**3.5 编程实现线性判别分析，并给出西瓜数据集3.0α上的结果。**

**①问题描述：**

根据数据集中西瓜的密度、含糖率及其是否为好瓜的分类结果，实现线性判别分析（LDA），并计算和绘制出最终结果。在对原始数据进行处理时，选择将好瓜标记为1，而坏瓜标记为0，以此进行接下来的数据处理。

**②实现过程：**

a.加载已导入txt文件中的数据集，根据分类结果对数据集进行分类，并创建图表绘制出数据集的可视化对象。

b.计算运算过程中可能用到的各均值向量。

c.建立运算过程中各符号所对应的矩阵。

d.计算类内散度矩阵。

e.计算类间散度矩阵。

f.计算矩阵。

g.计算的最大的d个特征值和对应特征向量，得到投影方向向量。

h.计算投影直线方程。

i.计算数据集在直线上投影点的位置。

**③实现结果：**

计算得投影方向向量为：

-0.194541249617840

-0.980894337937135

好瓜投影点坐标：

密度 含糖率

0.114158098281464 0.575595316951691

0.101043012751675 0.509467885556025

0.0743721971117934 0.374991253475597

0.0836927115221679 0.421986118727862

0.0620697897811472 0.312961418068800

0.0604774432427057 0.304932664749597

0.0466369063697385 0.235147442955388

0.0568027827016242 0.286404698440814

坏瓜投影点坐标：

密度 含糖率

0.0425706556690391 0.214645044123344

0.0601467679016070 0.303265370176255

0.0201493343456885 0.101594741535348

0.0318728967606747 0.160705989231672

0.0549065143455363 0.276843544200519

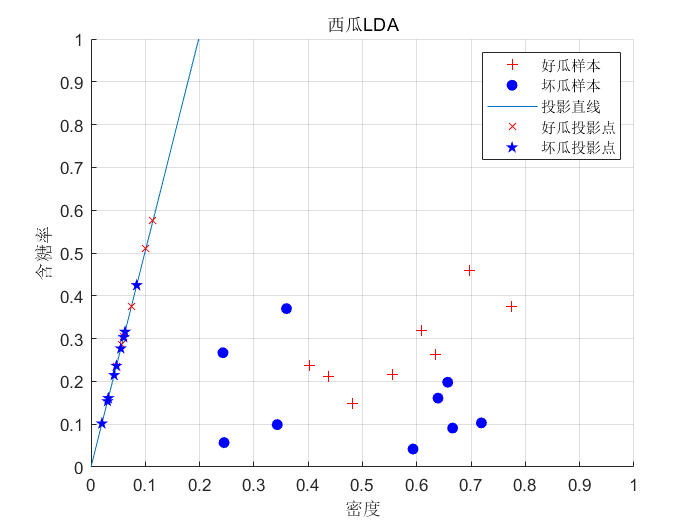
0.0626482508850661 0.315878070566229

0.0842296989998144 0.424693657501265

0.0304574798274072 0.153569330767774

0.0468664023755355 0.236304582292714

LDA可视化结果：



**④结果分析：**

该给定训练样例投影到直线上后，其同类样例的投影点都较为接近，但是异类样例的投影点没有尽可能远离，导致好瓜和坏瓜的判别容易产生混淆，可能将导致对西瓜分类的准确度不高等问题。