判别分析（discriminant analysis）是根据所研究的个体的观测指标来推断该个体所属类型的一种统计方法，在自然科学和社会科学的研究中经常会碰到这种统计问题。例如在地质找矿中我们要根据某异常点的地质结构、化探和物探的各项指标来判断该异常点属于哪一种矿化类型；医生要根据某人的各项化验指标的结果来判断该人属于什么病症；调查了某地区的土地生产率、劳动生产率、人均收入、费用水平、农村工业比重等指标，来确定该地区属于哪一种经济类型地区等等。该方法起源于 1921 年 Pearson 的种族相似系数法，1936 年 Fisher 提出线性判别函数，并形成把一个样本归类到两个总体之一的判别法。

例 15 某种产品的生产厂家有 12 家，其中 7 家的产品受消费者欢迎，属于畅销品，定义为 1 类；5 家的产品不大受消费者欢迎，属于滞销品，定义为 2 类。将 12 家的产品的式样，包装和耐久性进行了评估后，得分资料见表 1。

表1 生产厂家数据

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 厂家 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 式样 | 9 | 7 | 8 | 8 | 9 | 8 | 7 | 4 | 3 | 6 | 2 | 1 | 6 | 8 | 2 |
| 包装 | 8 | 6 | 7 | 5 | 9 | 9 | 5 | 4 | 6 | 3 | 4 | 2 | 4 | 1 | 4 |
| 耐久性 | 7 | 6 | 8 | 5 | 3 | 7 | 6 | 4 | 6 | 3 | 5 | 2 | 5 | 3 | 5 |
| 类别 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 待判 | 待判 | 待判 |

今有 3 家新的厂家，得分分别为（6，4，5），（8，1，3），（2，4，5），试对 3 个新厂家进行分类。

MATLAB源代码：

clc,clear

train=[9 7 8 8 9 8 7 4 3 6 2 1

8 6 7 5 9 9 5 4 6 3 4 2

7 6 8 5 3 7 6 4 6 3 5 2]';

sample=[6 4 5; 8 1 3; 2 4 5];

group=[ones(7,1);2\*ones(5,1)]; %已知样本的分类

[x1,y1]=classify(sample,train,group,'mahalanobis'); %马氏距离分类

fprintf('三个样本利用马氏距离判别结果分别为：');disp(x1');

fprintf('误判率为：');disp(y1);

[x2,y2]=classify(sample,train,group,'linear'); %线性分类

fprintf('三个样本利用线性判别结果分别为：');disp(x2');

fprintf('误判率为：');disp(y2);

[x3,y3]=classify(sample,train,group,'quadratic'); %二次分类

%函数classify的第二个返回值为误判率

fprintf('三个样本利用二次判别结果分别为：');disp(x3');

fprintf('误判率为：');disp(y3);

求得利用马氏距离、线性分类和二次分类方法都把厂家1，2分在第1类，厂家3分在第2类。