# 1.学习目标

|  |
| --- |
|  |

# 2.什么是朴素贝叶斯分类方法

|  |
| --- |
|  |
|  |
|  | |

## 所谓的朴素,是指特征和特征之间是相互独立的.

# 3.概率基础

|  |
| --- |
|  |
|  |
|  |
|  |

# 4.联合概率,条件概率与相互独立

## 上面的例子在第二条私有联合概率,第三条属于条件概率,第四条既符合联合概率,又符合条件概率

|  |
| --- |
|  |

# 5.贝叶斯公式

|  |
| --- |
|  |

## 5.1计算案例,计算小明是产品经理而且体重超重被女神喜欢的概率,导入上面的公式

|  |
| --- |
|  |
|  |

## 5.2朴素贝叶斯算法的应用场景:文本分类,垃圾邮件分类,因为我们如果需要用机器学习来处理文本我们必须把文本的单词作为特征,我们是假设单词之间是相互独立的.符合朴素贝叶斯算法条件.

|  |
| --- |
|  |

## 5.3文章分类计算

|  |
| --- |
|  |
|  |
| 不合适,所以需要拉普拉斯平滑系数 |

## 5.4. 拉普拉斯平滑系数

|  |
| --- |
|  |

# 6.算法API模块

|  |
| --- |
|  |

# 7实战案例:20类新闻分类

|  |
| --- |
|  |

## 流程分析

|  |
| --- |
|  |

## 实例代码

|  |
| --- |
|  |
| *""" 用朴素贝叶斯算法来做20类新闻分类,数据使用sklearn下载的数据 """* from sklearn.naive\_bayes import MultinomialNB from sklearn.datasets import fetch\_20newsgroups from sklearn.feature\_extraction.text import TfidfVectorizer from sklearn.model\_selection import train\_test\_split   # 1.获取数据 news = fetch\_20newsgroups(data\_home=".",subset="all") # 2.划分数据集 x\_train,x\_test,y\_train,y\_test = train\_test\_split(news.data,news.target,random\_state=6) # 3.特性工程-文本特征抽取,使用TfidfVectorizer比较好 transformer = TfidfVectorizer() x\_train = transformer.fit\_transform(x\_train) x\_test = transformer.transform(x\_test) # 4.贝叶斯预估器流程,如果有兴趣,你也可以添加模型调优流程 estimator = MultinomialNB() estimator.fit(x\_train,y\_train) # 5.模型评估 # 方法一：直接比对真实值和预估值 y\_predict = estimator.predict(x\_test) print("直接比对真实值和预测值：",y\_predict == y\_test) # # 方法二：计算准确率 score=estimator.score(x\_test,y\_test) print("准确率：",score) |

### 效果

|  |
| --- |
|  |

## 如果你想效果好一点可以修改一下random\_state的值

|  |
| --- |
|  |

### 效果

|  |
| --- |
|  |

## 此外也可以使用模型调优等等…