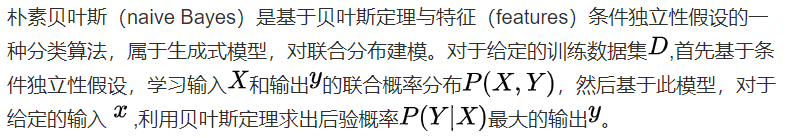
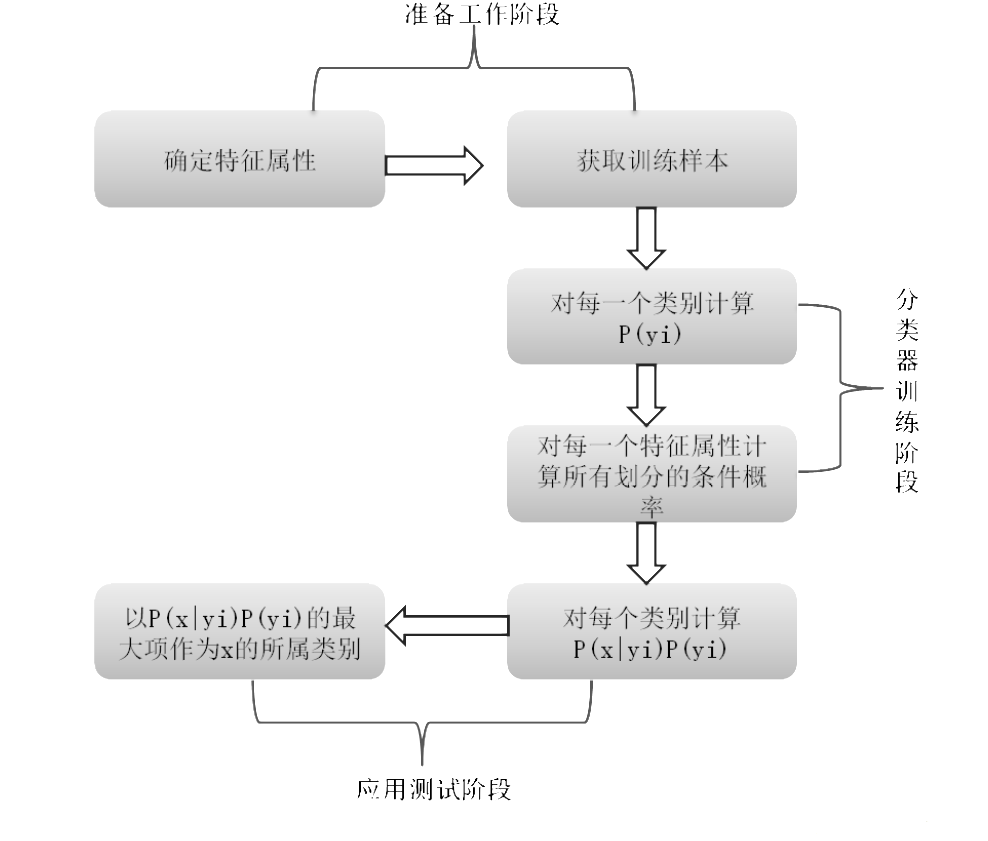
《机器学习》读书笔记

1. 算法概述（算法名称及原理）



1. 算法设计（流程图及主要分段代码，附详细代码注释）





1. 选用数据（数据集描述，包括来源，行数，列数，格式等）

数据集是uci 皮肤分割数据集

[http://archive.ics.uci.edu/ml/datasets/Skin+Segmentation#](http://archive.ics.uci.edu/ml/datasets/Skin+Segmentation)

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

描述信息如下：

* 共有245057行4列，
* 其中属性信息：其中前三列为B，G，R（x1，x2和x3要素）值，第四列为类标签（决策变量y）。

1. 评价方法（说明训练集和测试集分配方法及评价指标）

十折[交叉验证](https://baike.baidu.com/item/%E4%BA%A4%E5%8F%89%E9%AA%8C%E8%AF%81/8543100" \t "https://baike.baidu.com/item/%E5%8D%81%E6%8A%98%E4%BA%A4%E5%8F%89%E9%AA%8C%E8%AF%81/_blank)。将数据集分成十份，轮流将其中9份作为训练数据，1份作为测试数据，进行试验。每次试验都会得出相应的正确率（或差错率）。10次的结果的正确率（或差错率）的平均值作为对算法精度的估计，一般还需要进行多次10折交叉验证（例如10次10折交叉验证），再求其均值，作为对算法准确性的估计。

1. 实验结果截图





1. 实验结果分析及比较

对于鸢尾花数据集，与c4.5，cart，KNN算法相比，NB的验证结果要更优







1. 遇到的问题及解决方法，实践心得

我是按照文本分类的朴素贝叶斯算法进行理解之后选用两个数据集进行测试训练，但是最后的结果是1感觉有点奇怪，最终发现出现这样的原因是在编写代码的时候出了点细节上的错误。