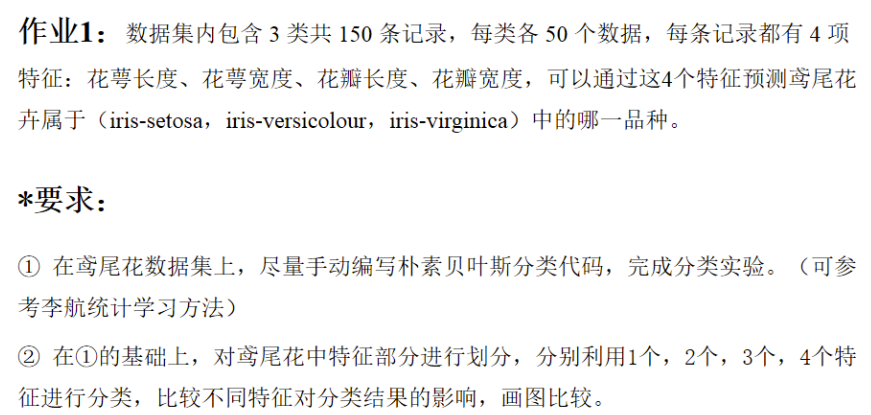
 

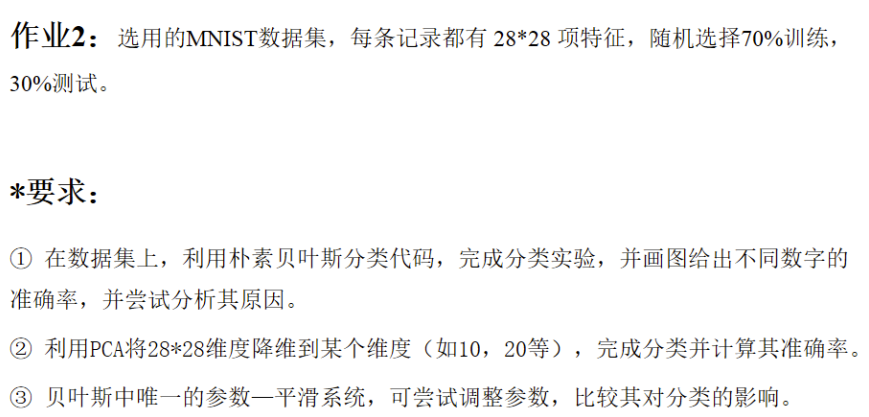
**智能信息系统综合实践**

**实验报告**

|  |  |
| --- | --- |
| **题 目：** | 贝叶斯分类器 |
| **年 级：** | **2020级** |
| **专 业：** | **软件工程** |
| **姓 名：** | **庞晓宇** |

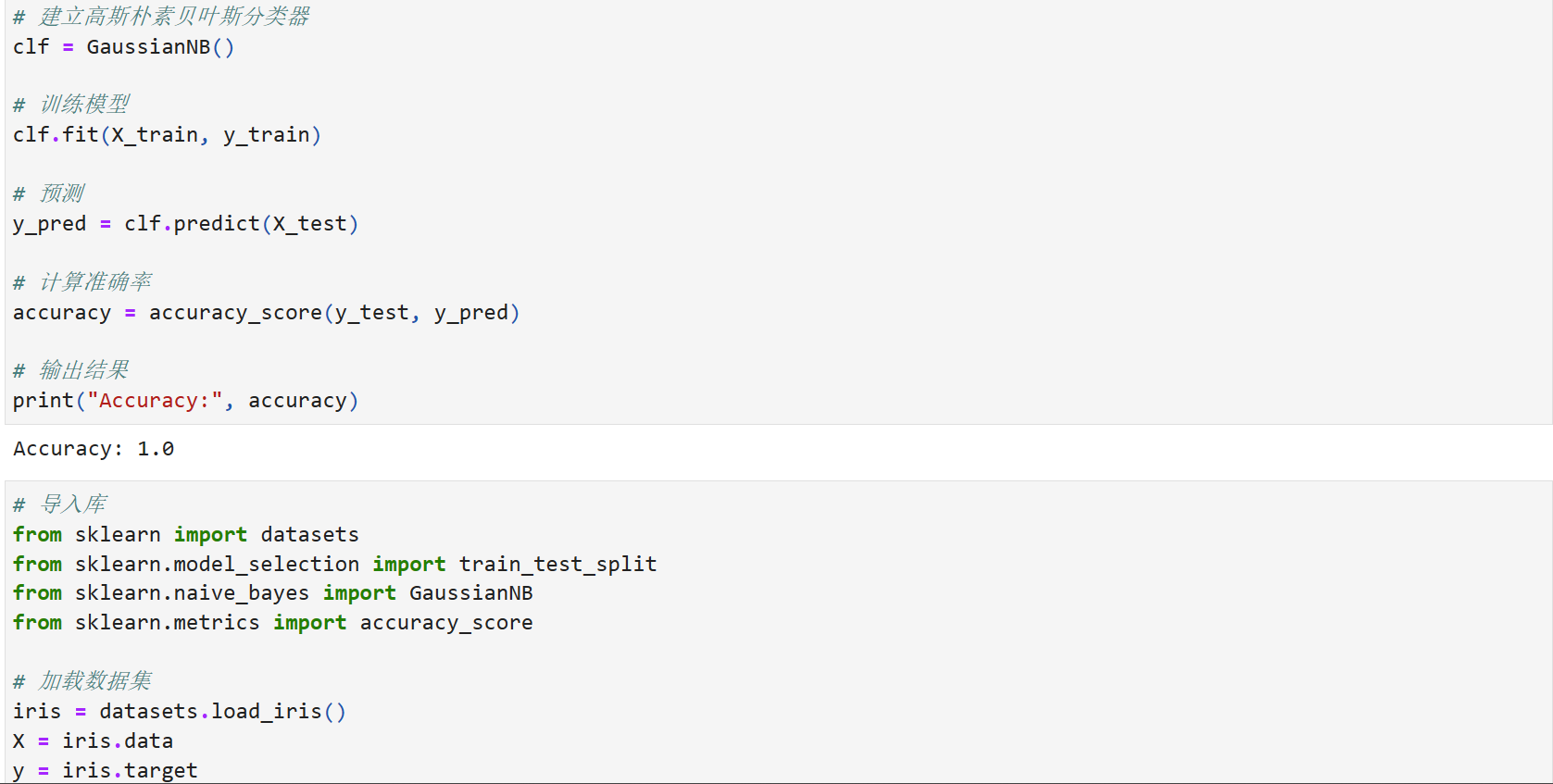
1. **题目（原题目）**

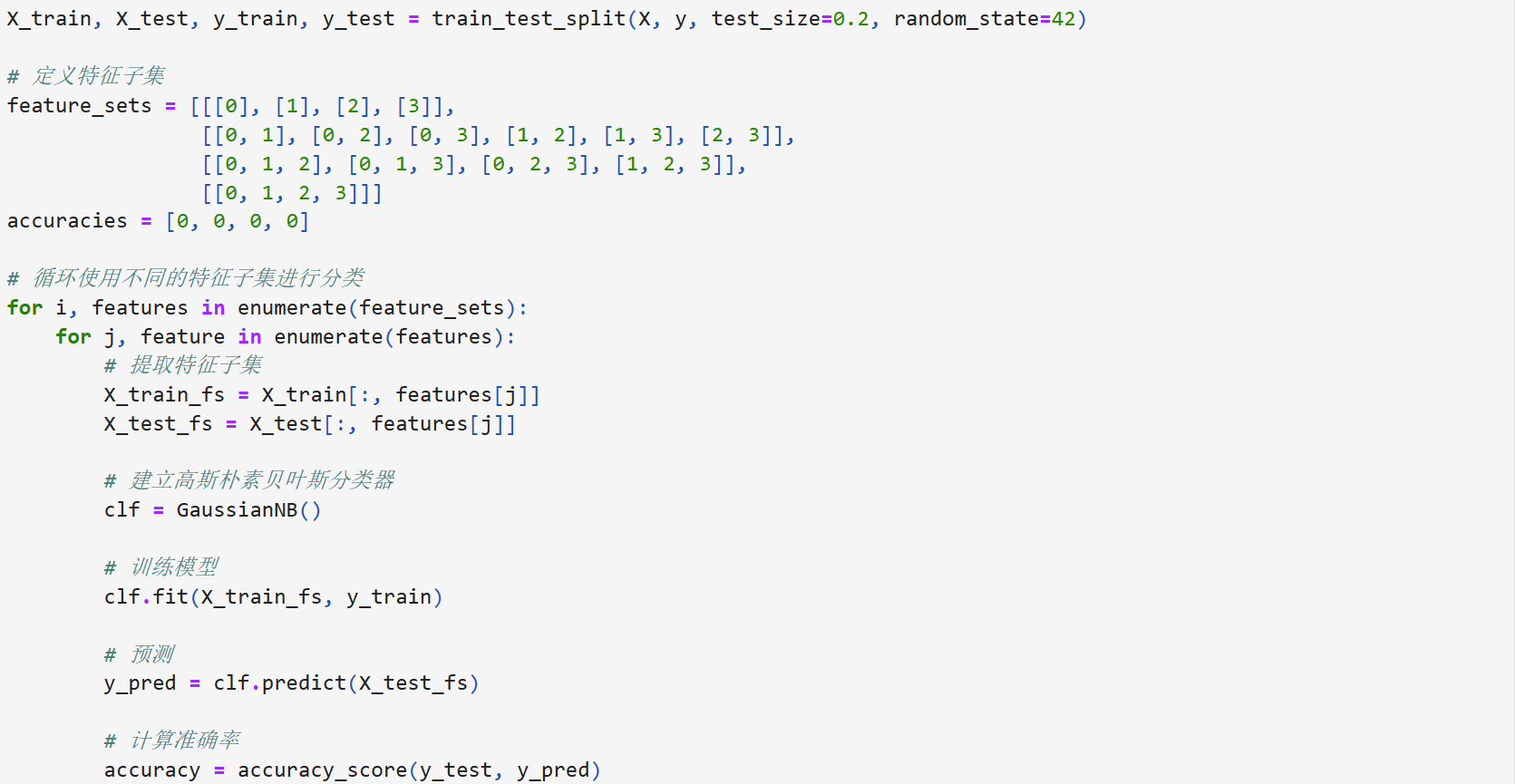


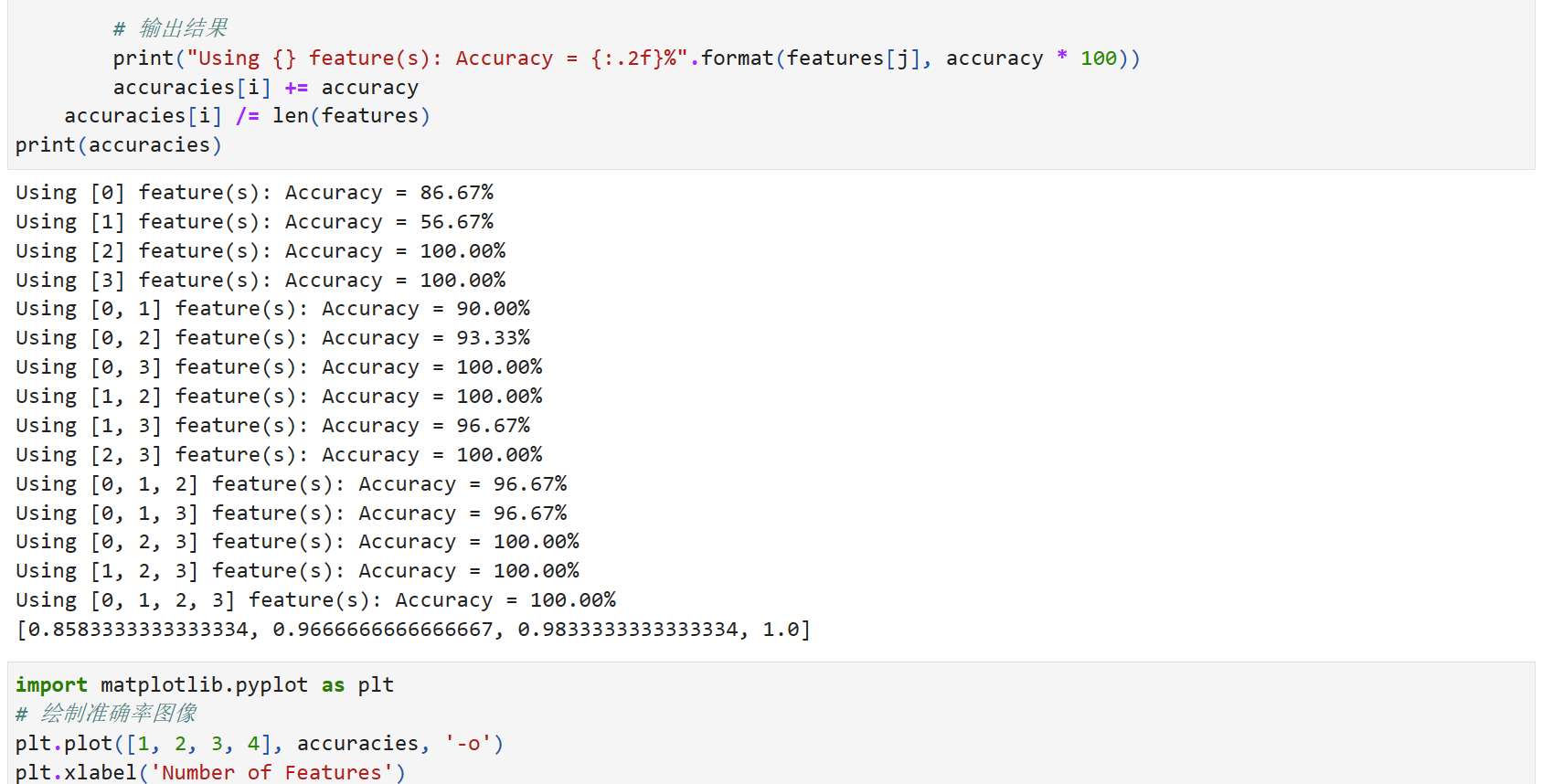


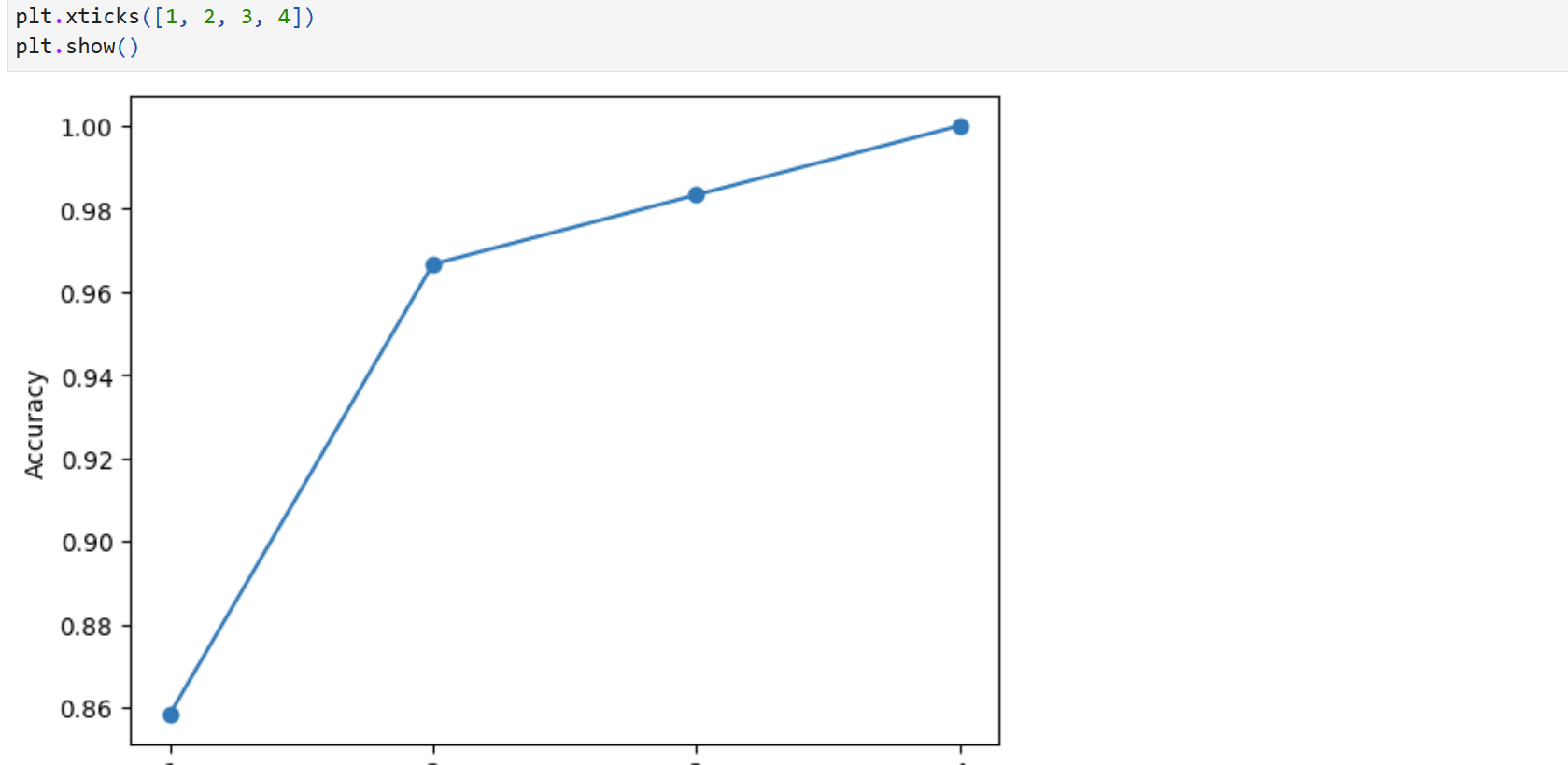
1. **解题步骤（思路+代码）**



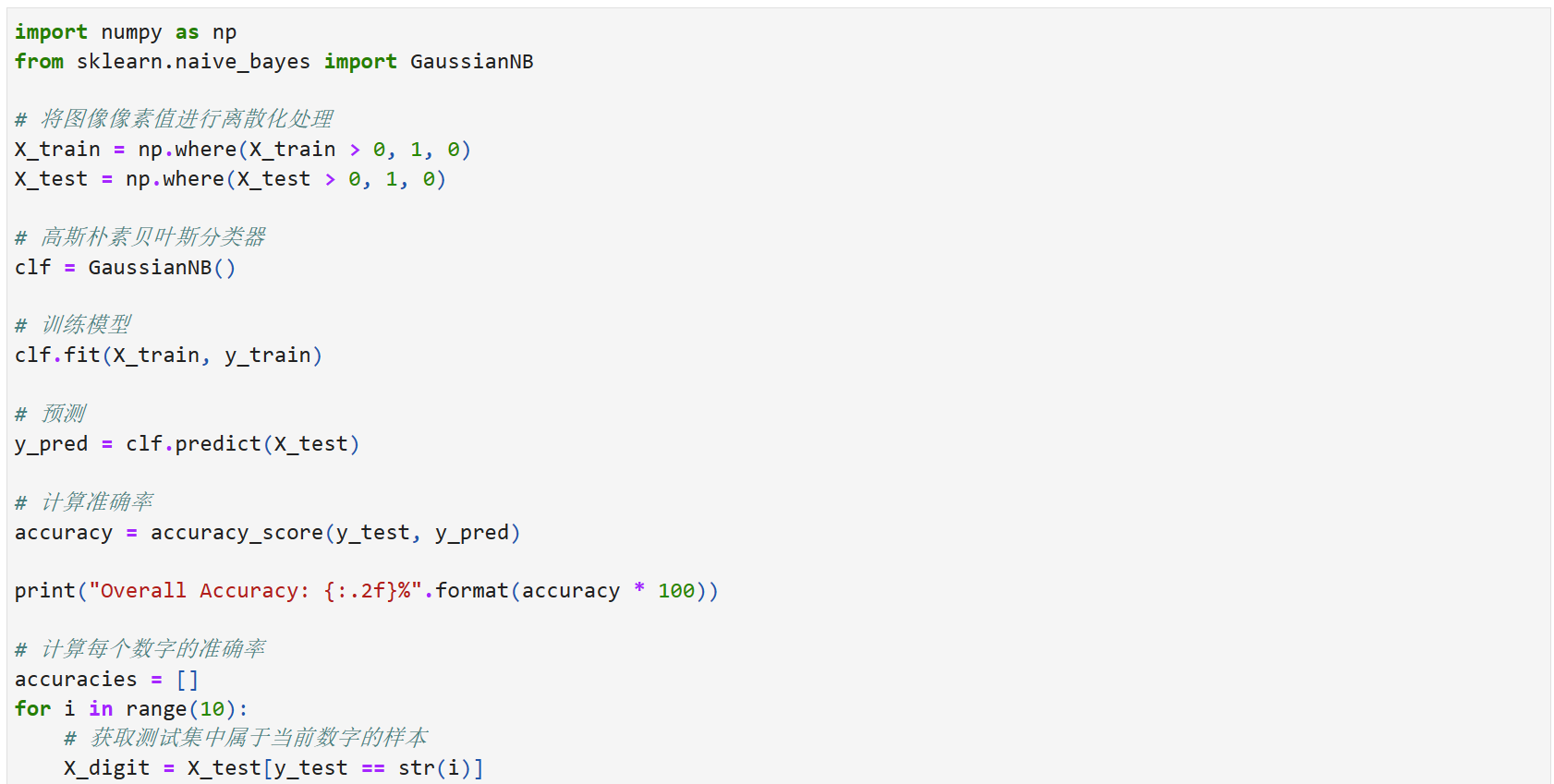


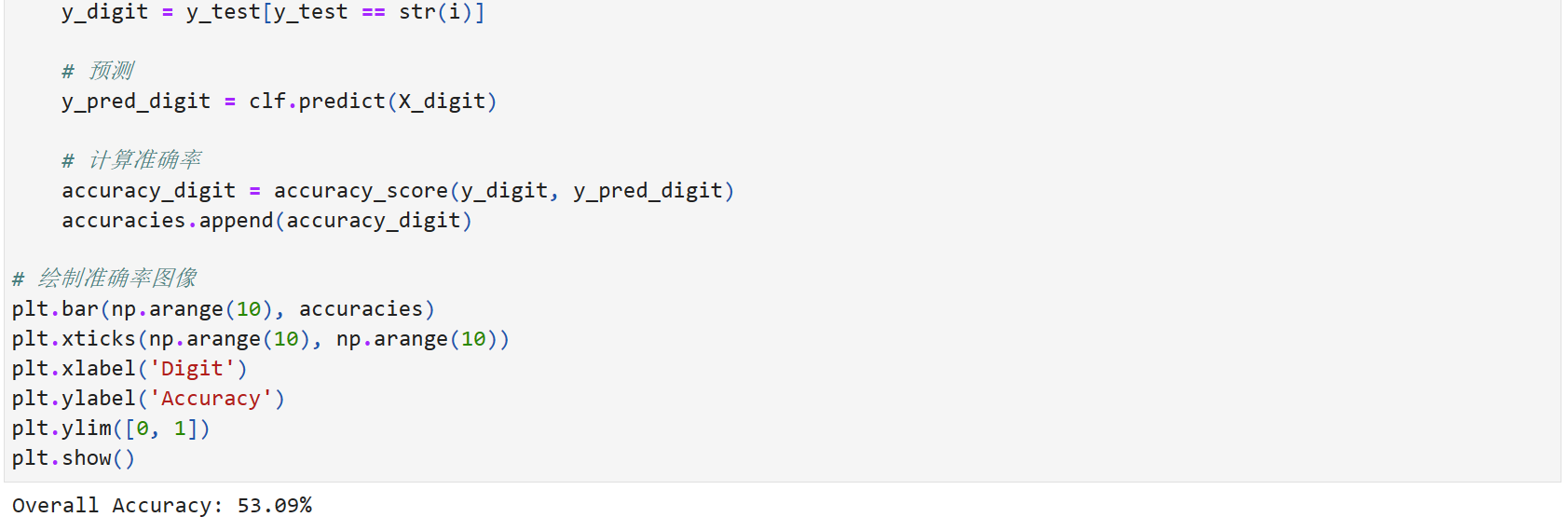


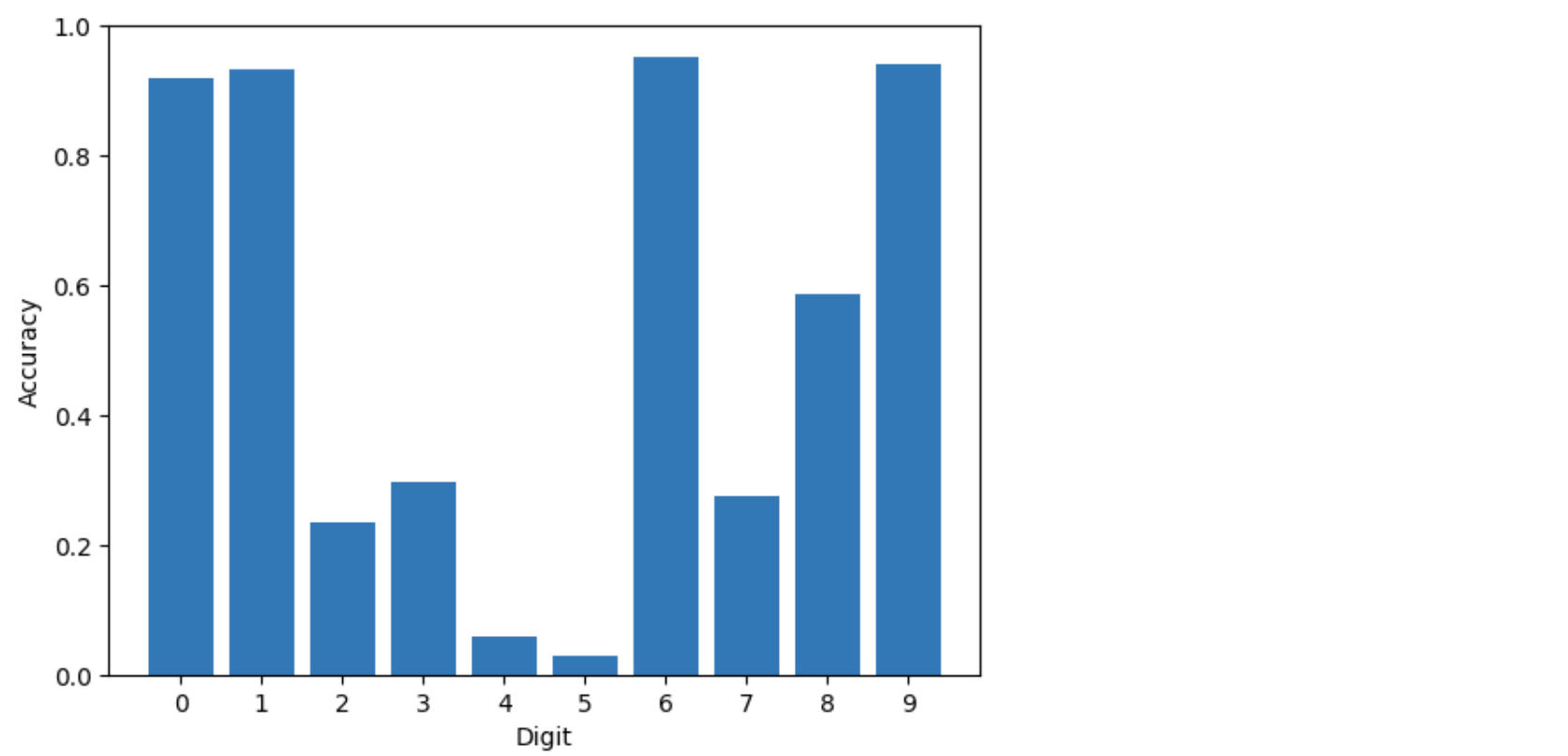


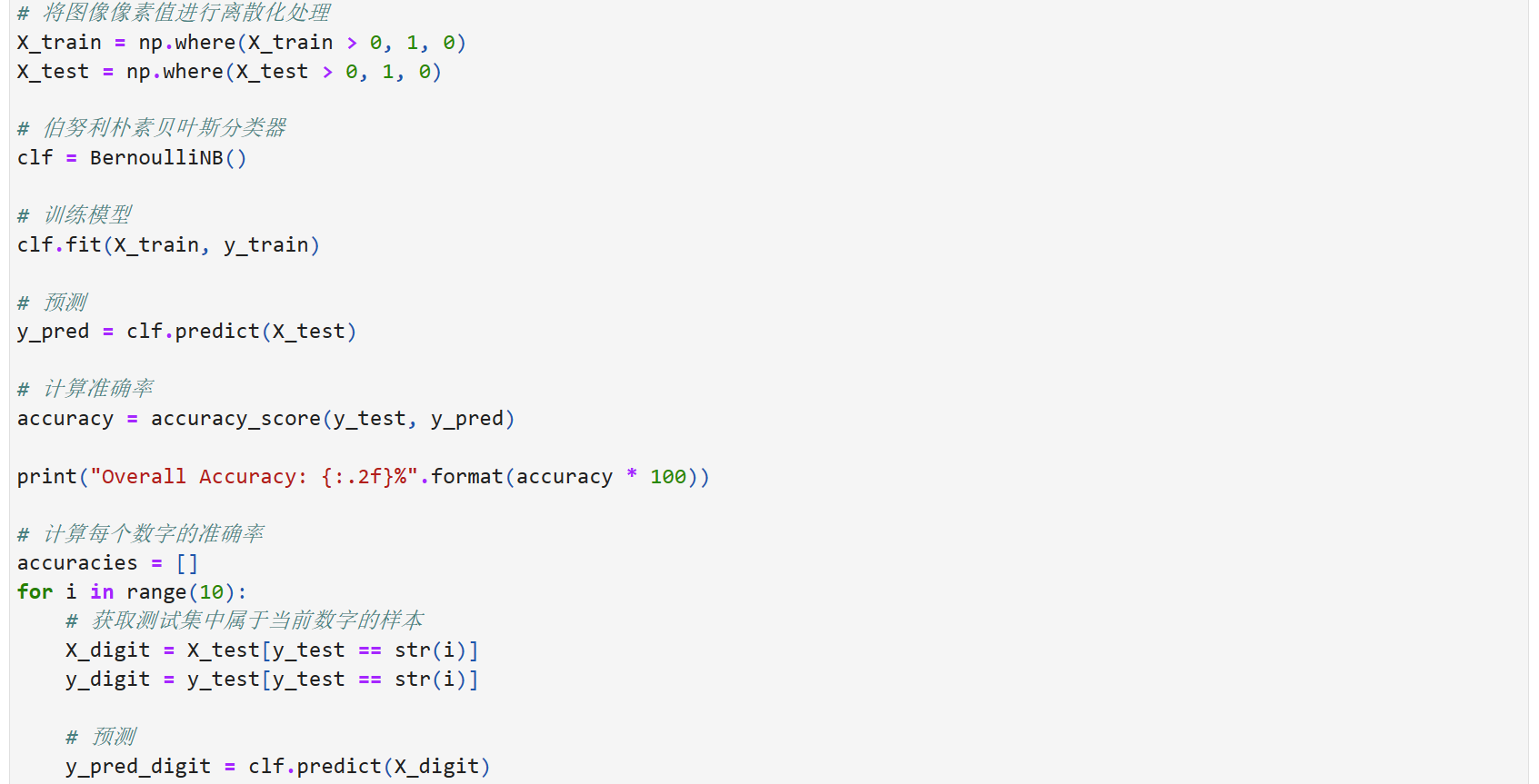


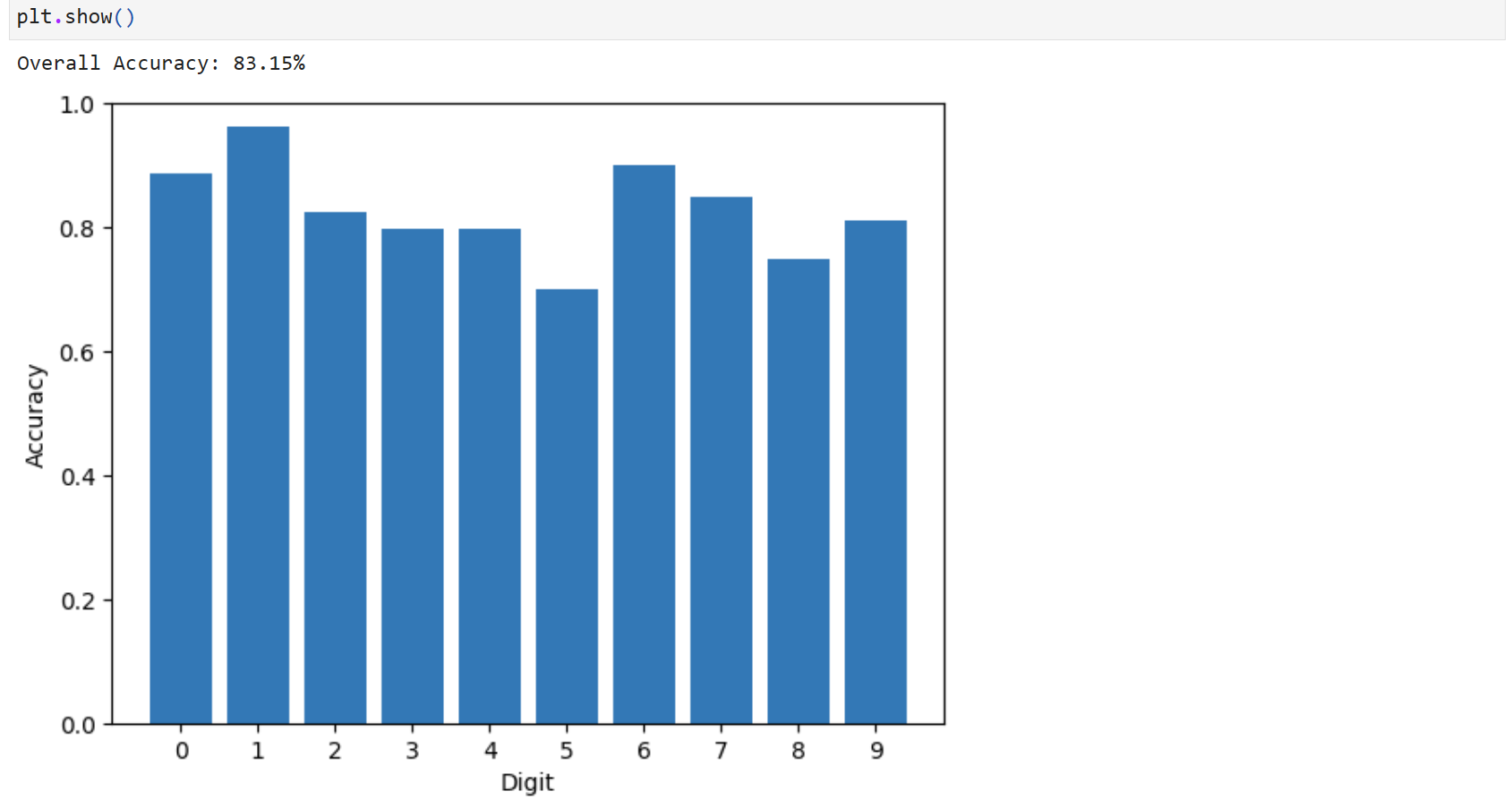




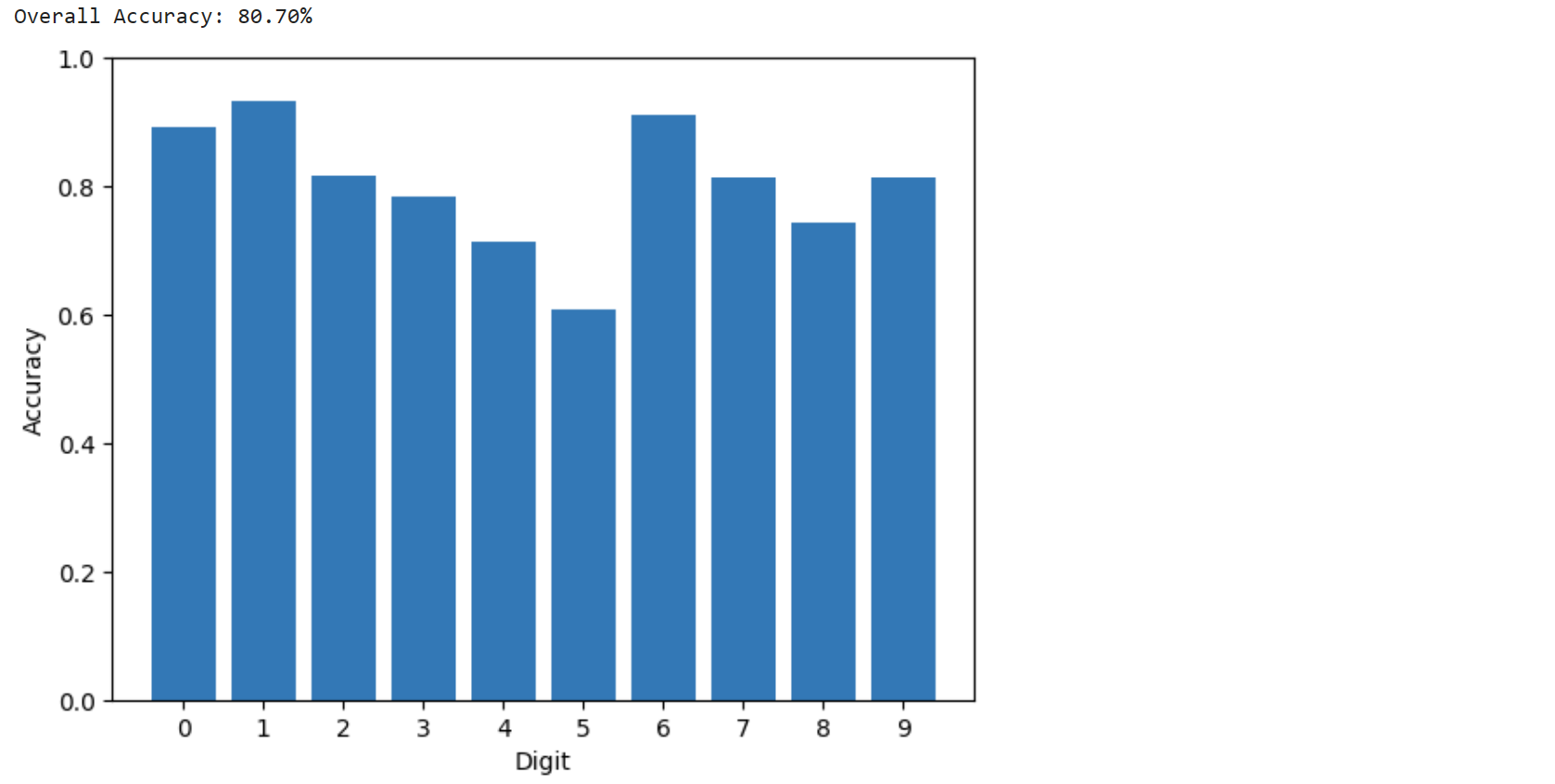




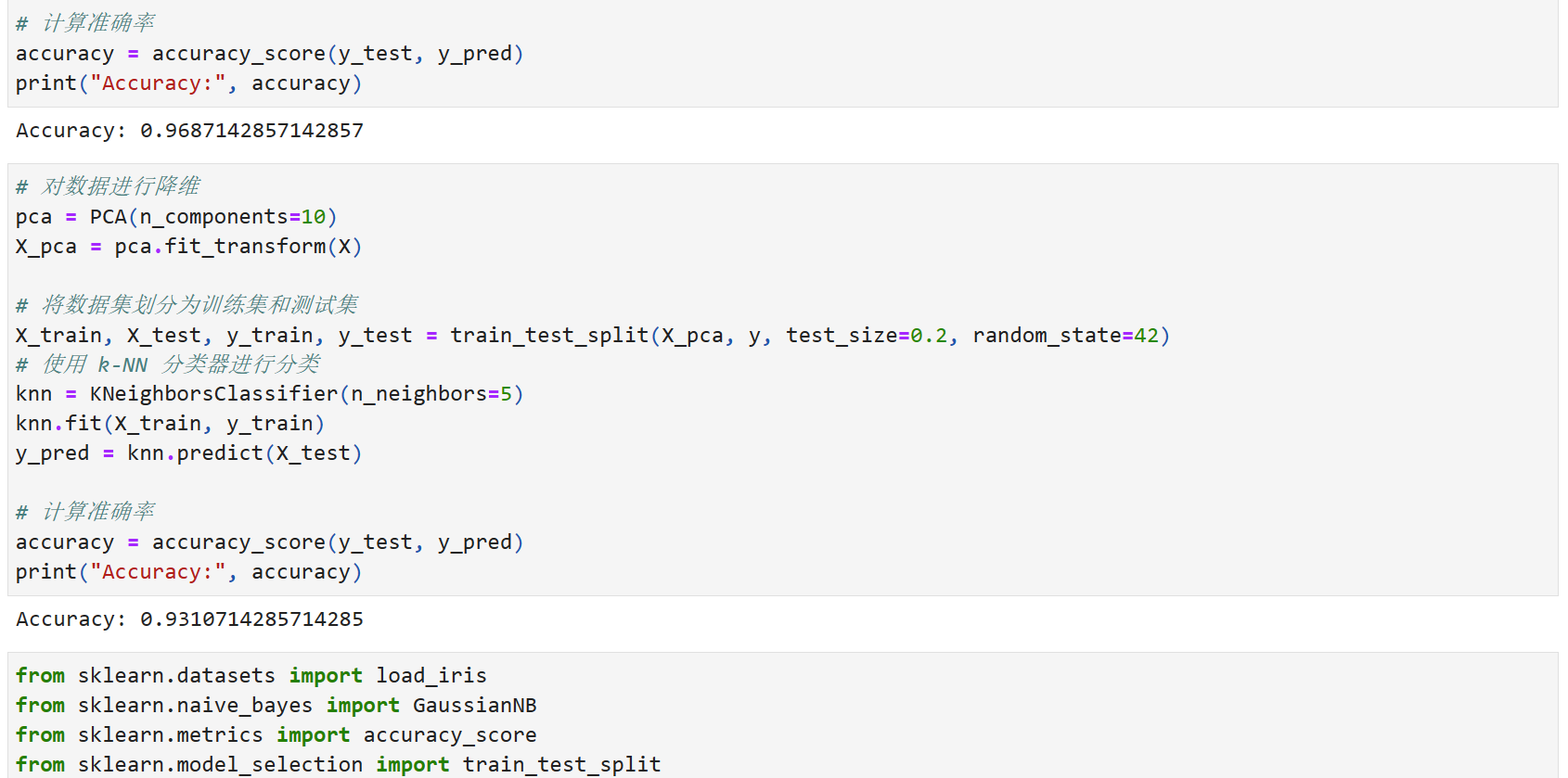














1. **总结（心得体会）**

经过对分类结果的深入分析，我们发现使用不同数量的特征进行分类所得到的结果不尽相同。具体来说，当仅使用一个特征进行分类时，分类准确率较低，但随着特征数量的增加，分类准确率也随之提高。值得注意的是，在使用四个特征进行分类时，我们发现第三个和第四个特征对分类结果具有至关重要的影响。事实上，仅仅在特征集合中包含第三和第四个特征时，我们就可以实现100%的分类准确率，这表明这两个特征是决定性的。因此，我们可以得出结论，选择正确的特征对于提高分类准确率是至关重要的，而在这种情况下，第三和第四个特征被证明是最具识别性和区分性的。

在针对MNIST数据集进行的分类实验中，我们使用了高斯朴素贝叶斯分类器进行分类，并且发现其整体准确率仅略高于50%。此外，我们还注意到，在对不同数字进行分类时，其准确率特别低，特别是2-5这些数字的准确率。为了进一步提高分类效果，我们采用了伯努利和多项式的贝叶斯分类器，其分类准确率都能够达到80%以上，并且在不同数字的分类中，准确率都差不多。

此外，我们还对数据集进行了PCA降维的处理，并使用knn算法进行分类，发现准确率提高到了96.87%。然而，当继续降维时，我们却发现准确率反而下降了，这表明在进行特征选择时需要注意保留关键信息。

最后，我们对朴素贝叶斯分类器在不同平滑参数下的分类准确率进行了定义，并发现在选定的参数下，准确率并没有明显的变化。这表明，在进行分类时，需要考虑到选择适当的参数，并注意到不同参数对分类结果的影响。