**** 成 绩

模式识别与机器学习实验报告

**实验二 贝叶斯分类器的综合设计**

**院（系）名称自动化科学与电气工程学院**

**专 业 名 称 自动控制与模式识别**

**学 生 姓 名 雷诗叶**

**学 生 学 号 15031148**

**2018年5月1日**

**目录**

[一、 实验绪论](#_Toc10986_WPSOffice_Level1) [1](#_Toc10986_WPSOffice_Level1)

[二、 实验原理](#_Toc24024_WPSOffice_Level1) [1](#_Toc24024_WPSOffice_Level1)

[三、 实验内容](#_Toc7615_WPSOffice_Level1) [1](#_Toc7615_WPSOffice_Level1)

[3.1步骤一](#_Toc24024_WPSOffice_Level2) [1](#_Toc24024_WPSOffice_Level2)

[3.2步骤二](#_Toc7615_WPSOffice_Level2) [4](#_Toc7615_WPSOffice_Level2)

[3.3步骤三](#_Toc20262_WPSOffice_Level2) [4](#_Toc20262_WPSOffice_Level2)

[3.4讨论](#_Toc21770_WPSOffice_Level2) [5](#_Toc21770_WPSOffice_Level2)

[四、 实验结论与收获](#_Toc20262_WPSOffice_Level1) [5](#_Toc20262_WPSOffice_Level1)

1. 实验绪论

线性感知器允许我们找到一个可以将两类分开的决策边界。当只有两类时，它们是非常有效的并且分离得很好。这样的分类器被称为判别分类器。

相反，生成分类器将每个样本视为一个随机特征向量，并通过其分布或密度函数明确地对每个类进行建模。分类时，应该为属于候选类的给定样本计算似然函数，以便将样本分如最有可能的一类中。换句话说，我们需要对于每个类ωi计算p(ωi | X)。但是，密度函数仅提供了特定样本的可能性，因为样本属于特定类别,即密度函数可以提供p(X| ωi)。贝叶斯规则为我们提供了一种从密度函数和相关信息计算给定样本类别可能性的方法。

1. 实验原理

贝叶斯方法的本质是提供一个数学规则解释如何根据新的证据改变现有的观念。换句话说，它使我们能够将新数据与他们现有的知识或专业知识结合起来。典型的例子是想象一个早熟的新生儿观察他的第一个日落，并想知道太阳是否会再次升起。他事先对两种可能的结果分配了相同的可能性，并通过在袋子里放置一个白色和黑色弹子来表示。第二天日出后，孩子把另一个白色弹子放在袋子里。随机抽出白色弹子的概率（即孩子对未来日出的信念度）因此从一半变到到三分之二。下一天日出后，孩子把另一个白色弹子放在袋子里，而概率从三分之二到四分之三。渐渐地，最初太阳可能不是每天早晨升起的观念被转变成为太阳几乎确定性地每天都会升起。

就分类而言，贝叶斯定理允许我们将先验概率以及观察到的证据结合起来得到后验概率。此外，条件概率代表假定证据后事件发生的概率。根据贝叶斯定理，如果p(ωi)、p(ωi | X)，i=1,2,·,c以及X都是已知的，那后验概率可由下式得出：



我们把一系列决策行为称为，决策ai的条件风险可以被计算为：



因此最小风险贝叶斯决策即为：

1. 实验内容

3.1步骤一

1. 输入数据x1，x2两类样本，通过贝叶斯方法寻找最小错误率决策以及最小风险决策。步骤流程图如图3.1。

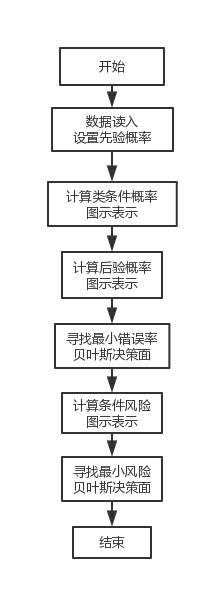


图3.1 程序实现流程图

1. 运行程序得到最终结果如下图

结果：decision boundary is : 1.9295。当观测到的特征值x > 1.9295，则细胞会被判定为abnormal，若x <= 1.9295 时，细胞被判定为normal

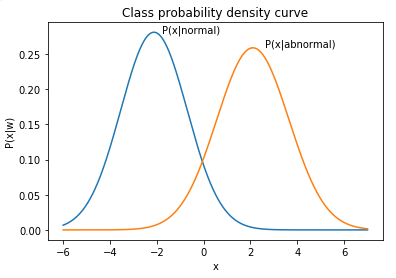


图3.2 类条件概率密度曲线

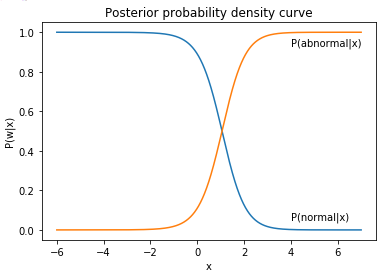


图3.3 后验概率曲线

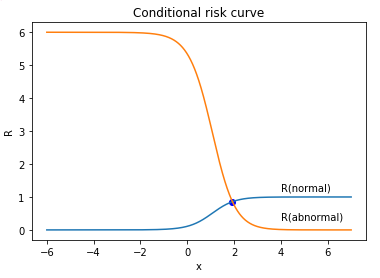


图3.4 条件风险曲线

3.2步骤二

建立二维数据集x1，x2，x3，均符合不同的正态分布。每类数据集中各有60个数据点，结果如下图所示。

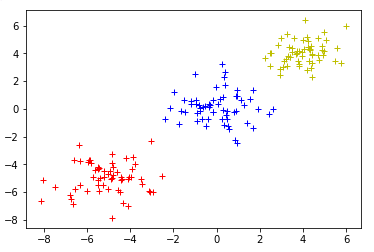


图3.5 初始数据集散点图

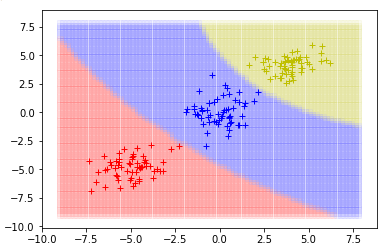


图3.6 类条件概率

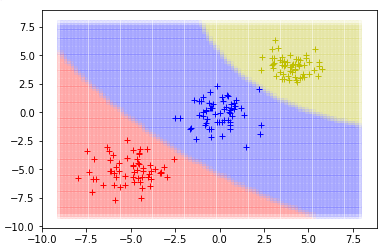


图3.7 后验概率

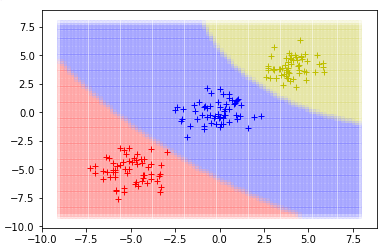


图3.8 条件风险分布

3.3讨论

最小错误率贝叶斯决策与最小风险贝叶斯决策的区别？

最小错误率贝叶斯决策即是简单的根据概率来进行分类。例如对于某个待测样本x，算出x属于哪类的可能性最大即把它分到哪类。

而最小风险贝叶斯决策则是将错误划分的代价考虑在内。例如将一个正常人判断为癌症患者和将癌症患者判断为正常人的代价是不一样的。尽管有人的检测结果基本符合正常人的分类，但我们还是会倾向于将其划分到癌症患者一类。

对于贝叶斯决策，一般认为先验概率是已知的，依据由贝叶斯公式计算出后验概率，根据后验概率大小进行分类，认为把待测样本决策为后验概率大的一类错误率较小，即为最小错误率贝叶斯决策。最小错误率贝叶斯决策可以看做是最小风险贝叶斯决策的特例，即各类之间错误划分的损失函数均相等。

需要指出，最小风险贝叶斯决策中的决策表是需要人为确定的，决策表不同会导致决策结果的不同。因此，在实际应用中，需要认真分析所研究问题的内在特点和分类的目的，与应用领域的专家共同设计出适当的决策表，才能保证模式识别发挥有效的作用。

1. 实验结论与收获

本次实验通过python编程实现了最小风险贝叶斯决策，以及最小错误率贝叶斯决策（特殊的最小风险贝叶斯决策），使自身对于贝叶斯决策的基本原理以及运行过程有了比较深入的了解。

最小风险的贝叶斯决策的一般步骤为：

1. 若未知类概率密度函数，则通过样本分别及已知条件计算出类概率密度函数；
2. 根据已知的先验概率以及类概率密度函数，利用贝叶斯公式求出后验概率；
3. 根据决策表以及后验概率计算条件风险
4. 决策：在各种决策中选取风险最小的决策。画出决策面

令决策表中错误分类的损失均为1，则为最小错误率贝叶斯决策。

通过本次实验，我对于贝叶斯分类器的基本原理有了深入的了解，并掌握了如何实现最小风险贝叶斯决策。