暨南大學
JINAN UNIVERSITY

机 器 学 习

第五章：朴素贝叶斯

黄斐然

2022/3/21

挂科预测

- 高数考前最后一周，同学们开始通过不同的行为（喝酒、逛街、学习）为考试做准备。

同学编号	喝酒	逛街	学习	挂科
1	1	1	0	1
2	0	0	1	0
3	1	0	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	1	0

例：挂科预测

- 高数考前最后一周，同学们开始通过不同的行为（喝酒、逛街、学习）为考试做准备。

同学编号	喝酒	逛街	学习	挂科
1	1	1	0	1
2	0	0	1	0
3	1	0	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	1	0
7	0	1	0	0
8	0	0	1	1

学霸

学渣

例：挂科预测

- 问题：同学9（没喝酒，没逛街，学习）挂科的概率有多高?

同学编号	喝酒	逛街	学习	挂科
1	1	1	0	1
2	0	0	1	0
3	1	0	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	1	0
7	0	1	0	0
8	0	0	1	1
9	0	0	1	？？？

例：挂科预测

- 问题：同学9（没喝酒，没逛街，学习）挂科的概率有多高?

同学编号	喝酒	逛街	学习	挂科
1	1	1	0	1
2	0	0	1	0
3	1	0	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	1	0
7	0	1	0	0
8	0	0	1	1
9	0	0	1	？？？

50%?

变量表示

- 问题：同学9（没喝酒，没逛街，学习）挂科的概率有多高?
  - A表示挂科
  - B表示喝酒
  - C表示逛街
  - D表示学习

同学编号	喝酒	逛街	学习	挂科

概率与条件概率

- 概率
  - 亦称“或然率”，它是反映随机事件出现的可能性大小。
- 条件概率：
  - 一个事件发生后另一个事件发生的概率。一般的形式为P(x|y)表示y发生的条件下x发生的概率。

概率与条件概率

概率

- P(A=1) = 1/2
- P(A=0) = 1/2

同学编号	喝酒(B)	逛街(C)	学习(D)	挂科(A)
1	1	1	0	1
2	0	0	1	0
3	1	0	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	1	0
7	0	1	0	0
8	0	0	1	1

概率与条件概率

喝酒对挂科的影响

- $P(A=1|B=1) = 2/3$
- $P(A=1|B=0) = 2/5$
- 也可以使用条件概率公式计算

  $P(A=1|B=1) = P(A=1 \wedge B=1)/P(B=1)$

同学编号	喝酒(B)	逛街(C)	学习(D)	挂科(A)
1	1	1	0	1
2	0	0	1	0
3	1	0	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	1	0
7	0	1	0	0
8	0	0	1	1

概率与条件概率

逛街对挂科的影响

- $P(A=1|C=1) = 2/4$
- $P(A=1|C=0) = 2/4$

同学编号	喝酒(B)	逛街(C)	学习(D)	挂科(A)
1	1	1	0	1
2	0	0	1	0
3	1	0	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	1	0
7	0	1	0	0
8	0	0	1	1

概率与条件概率

学习对挂科的影响

- $P(A=1|D=1) = 1/4$
- $P(A=1|D=0) = 3/4$

同学编号	喝酒(B)	逛街(C)	学习(D)	挂科(A)
1	1	1	0	1
2	0	0	1	0
3	1	0	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	1	0
7	0	1	0	0
8	0	0	1	1

概率与条件概率

没喝酒，没逛街，学习的条件下挂科的概率

- P(A=1|B=0,C=0,D=1) = ?

- 1、直接计算
  - P(A=1|B=0,C=0,D=1) =1/2 ❌
  - P(A=1|B=1,C=1,D=1)=?

同学编号	喝酒(B)	逛街(C)	学习(D)	挂科(A)
1	1	1	0	1
2	0	0	1	0
3	1	0	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	1	0
7	0	1	0	0
8	0	0	1	1

概率与条件概率

- 没喝酒，没逛街，学习的条件下挂科的概率
  - $P(A=1|B=0,C=0,D=1) = ?$
  - 2、直接乘式
    - $P(A=1|B=0,C=0,D=1) =$
    - $P(A=1|B=0)*P(A=1|C=0)*P(A=1|D=1)$

同学编号	喝酒(B)	逛街(C)	学习(D)	挂科(A)
1	1	1	0	1
2	0	0	1	0
3	1	0	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	1	0
7	0	1	0	0
8	0	0	1	1

条件概率公式

条件概率公式：

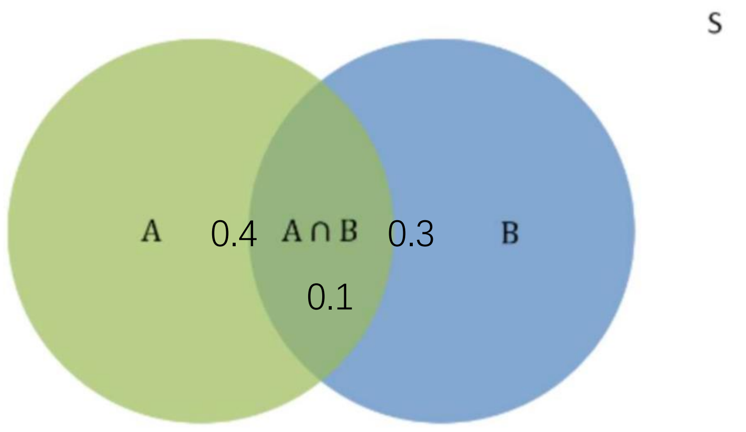
假设A和B为两个不相互独立的事件

- 由上图可以看出，在事件B已经发生的情况下，事件A发生的概率为事件A和事件B的交集除以事件B：

$$P(A|B)=P(A\cap B)/P(B)$$

- 同理，在事件A已经发生的情况下，事件B发生的概率为事件A和事件B的交集除以事件A：

$$P(B|A)=P(A\cap B)/P(A)$$



贝叶斯公式的推导

贝叶斯定理的推导:

- 由公式$P(A|B)=P(A \cap B)/P(B)$推出

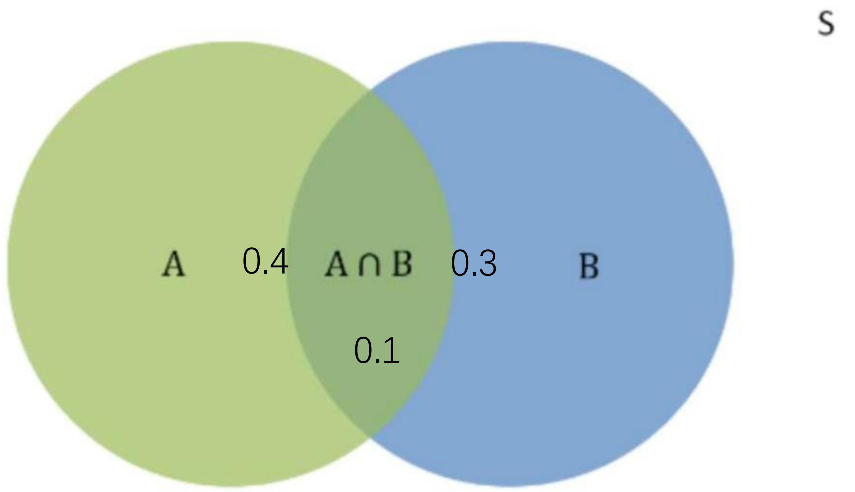
$$P(A \cap B)=P(A|B)P(B)$$

- 由公式$P(A|B)=P(A \cap B)/P(B)$推出:

$$P(A \cap B)=P(B|A)P(A)$$

由两式得到**贝叶斯公式**:

$$P(A|B)=P(B|A)P(A)/P(B)$$



一般化的贝叶斯公式

- 更一般化的情况，假设事件 A 本身又包含多种可能性，即 A 是一个集合：$A_{(1,2,3,\ldots,n)}$，那么对于集合中任意的 A_i，
- 由公式P(A|B)=P(B|A)P(A)/P(B)，得出一般化的贝叶斯公式：

$$P(A_i|B)=\frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$

多因素的贝叶斯公式

- 但在实际应用中，很少有一件事只受一个特征影响的情况，往往**影响一件事的因素有多个**。假设，影响 B 的因素有 n 个，分别是 $b^{(1)},b^{(2)},\dots,b^{(n)}$。
- 如果考虑多个影响B的因素，则 P(A|B) 可以写为：
- $P(A_i|B) = P\left(A_i \middle| b^{(1)}, b^{(2)}, \dots, b^{(n)}\right) = \frac{P(b^{(1)},b^{(2)},\dots,b^{(n)}|A_i)P(A_i)}{P(B)}$

朴素贝叶斯公式

- 影响 B 的因素有 n 个，假定特征之间独立（朴素），则 P(A|B) 可以写为：

$$P(A_i|B) = P\left(A_i \middle| b^{(1)}, b^{(2)}, \ldots, b^{(n)}\right) = \frac{P(A_i)\prod_{j=1}^{n} P(b^{(j)}|A_i)}{P(B)}$$

朴素贝叶斯分类公式

- 上式中的 $b^{(1)}$ 到 $b^{(n)}$是特征（Feature），而 A 则是最终的类别（Class），所以，我们换一个写法：

- $P\left(y_i \middle| x^{(1)}, x^{(2)}, \ldots, x^{(n)}\right) = \frac{P(y_i)\prod_{j=1}^{n} P(x^{(j)}|y_i)}{P(X)}$

朴素贝叶斯分类公式

- $P\left(\mathrm{y}_i \middle| x^{(1)}, x^{(2)}, \ldots, x^{(n)}\right) = \frac{P(y_i)\prod_{j=1}^{n} P(x^{(j)}|y_i)}{P(X)}$
- 求出$P\left(\mathrm{y}_1 \middle| x^{(1)}, x^{(2)}, \ldots, x^{(n)}\right)$, $P\left(\mathrm{y}_2 \middle| x^{(1)}, x^{(2)}, \ldots, x^{(n)}\right)$, ……的分子
- 朴素贝叶斯分类概率即为：$\frac{\breve{P}\left(\mathrm{y}_i \middle| x^{(1)}, x^{(2)}, \ldots, x^{(n)}\right)}{\sum_i \breve{P}\left(\mathrm{y}_i \middle| x^{(1)}, x^{(2)}, \ldots, x^{(n)}\right)}$

回顾问题？

- 不喝酒、不逛街、好好学习的同学会不会挂科?

$$P(y=1|x^{(1)}=0,x^{(2)}=0,x^{(3)}=1)$$

编号	喝酒$x^{(1)}$	逛街$x^{(2)}$	学习$x^{(3)}$	挂科y
1	1	1	0	1
2	0	0	1	0
3	1	0	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	1	0
7	0	1	0	0
8	0	0	1	1

回顾问题？

- 不喝酒、不逛街、好好学习的同学会不会挂科?

$$P(y=1|x^{(1)}=0,x^{(2)}=0,x^{(3)}=1)$$

编号	喝酒$x^{(1)}$	逛街$x^{(2)}$	学习$x^{(3)}$	挂科y
1	1	1	0	1
2	0	0	1	0
3	1	0	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	1	0
7	0	1	0	0
8	0	0	1	1

朴素贝叶斯公式

$$P\left(y_i|x^{(1)},x^{(2)},\dots,x^{(n)}\right)=\frac{P(y_i)\prod_{j=1}^{n}P(x^{(j)}|y_i)}{P(X)}$$

回顾问题？

- 不喝酒、不逛街、好好学习的同学会不会挂科?

$$P(y=1|x^{(1)}=0,x^{(2)}=0,x^{(3)}=1)$$

编号	喝酒$x^{(1)}$	逛街$x^{(2)}$	学习$x^{(3)}$	挂科y
1	1	1	0	1
2	0	0	1	0
3	1	0	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	1	0
7	0	1	0	0
8	0	0	1	1

朴素贝叶斯公式

$$\breve{P}(y_i|x^{(1)},x^{(2)},x^{(3)})$$

$$=P(y_i)P(x^{(1)}|y_i)\,P(x^{(2)}|y_i)\,P(x^{(3)}|y_i)$$

贝叶斯公式求解

- 不喝酒、不逛街、好好学习的同学会不会挂科？

$$P(y=1|x^{(1)}=0,x^{(2)}=0,x^{(3)}=1)$$

编号	喝酒$x^{(1)}$	逛街$x^{(2)}$	学习$x^{(3)}$	挂科y
1	1	1	0	1
2	0	0	1	0
3	1	0	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	1	0
7	0	1	0	0
8	0	0	1	1

方法：使用贝叶斯公式

$$\breve{P}(y=1|x^{(1)}=0,x^{(2)}=0,x^{(3)}=1)$$

$$=P(y=1)*P(x^{(1)}=0|y=1)*P(x^{(2)}=0|y=1)*P(x^{(3)}=1|y=1)$$

$$\breve{P}(y=0|x^{(1)}=0,x^{(2)}=0,x^{(3)}=1)$$

$$=P(y=0)*P(x^{(1)}=0|y=0)*P(x^{(2)}=0|y=0)*P(x^{(3)}=1|y=0)$$

代入计算

式一

$$P(y=1)*P(x^{(1)}=0|y=1)*P(x^{(2)}=0|y=1)*P(x^{(3)}=1|y=1)$$

编号	喝酒$x^{(1)}$	逛街$x^{(2)}$	学习$x^{(3)}$	挂科y
1	1	1	0	1
2	0	0	1	0
3	1	0	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	1	0
7	0	1	0	0
8	0	0	1	1

- $P(y=1)=\frac{1}{2}$
- $P\left(x^{(1)}=0\middle|y=1\right)=\frac{1}{2}$
- $P\left(x^{(2)}=0\middle|y=1\right)=\frac{1}{2}$
- $P\left(x^{(3)}=1\middle|y=1\right)=\frac{1}{4}$
- 结果：2/64

代入计算

式二

$$P(y=0)*P(x^{(1)}=0|y=0)*P(x^{(2)}=0|y=0)*P(x^{(3)}=1|y=0)$$

编号	喝酒$x^{(1)}$	逛街$x^{(2)}$	学习$x^{(3)}$	挂科y
1	1	1	0	1
2	0	0	1	0
3	1	0	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	1	0
7	0	1	0	0
8	0	0	1	1

- $P(y=0)=\frac{1}{2}$
- $P\left(x^{(1)}=0\middle|y=0\right)=\frac{3}{4}$
- $P\left(x^{(2)}=0\middle|y=0\right)=\frac{1}{2}$
- $P\left(x^{(3)}=1\middle|y=0\right)=\frac{3}{4}$
- 结果：9/64

- 综合，挂科概率 = 2/11 = 0.1818

什么是朴素贝叶斯

- 分类算法常用的有很多种，朴素贝叶斯算法是其中一个比较常用的，之所以称为朴素贝叶斯算法主要是因为该算法最基本的原理是基于贝叶斯定理的，称为朴素是因为该算法成立的前提是特征之间必须得是独立的。
- 朴素贝叶斯（Naive Bayes）算法理论基础是基于贝叶斯定理和条件独立性假设的一种分类方法。

先验概率

- 事件发生前的预判概率。
- 可以是基于历史数据的统计，可以由背景常识得出，也可以是人的主观观点给出。
- 一般都是单独事件概率，如P(x),P(y)。

$$P\left(y_i \middle| x^{(1)}, x^{(2)}, \dots, x^{(n)}\right) = \frac{\boxed{P(y_i)} \prod_{j=1}^{n} P(x^{(j)} | y_i)}{P(X)}$$

后验概率

后验概率：

- 事件发生后求的反向条件概率；或者说，基于先验概率求得的反向条件概率。概率形式与条件概率相同。

$$P\left(y_i \middle| x^{(1)}, x^{(2)}, \dots, x^{(n)}\right) = \frac{P(y_i)\prod_{j=1}^{n} \boxed{P(x^{(j)}|y_i)}}{P(X)}$$

如何处理概率为0的情况？

拉普拉斯平滑

- 为了解决零概率的问题，法国数学家拉普拉斯最早提出用加1的方法估计没有出现过的现象的概率，所以加法平滑也叫做拉普拉斯平滑。
- 假定训练样本很大时，每个分量x的计数加1造成的估计概率变化可以忽略不计，但可以方便有效的避免零概率问题。

$$P(A_i = a|C = c) = \frac{N_{ac} + 1}{N_c + N_i}$$

如何处理特征连续的情况？

- 怎么求$P\left(x^{(1)}=0.23 \middle| y=0\right)$?
  - 高斯朴素贝叶斯模型。在高斯朴素贝叶斯模型中，特征向量X的特征 通常为 连续型变量，并且假定所有特征的取值是符合高斯分布的多项分布朴素贝叶斯模型
  - 多项分布朴素贝叶斯模型。特征向量X的特征通常为离散型变量，并且假定所有特征的取值是符合多项分布的，可用于文本分类。
  - 伯努利朴素贝叶斯模型。在伯努利朴素贝叶斯模型中，每个特征的取值是布尔型，或以0和1表示，所以伯努利模型中，每个特征值为0或者1。

使用sklearn实现

使用sklearn实现：

- from sklearn.naive_bayes import CategoricalNB 或 MultinomialNB 或 BernoulliNB 或 GaussianNB
- MultinomialNB(alpha=1.0, fit_prior=True, class_prior=None)
- 重要参数：
  - alpha ：alpha即为拉普拉斯平滑常数，如果你没有特别的需要，用默认的1即可。如果发现拟合的不好，需要调优时，可以选择稍大于1或者稍小于1的数。
  - fit_prior：是否要考虑先验概率，如果True，则可以自己用第三个参数class_prior输入先验概率。

使用sklearn实现

使用sklearn实现：

In [8]:

```python
from sklearn.naive_bayes import MultinomialNB
from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import KFold
from sklearn import metrics
import numpy as np

dataset = load_breast_cancer()
X = dataset['data']
y = dataset['target']

avg_scores = []
for i in range(5):

    kf = KFold(n_splits=10,shuffle=True)
    score = []
    for train_inx, test_inx in kf.split(X):
        clf = MultinomialNB(alpha=1.0, fit_prior=True, class_prior=None
                            ).fit(X[train_inx],y[train_inx])
        y_pre = clf.predict(X[test_inx])
        y_test = y[test_inx]
        score.append(metrics.accuracy_score(y_test,y_pre))
    avg_scores.append(np.mean(score))

print(np.mean(avg_scores))
```

```
0.8941729323308272
```

优缺点

- 朴素贝叶斯优点：
  - 所需估计的参数少，对于缺失数据不敏感。
  - 有着坚实的数学基础，以及稳定的分类效率。
- 朴素贝叶斯缺点：
  - 假设属性之间相互独立，这往往并不成立。（喜欢吃番茄、鸡蛋，却不喜欢吃番茄炒蛋）。
  - 需要知道先验概率。

第二次作业

编号	收入	学历	年龄	贷款
1	高	研究生	40-60	是
2	高	本科	>60	否
3	高	专科	20-40	是
4	中	研究生	40-60	是
5	中	本科	40-60	是
6	中	专科	20-40	是
7	中	研究生	>60	是
8	低	本科	>60	否
9	低	本科	20-40	是
10	低	专科	40-60	否
11	低	专科	>60	否

计算收入中等，20-40本科生是否应该贷款
（计算过程+代码）

问题？

暨南大學

JINAN UNIVERSITY