# 实验报告一

学号：

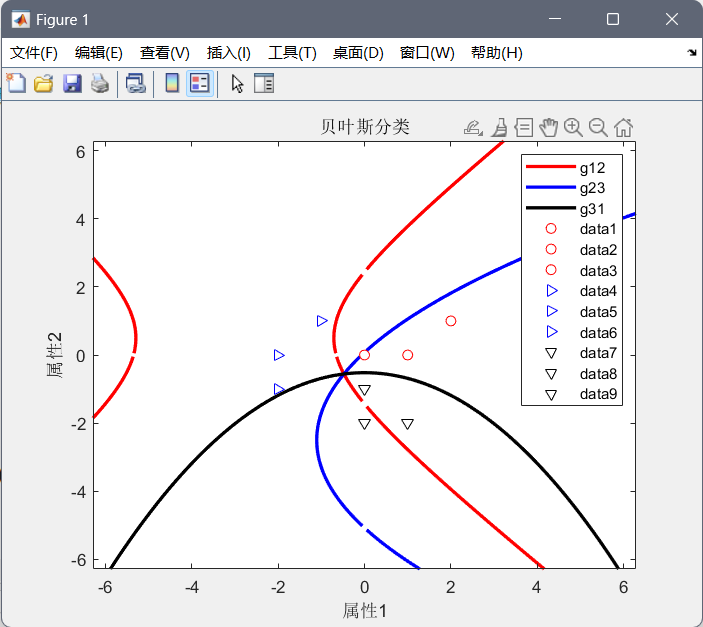
姓名：***Steven***

实验名称：基于最小错误率的贝叶斯分类器

实验内容：使用MATLAB编程环境，设计基于最小错误率的贝叶斯分类器，并进行实验。

实验要求及结果：

1. 运行demo\_1.m文件，给出测试样本x\_test=[-2, 2]的分类结果，用不同颜色画出三类模式的分界线，用不同颜色和形状画出三类模式的训练样本，请分别贴出分类结果的截图、绘制的分界线和训练样本示意图。



(图1.1 分界线和训练样本示意图)

文本

描述已自动生成

(图1.2 x\_test分类结果)

1. 改写上述文件中给出的示例程序，采用iris数据集进行贝叶斯分类实验，给出测试样本x\_test =[6, 3.5, 4.5, 2.5]的分类结果（请给出分类结果的截图）。

文本

描述已自动生成

(图2.1 x\_test分类结果)

问题：

请简述基于最小错误率的贝叶斯分类器的工作原理。

对于分类问题，获取不同类的先验概率和类条件概率，利用贝叶斯公式计算观测样本X属于不同类的后验概率，认为属于后验概率最大的类。

附采用iris数据集进行贝叶斯分类实验的代码：

|  |
| --- |
| close all;  clear all;  clc;  %读取数据  fileID = fopen("Iris.data"); % 打开文件  data = textscan(fileID, '%f%f%f%f%s', 'Delimiter', ',');  attrib1 = data{1, 1}; attrib2 = data{1, 2}; attrib3 = data{1, 3}; attrib4 = data{1, 4}; class = data{1, 5};  attrib = [attrib1, attrib2, attrib3, attrib4];  label\_set = char('Iris-setosa', 'Iris-versicolor', 'Iris-virginica');  label = zeros(150, 1);  label(strcmp(class, 'Iris-setosa')) = 1;  label(strcmp(class, 'Iris-versicolor')) = 2;  label(strcmp(class, 'Iris-virginica')) = 3;  x1 = zeros(length(find(label == 1)), 4);  x2 = zeros(length(find(label == 2)), 4);  x3 = zeros(length(find(label == 3)), 4);  for i = 1:length(label)  if label(i) == 1  x1(i, :) = attrib(i, :);  elseif label(i) == 2  x2(i, :) = attrib(i, :);  else  x3(i, :) = attrib(i, :);  end  end  % 求取各类的均值，协方差矩阵及其逆矩阵  u1 = mean(x1); u2 = mean(x2); u3 = mean(x3); % 三类训练样本的均值  c1 = cov(x1); c2 = cov(x2); c3 = cov(x3); % 三类训练样本的协方差矩阵  t1 = diag(c1); t2 = diag(c2); t3 = diag(c3);  c1 = diag(t1); c2 = diag(t2); c3 = diag(t3); % 将三类训练样本的协方差矩阵简化为对角矩阵，这样做是为了满足可逆性，方便后续的求逆操作  inv\_c1 = inv(c1); inv\_c2 = inv(c2); inv\_c3 = inv(c3); % 三类训练样本的协方差矩阵的逆矩阵  d1 = det(c1); d2 = det(c2); d3 = det(c3); % 三类训练样本的协方差矩阵的行列式  % 给定一个测试样本x\_test,根据公式(2-39)判断x\_test的类别归属  x\_test = [6, 3.5, 4.5, 2.5];  p1 = -0.5 \* (x\_test - u1) / c1 \* (x\_test - u1)' - 0.5 \* log(d1);  p2 = -0.5 \* (x\_test - u2) / c2 \* (x\_test - u2)' - 0.5 \* log(d2);  p3 = -0.5 \* (x\_test - u3) / c3 \* (x\_test - u3)' - 0.5 \* log(d3);  [~, max\_id] = max([p1, p2, p3]);  fprintf('x\_test属于第%d类，为%s\n', max\_id, label\_set(max\_id, :)); |