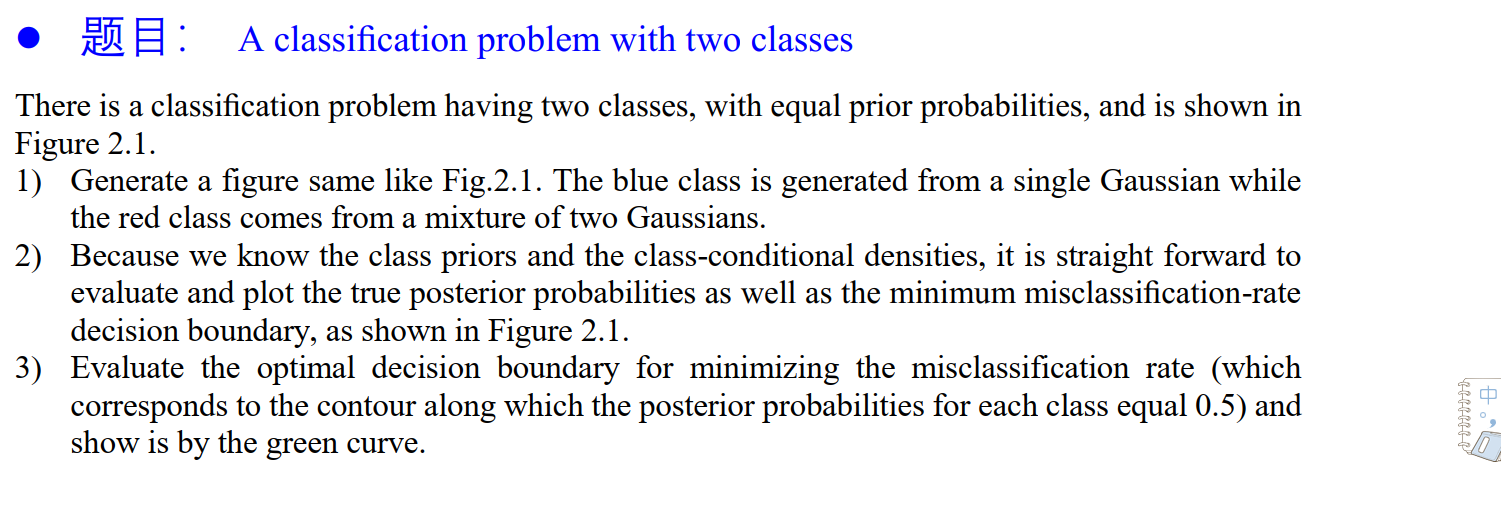
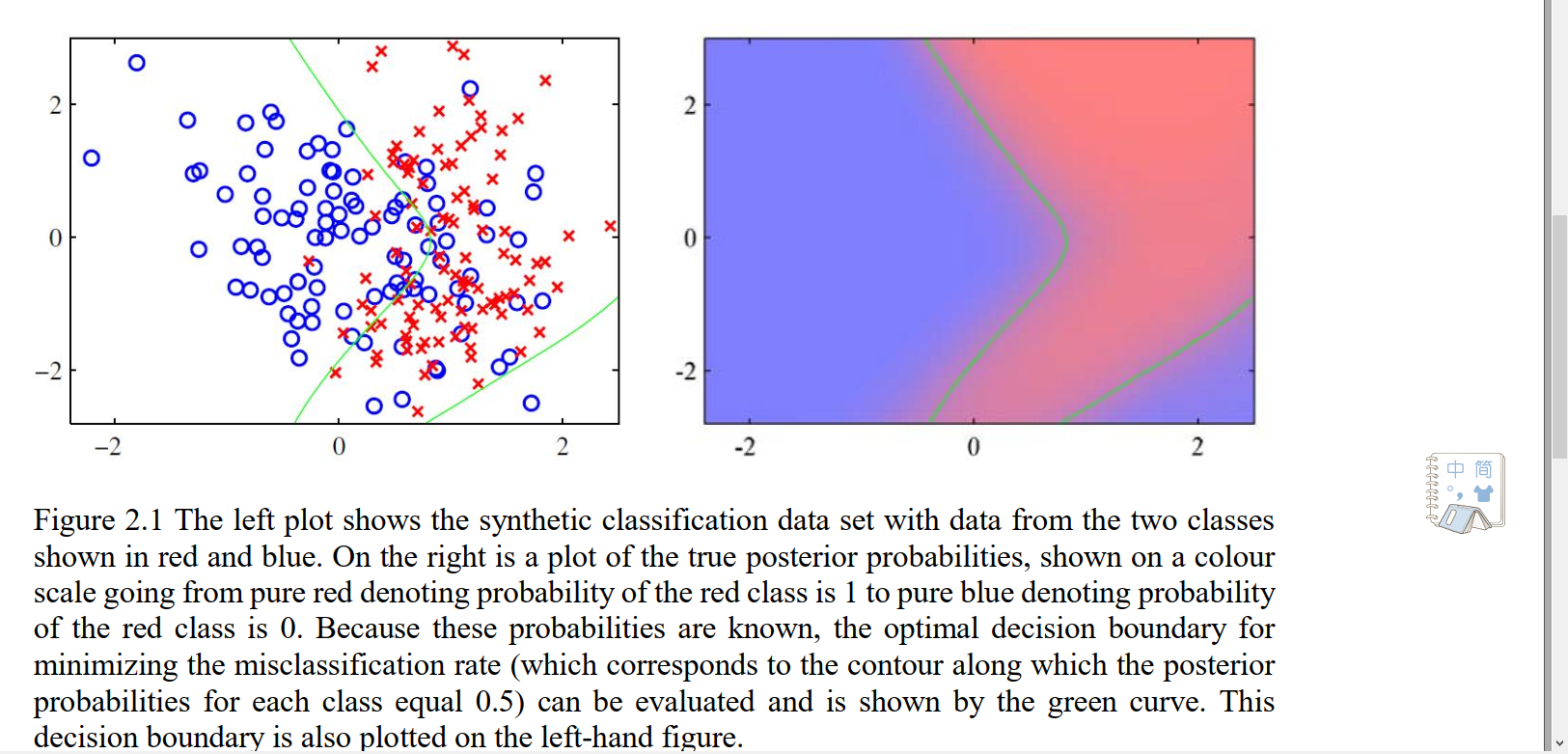
# Machine Learning Homework 4

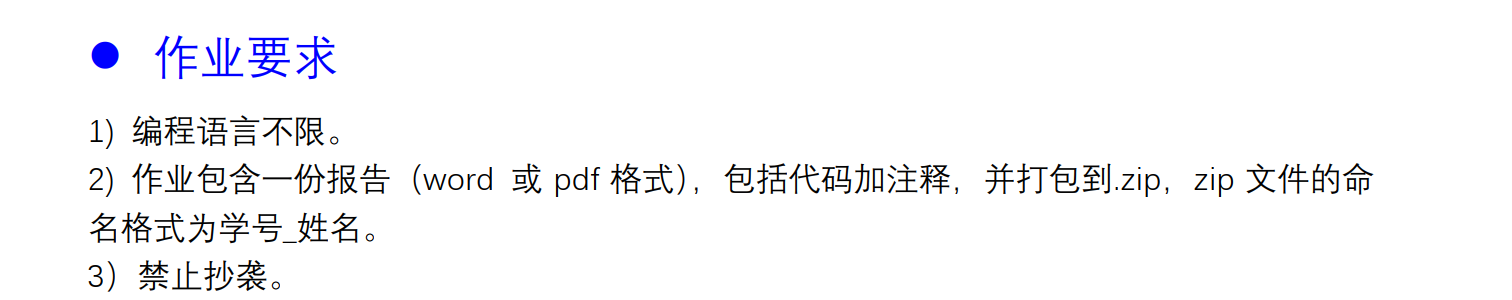
专业 :计算机技术 姓名 ：严晓威 学号 ：18214679

1. 实验题目



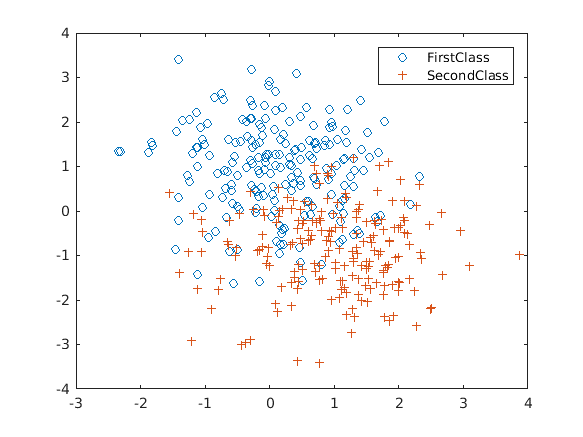


1. 实验要求

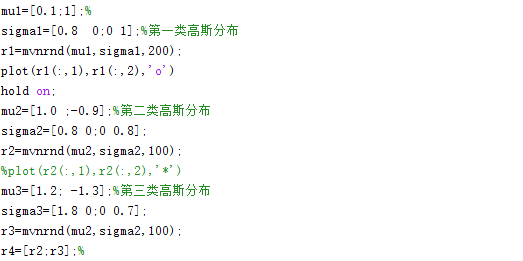


1. 实验过程及代码

首先产生第一类的一个二维高斯分布记为c0（共有200个样本点），然后分别产生两个二维高斯分布分别记为c2,c3分别含有100个样本点，并将其混合分布记为c1作为第二个高斯分布，然后画出的结果如下所示

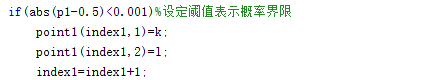


蓝色代表第一类高斯分布，此二维高斯分布的mu=和,sigma=,红色代表第二个分布(两个高斯混合分布)，c2的mu=和sigma=然后c3的mu=，sigma=,作为判别依据的概率密度函数如下所示，点属于c0的概率，，



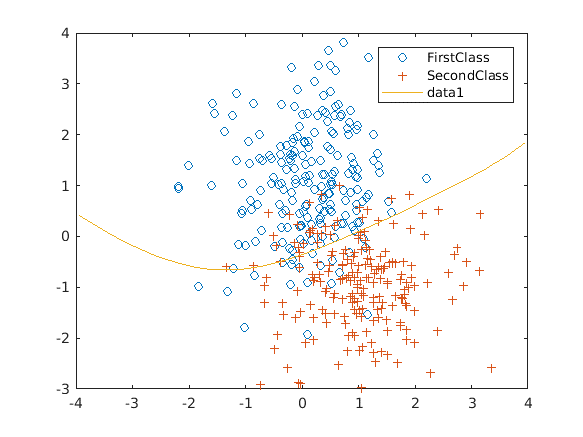


点属于c1的概率则为,其中先验概率由样本数决定，由于我们上述产生的点c0和c1分别有200个样本点所以先验概率分别为0.5，后验概率通过把样本点代入到高斯分布的概率密度分布函数中求取，c0的直接带入pdf求解，而c1是由c2，c3两个高斯分布组成所以c1的概率拆分为两个高斯分布的概率密度分布函数各取一半（为0.5）产生。所以就得到了产生了计算上述概率的方法，我们要找的就是使上述概率值为0.5的点，这样的话表示该点属于c0或者c1的概率都是0.5，也就是决策边界所带有的特征，所以就产生了两个聚类的决策边界。然后把符合这些特征的采样值画出来就可以可视化决策边界。我们根据上图的样本分布区间将x[-4:4]和y[-4,4]区间上分别以0.02的间隔采样，共产生400\*400个点，然后对这些点计算相应的概率，带入到此公式中，要是概率值接近某个设定的

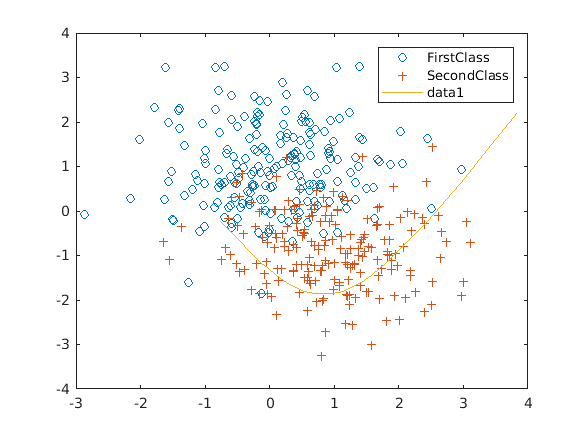


阈值，则可以判定该点可以作为决策边界上。将阈值设为0.001所得到决策边界结果如下图所示。

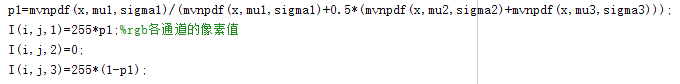


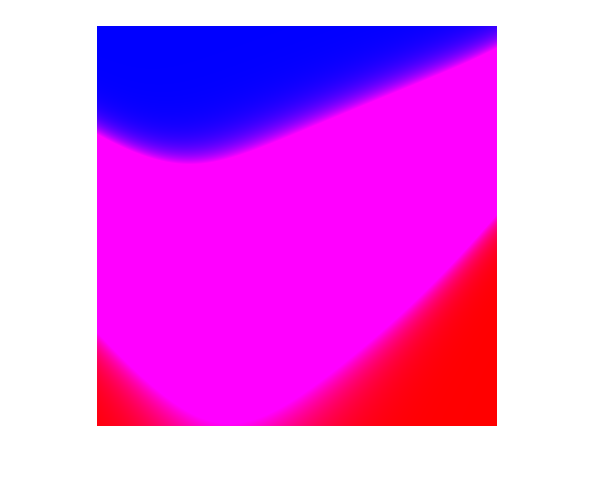


可以看到决策边界基本符合预期，成功将一类二类区分出来。下面一张图为混合高斯分布的决策边界。



然后根据网格中各个节点的概率判断他是属于哪一个类，然后将其绘制成不同的颜色。采用rbg三通道，蓝色表示该区域属于第一类，红色属于第二类，粉色表示该区域属于第一类和第二类的。





上图和两个决策边界的轮廓极为相似，这也从另一方面验证了本方法的合理性。

1. 实验结果与分析

本次作业主要是使用概率分布模型来解决问题，通过充分利用高斯分布的各种先验信息，通过贝叶斯的相关内容来，一开始考虑可以使用logistic regression的方法来进行分类，但是这样的话就没有体现出高斯分布的信息，而是单纯的把此问题看成一个二分类问题，通过迭代优化的思路来优化产生最优决策边界，这样就体现不出各种先验信息的作用。所以本次主要还是使用概率的知识来解决问题。

1. 总结