

拉普拉斯修正的朴素贝叶斯分类器

分析:

先通过一个字典保存了贝叶斯分类器的一些信息。

- 1.对于类别型的变量，保存其是好瓜和不是好瓜的概率。
- 2.对于数值型的属性，保存其是好瓜和不是好瓜的均值和方差。
- 3.保存了类别的先验概率。即 $P(c=\text{好瓜})$ 和 $P(c=\text{坏瓜})$ 。

代码:

见源码

结果:

字典:

```
'色泽': {
    '浅白': {'是': 0.18181818181818182, '否': 0.4166666666666667},
    '青绿': {'是': 0.36363636363636365, '否': 0.3333333333333333},
    '乌黑': {'是': 0.45454545454545453, '否': 0.25}
},

'根蒂': {
    '硬挺': {'是': 0.09090909090909091, '否': 0.25},
    '蜷缩': {'是': 0.5454545454545454, '否': 0.3333333333333333},
    '稍蜷': {'是': 0.36363636363636365, '否': 0.4166666666666667}
},

'敲声': {
    '沉闷': {'是': 0.2727272727272727, '否': 0.3333333333333333},
    '浊响': {'是': 0.6363636363636364, '否': 0.4166666666666667},
    '清脆': {'是': 0.09090909090909091, '否': 0.25}
},

'纹理': {
    '清晰': {'是': 0.7272727272727273, '否': 0.25},
```

```
'模糊': {'是': 0.09090909090909091, '否': 0.3333333333333333},  
'稍糊': {'是': 0.18181818181818182, '否': 0.4166666666666667}  
},
```

```
'脐部': {  
    '凹陷': {'是': 0.5454545454545454, '否': 0.25},  
    '平坦': {'是': 0.09090909090909091, '否': 0.4166666666666667},  
    '稍凹': {'是': 0.36363636363636365, '否': 0.3333333333333333}  
},
```

```
'触感': {  
    '硬滑': {'是': 0.7, '否': 0.6363636363636364},  
    '软粘': {'是': 0.3, '否': 0.36363636363636365}  
},
```

```
'密度': {  
    '是': {'平均值': 0.57375, '方差': 0.014608437499999998},  
    '否': {'平均值': 0.49611111111111117, '方差': 0.03370254320987655}  
},
```

```
'含糖量': {  
    '是': {'平均值': 0.27875, '方差': 0.008912437500000002},  
    '否': {'平均值': 0.15422222222222222, '方差': 0.010328617283950618}  
},
```

```
'好瓜': {'是': 0.47368421052631576, '否': 0.5263157894736842}}
```

对测 1 的预测:

```
p1 = 0.02180124640594357  
p0 = 4.915834021416594e-05  
pre = 好瓜
```

分类器的准确率:

```
train data set acc = 0.8235294117647058
```