

朴素贝叶斯算法实现

实验目标

通过本案例的学习和课后作业的练习：

1. 了解朴素贝叶斯算法的基本思想；
2. 能够使用SKlearn实现朴素贝叶斯算法。

你也可以将本案例相关的 ipynb 学习笔记分享到 [AI Gallery Notebook \(https://marketplace.huaweicloud.com/markets/aihub/notebook/list/\)](https://marketplace.huaweicloud.com/markets/aihub/notebook/list/) 版块获得成长值 (https://marketplace.huaweicloud.com/markets/aihub/article/detail/?content_id=9b8d7e7a-a150-449e-ac17-2dcf76d8b492)，分享方法请查看[此文档 \(https://marketplace.huaweicloud.com/markets/aihub/article/detail/?content_id=8afec58a-b797-4bf9-acc4-76ed512a3acb\)](https://marketplace.huaweicloud.com/markets/aihub/article/detail/?content_id=8afec58a-b797-4bf9-acc4-76ed512a3acb)。

案例内容介绍

朴素贝叶斯分类器是一系列以假设特征之间强（朴素）独立下运用贝叶斯定理为基础的简单概率分类器。该分类器模型会给问题实例分配用特征值表示的类标签，类标签取自有限集合。它不是训练这种分类器的单一算法，而是一系列基于相同原理的算法：所有朴素贝叶斯分类器都假定样本每个特征与其他特征都不相关。

本案例推荐的理论学习视频：

- [《AI技术领域课程--机器学习》朴素贝叶斯 \(https://education.huaweicloud.com/courses/course-v1:HuaweiX+CBUCNxE086+Self-paced/courseware/099a890c59f64d44b81440581d22fa37/d132e0a0c98e4e64acf66fc92c5e7892/\)](https://education.huaweicloud.com/courses/course-v1:HuaweiX+CBUCNxE086+Self-paced/courseware/099a890c59f64d44b81440581d22fa37/d132e0a0c98e4e64acf66fc92c5e7892/)

注意事项

1. 如果您是第一次使用 JupyterLab，请查看 [《ModelArts JupyterLab使用指导》 \(https://marketplace.huaweicloud.com/markets/aihub/article/detail/?content_id=03676d0a-0630-4a3f-b62c-07fba43d2857\)](https://marketplace.huaweicloud.com/markets/aihub/article/detail/?content_id=03676d0a-0630-4a3f-b62c-07fba43d2857) 了解使用方法；
2. 如果您在使用 JupyterLab 过程中碰到报错，请参考 [《ModelArts JupyterLab常见问题解决办法》 \(https://marketplace.huaweicloud.com/markets/aihub/article/detail/?content_id=9ad8ce7d-06f7-4394-80ef-4dbf6cfb4be1\)](https://marketplace.huaweicloud.com/markets/aihub/article/detail/?content_id=9ad8ce7d-06f7-4394-80ef-4dbf6cfb4be1) 尝试解决问题。

实验步骤

本案例随机生成数据集，通过sklearn调用朴素贝叶斯的方法。

1、导入依赖库

```
In [1]: import numpy as np
        from sklearn.naive_bayes import BernoulliNB
```

2、创建数据集

```
In [2]: X = np.random.randint(2, size=(6, 100))
        Y = np.array([1, 2, 3, 4, 4, 5])
```

3、定义模型

```
In [3]: clf = BernoulliNB()
```

4、训练模型

```
In [4]: clf.fit(X, Y)
```

```
Out[4]: BernoulliNB(alpha=1.0, binarize=0.0, class_prior=None, fit_prior=True)
```

5、输出预测结果

```
In [5]: print(clf.predict(X[2:3]))
```

```
[3]
```

以上是朴素贝叶斯的实现方法，受限于篇幅原因，本案例未完全覆盖朴素贝叶斯的全部操作，欢迎你将更全面的朴素贝叶斯学习笔记分享到 [AI Gallery Notebook \(https://marketplace.huaweicloud.com/markets/aihub/notebook/list/\)](https://marketplace.huaweicloud.com/markets/aihub/notebook/list/) 版块获得 [成长值 \(https://marketplace.huaweicloud.com/markets/aihub/article/detail/?content_id=9b8d7e7a-a150-449e-ac17-2dcf76d8b492\)](https://marketplace.huaweicloud.com/markets/aihub/article/detail/?content_id=9b8d7e7a-a150-449e-ac17-2dcf76d8b492)，分享方法请查看 [此文档 \(https://marketplace.huaweicloud.com/markets/aihub/article/detail/?content_id=8afec58a-b797-4bf9-acca-76ed512a3acb\)](https://marketplace.huaweicloud.com/markets/aihub/article/detail/?content_id=8afec58a-b797-4bf9-acca-76ed512a3acb)。

作业

请你利用本实验中学到的知识点，完成以下编程题：

1. 请你尝试修改 `BernoulliNB()` 函数的 `alpha`（平滑因子）参数的不同取值，看看该参数的修改对模型会有怎样的影响。<https://marketplace.huaweicloud.com/markets/aihub/notebook/detail/?id=5d08aa2e-e3aa-4149-b796-77c43df8e349>
2. 请你尝试修改 `BernoulliNB()` 函数的 `binarize`（样本特征二值化的阈值）参数的不同取值，看看该参数的修改对模型会有怎样的影响。<https://marketplace.huaweicloud.com/markets/aihub/notebook/detail/?id=579d7aae-b0ad-4783-99e6-470a39aa2c27>
3. 使用鸢尾花数据集，基于Sklearn框架实现一个朴素贝叶斯鸢尾花分类模型。<https://marketplace.huaweicloud.com/markets/aihub/notebook/detail/?id=6c1785bb-0cd0-44b2-aa59-13237026a142>