姓名和学号：李胜志 2210180232

章节名称：ch05\_P123

知识目标：掌握根据具体情景选择合适的朴素贝叶斯算法；掌握机器学习常用的库的使用方法；掌握朴素贝叶斯算法训练模型的方法。

能力目标：能够使用朴素贝叶斯算法训练分类模型。

素质目标：养成分析问题、事前规划的良好习惯。

知识重点：模型的选择与优化。

知识难点：根据具体情境选择合适的朴素贝叶斯算法。

==========================================================

实训目的：

1. 掌握根据具体情景选择合适的朴素贝叶斯算法。
2. 掌握机器学习常用的库的使用方法。
3. 掌握朴素贝叶斯算法训练模型的方法。

实训内容：

为预测未来一天是否下雨，某机构收集了过去7天的天气情况，如表5-3所示。除“序号”列外，其他列中1表示是，0表示否。使用朴素贝叶斯算法训练模型，预测未来一天（刮风，不闷热，多云）会不会下雨。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 是否刮风 | 是否闷热 | 是否多云 | 是否下雨 |
| 1 | 0 | 1 | 0 | 0 |
| 2 | 1 | 1 | 1 | 1 |
| 3 | 0 | 1 | 1 | 1 |
| 4 | 0 | 0 | 0 | 0 |
| 5 | 0 | 1 | 1 | 1 |
| 6 | 0 | 1 | 0 | 0 |
| 7 | 1 | 0 | 0 | 0 |

1. 启动Jupyter Notebook，以Python3工作方式新建 Jupyter Notebook 文档，并重命名为“item5-sx.ipynb”。
2. 数据准备。
3. 导入pandas库。
4. 输入训练集的数据与标签。
5. 将字典转成DataFrame类型并划分特征和标签

import pandas as pd

# 输入训练集数据，用字典存储数据

mydata = {"wind":[0,1,0,0,0,0,1],"hot":[1,1,1,0,1,1,0],"cloud":[0,1,1,0,1,0,0],"rain":[0,1,1,0,1,0,0]}

mydata\_frame = pd.DataFrame(mydata)

x = mydata\_frame.drop("rain", axis=1).values.reshape(-1, 3)

y = mydata\_frame["rain"]

1. 训练模型。
2. Sklearn的naive\_bayes模块提供了3种朴素贝叶斯算法，分别是高斯朴素贝叶斯算法、多项式朴素贝叶斯算法和伯努利朴素贝叶斯算法，选择合适的算法模块。
3. 使用选择好的朴素贝叶斯算法定义一个分类模型。
4. 使用数据集对模型进行训练。

# 训练集中特征变量属于布尔值，因此选择伯努利朴素贝叶斯算法

from sklearn.naive\_bayes import BernoulliNB

# 创建模型并训练

bnl = BernoulliNB()

bnl.fit(x, y)

1. 模型预测。
2. 使用训练好的模型预测未来一天(刮风，不闷热，多云)会不会下雨。
3. 输出预测结果。如果这一天不会下雨，则输出“这是一个好天气”，否则输出“要下雨了，快做好准备吧”。

y\_predict = bnl.predict([[1,0,1]])

print(y\_predict)

if y\_predict == 0:

print("这是一个好天气")

else:

print("要下雨了,快做好准备吧")

运行结果：

[1]

要下雨了,快做好准备吧

因此可以得出这一天会下雨