**作业1 生成式分类器 & 判别式分类器**

作业提交截止日期 **10月9日 23:55**

1. 秋天到了，很多浆果熟了。我们知道有些浆果有毒，有些可以食用。下表给出了一些达人传授的知识（注：该表纯属虚构，不反映现实生活）。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 样本索引 | 是否是黑红色R | 是否成簇生长C | 是否有斑点S | 是否有毒Y |
| 1 | 0 | 0 | 1 | 1 |
| 2 | 1 | 1 | 0 | 1 |
| 3 | 0 | 1 | 1 | 1 |
| 4 | 0 | 0 | 0 | 0 |
| 5 | 0 | 1 | 0 | 0 |
| 6 | 1 | 0 | 1 | 0 |
| 7 | 0 | 0 | 1 | 1 |
| 8 | 1 | 0 | 0 | 1 |

1. 根据上述数据（1-8）训练朴素贝叶斯分类器。计算给定类别的情况下各属性取值的条件概率，例如、，以及各类的先验概率。
2. 给定如下样本，根据（1）中训练好的朴素贝叶斯分类器，预测它们是否有毒。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 样本索引 | 是否是黑红色R | 是否成簇生长C | 是否有斑点S | 是否有毒Y |
| 9 | 1 | 0 | 0 | ？ |

（3）假设将有毒的浆果误判为可食用的代价是20，将可食用的浆果误判为有毒的代价是10，请根据最小风险原则，判断(2)中的浆果应该可食用。

1. 设以下两个类别具有正态概率密度函数，并用样本

1. 求这两类模式之间的贝叶斯判别界面的方程式。

(2) 绘出判别界面。

1. 给定两个类别的样本，每类包括5个样本，每个样本有3维特征。
2. 如果它们是线性可分的，权向量的维度是多少？
3. 假如要建立二次多项式判别函数，权向量的维度是多少？
4. 给定两类样本，其中

类别1有5个样本：

类别2有6个样本：

1. 求这两类的Fisher判别投影向量、判别函数，并在图上画出；
2. 计算每个样本点投影后的坐标。
3. 对给定的如下样本，根据Gini指数，构造二叉决策树：

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| 3 | 2 | 5 | 1 |
| 5 | 7 | 7 | 0 |
| 8 | 6 | 9 | 1 |
| 1 | 3 | 2 | 0 |
| 6 | 4 | 8 | 1 |
| 4 | 5 | 6 | 0 |

1. 根结点的Gini指数为多少？
2. 请给出第一个决策规则（选择的特征及阈值）。