**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。

答：类别先验：、

各属性值的类条件概率估计采用Laplace平滑估计

、

、

、

、

、

、

1. 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 | ？ |

答：

所以该浆果有毒。

（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。

答：根据题意，

所以判定该浆果有毒。

1. 设以下两个类别具有正态概率密度函数，并用样本

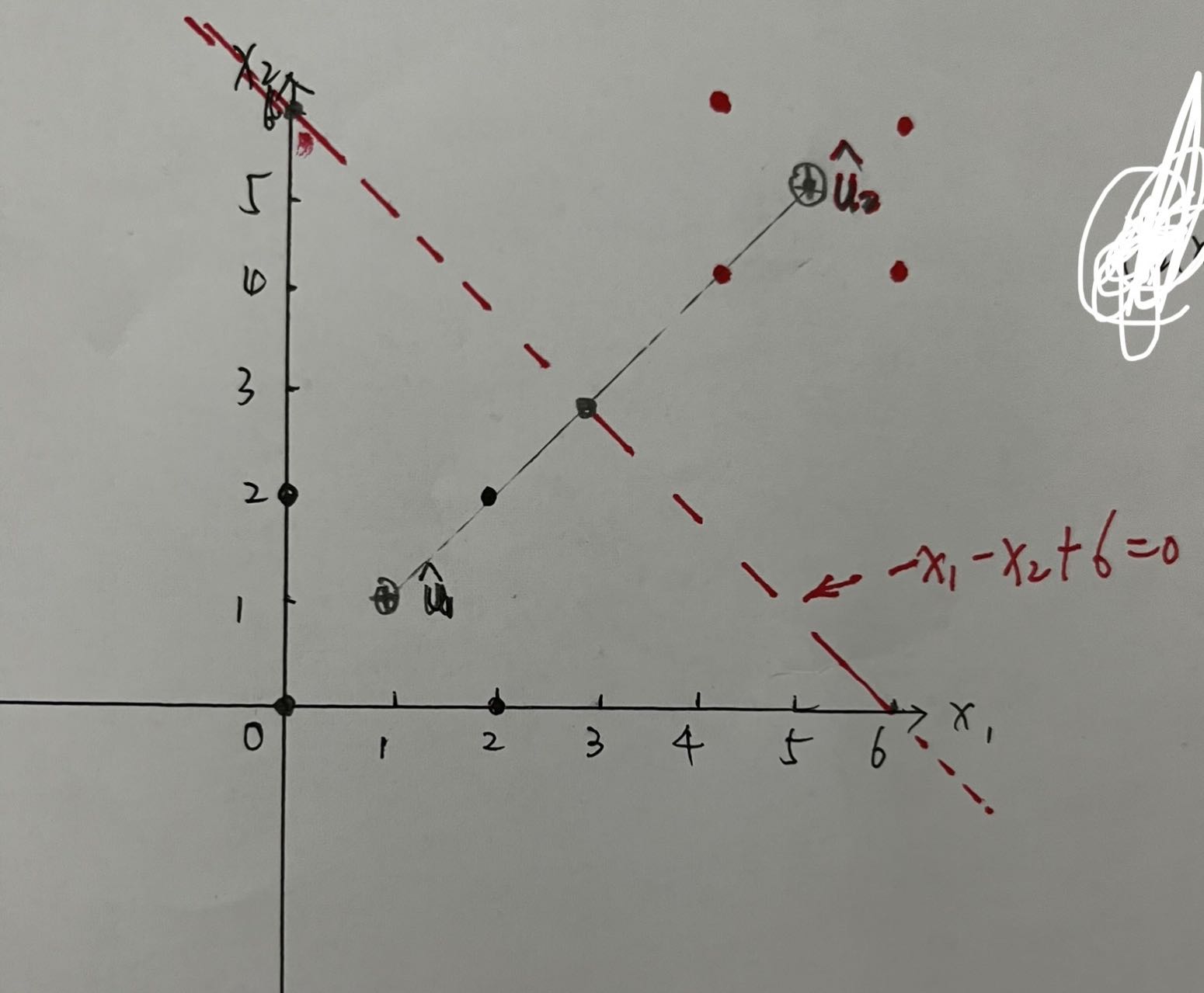
1. 求这两类模式之间的贝叶斯判别界面的方程式。

答：

，所以

所以判别函数为的线性函数：

(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？

答：线性模型的权向量的维度同输入模型的维度为3。还有额外的一个偏差参数。

1. 假如要建立二次多项式判别函数，权向量的维度是多少？

答：维向的次多项式，权系数的总项数为

1. 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；

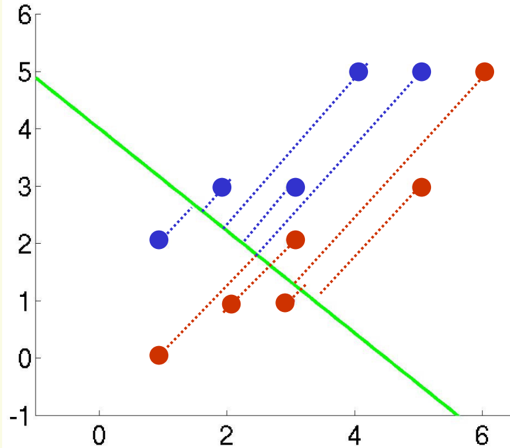
答：

分类阈值为：

判别面方程为：

即。

1. 计算每个样本点投影后的坐标。



1. 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？

答：（1）

1. 请给出第一个决策规则（选择的特征及阈值）。

答：根据特征 的取值，对所有样本进行排序：得到1，3，4，5，6，8，所以可能的阈值为：，3.5， 4.5，5.5，7。

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

根据每个可能的阈值将数据集划分成左右分支，划分后的阈值分别为：

所以对特征，决策规则为时最佳，划分后的Gini指数为。

对特征类似处理，所以可能的阈值为：，3.5， 4.5，5.5，6.5, 根据每个可能的阈值将数据集划分成左右分支，划分后的阈值分别为：

这里最小的Gini指数大于已有的最佳决策的Gini指数，所以不考虑特征。

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

对特征类似处理，所以可能的阈值为：，5.5， 6.5，7.5，8.5, 根据每个可能的阈值将数据集划分成左右分支，划分后的阈值分别为：

这里最小的Gini指数等于已有的最佳决策的Gini指数，所以第一次的最佳决策规则为或者。

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