好瓜大概率声音沉闷，坏瓜大概率声音清脆，由条件概率表述为

$$\begin{aligned} &P(\text{声音沉闷}|\text{好瓜}) \\ &P(\text{声音清脆}|\text{坏瓜}) \end{aligned} \tag{4}$$

同时我们通常可以知道今年度好瓜或者坏瓜的概率和今年度西瓜敲击声音沉闷或者的概率，这些属于先验概率，那么根据贝叶斯公式我们可以将声音沉闷或者清脆来推断好瓜的概率表示出来

$$\begin{aligned} P(\text{好瓜}|\text{声音沉闷}) &= \frac{P(\text{声音沉闷}|\text{好瓜}) \cdot P(\text{好瓜})}{P(\text{声音沉闷})} \\ P(\text{坏瓜}|\text{声音清脆}) &= \frac{P(\text{声音清脆}|\text{坏瓜}) \cdot P(\text{坏瓜})}{P(\text{声音清脆})} \end{aligned} \tag{5}$$

注意到式子的左边是我们不太敢确定的事儿，但是式子的右边都是大概率可以确定的事儿，$P(\text{声音清脆}|\text{坏瓜})$ 和 $P(\text{声音沉闷}|\text{好瓜})$ 都是我们能确定的事实，$P(\text{声音清脆})$,$P(\text{声音沉闷})$,$P(\text{好瓜})$,$P(\text{坏瓜})$ 这些都是属于先验的概率，也就是说我们可以依照贝叶斯公式，通过已经能确定的事实去推断不能准确判断的事实。

0.2. 条件独立性假设

再次回到贝叶斯决策的基本问题，假如有一堆分类属性 $X = \{X_1, X_2, \cdots, X_m\}$，描述待分类事件的特征如敲击西瓜的声音，我们西方通过这对分类属性去判断事件属于分类集合某一个元素 $Y_i \in Y$ 的概率。不一样的是我们现在发现 X 的元素不唯一，这就意味着我们不能仅仅依靠听敲击西瓜的声音去判断西瓜的好坏，影响判断西瓜品质的还有颜色、气味、生长环境等等。因此，我们

的后验条件概率就应该写成

$$\begin{aligned} &P(Y_1|X_1=c_1,X_2=c_2,\cdots,X_m=c_m)\\ &P(Y_2|X_1=c_1,X_2=c_2,\cdots,X_m=c_m)\\ &\qquad\vdots\\ &P(Y_n|X_1=c_1,X_2=c_2,\cdots,X_m=c_m) \end{aligned} \tag{6}$$

这就意味着对于任意 $Y_i \in Y$

$$P(Y_i|X_1,X_2,\cdots,X_m)=\frac{P(X_1,X_2,\cdots,X_m|Y_i)P(Y_i)}{P(X_1,X_2,\cdots,X_m)} \tag{7}$$

这对我们是很头疼的，因为我们要找出去计算不同属性同时发生的先验概率，不但会给数据统计造成麻烦而且不利于计算机计算。因此在朴素贝叶斯算法里面，我们往往采用**条件独立性假设**。即假设 $X_1,X_2,\cdots,X_m$ 之间是相互独立的，这就意味着

$$P(Y_i|X_1,X_2,\cdots,X_m)=\prod_{j\in\{1,2,\cdots,m\}}\frac{P(X_j|Y_i)P(Y_i)}{P(X_j)}=\frac{P(Y_i)\prod\limits_{j\in\{1,2,\cdots,m\}}P(X_j|Y_i)}{C} \tag{8}$$

其中 $C=P(X_1)P(X_2)\cdots P(X_m)$，一般来说这是先验概率，是一个常数，在贝叶斯估计中一般称为”evidence”。

0.3. 决策

我们的目标是最大化 $P(Y_i|X_1,X_2,\cdots,X_m)$，由于分母是常数，只要最大化分子，即下列优化问题

$$Y_i=\arg\max P(Y_i)\prod_{j=1}^{m}P(X_j|Y_i) \tag{9}$$

即选择分类 Y_i 使的 $P(Y_i|X_1,X_2,\cdots,X_m)$ 最大。

0.4. 估计

我们可以通过极大似然估计的方式来估计这个问题，即对优化问题取对数，然后求对 Y_i 的极值点，但是用极大似然估计很可能会出现要估计的概率值为 0 的情况，比如在学习问题中，训练集中某个量没有出现过，这会影响到后验概率的计算结果，使分类产生偏差。解决这一个问题的方式是通过贝叶斯估计，条件概率的贝叶斯估计如下

$$P_\lambda(X^{(j)}=a_{ji}|Y=c_k)=\frac{\sum\limits_{i=1}^{N}I(x^{(j)_i=a_{ji},y_i=c_k})+\lambda}{\sum\limits_{i=1}^{N}I(y_i=c_k)+S_j\lambda} \tag{10}$$

其中 I 是指示函数，$\lambda>0$，等价于随机变量每个取值的频数上赋予一个正数 $\lambda>0$，当 $\lambda=0$ 的时候就是极大似然估计，当 $\lambda=1$ 时称为**拉普拉斯平滑** *(Laplacian smoothing)*。显然对任

意 $l=1,2,\cdots,S_j$，$k=1,2,\cdots,K$，有

$$\begin{aligned} &P_\lambda(X^{(j)}=aji|Y=c_k)>0 \\ &\sum_{i=1}^{S_j}P(X^{(j)}=a_{ji}|Y=c_k)=1 \end{aligned} \tag{11}$$

表明贝叶斯估计确实是一种概率分布，同一，先验概率的贝叶斯估计是

$$P_\lambda(Y=c_k)=\frac{\sum\limits_{i=1}^{N}I(y_i=c_k)+\lambda}{N+K\lambda} \tag{12}$$

References

[1] I. Surname, I. Surname, and I. Surname. “The Title of the Article”. In: *The Title of the Journal* 1.2 (2000), pp. 123–456.

Chapter A

Source Code Example

Adding source code to your report/thesis is supported with the package `listings`. *An example can be found below. Files can be added using* `\lstinputlisting[language=<language>]{<filename>}`.

```
"""
ISA Calculator: import the function, specify the height and it will return a
list in the following format: [Temperature,Density,Pressure,Speed of Sound].
Note that there is no check to see if the maximum altitude is reached.
"""

import math
g0 = 9.80665
R = 287.0
layer1 = [0, 288.15, 101325.0]
alt = [0,11000,20000,32000,47000,51000,71000,86000]
a = [-.0065,0,.0010,.0028,0,-.0028,-.0020]

def atmosphere(h):
    for i in range(0,len(alt)-1):
        if h >= alt[i]:
            layer0 = layer1[:]
            layer1[0] = min(h,alt[i+1])
            if a[i] != 0:
                layer1[1] = layer0[1] + a[i]*(layer1[0]-layer0[0])
                layer1[2] = layer0[2] * (layer1[1]/layer0[1])**(-g0/(a[i]*R))
            else:
                layer1[2] = layer0[2]*math.exp((-g0/(R*layer1[1]))*(layer1[0]-layer0[0]))
    return [layer1[1],layer1[2]/(R*layer1[1]),layer1[2],math.sqrt(1.4*R*layer1[1])]
```

Chapter B

Task Division Example

If a task division is required, a simple template can be found below for convenience. Feel free to use, adapt or completely remove.

表 B.1: Distribution of the workload

	Task	Student Name(s)
	Summary	
Chapter 1	Introduction	
Chapter 2		
Chapter 3		
Chapter *		
Chapter *	Conclusion	
	Editors	
	CAD and Figures	
	Document Design and Layout	