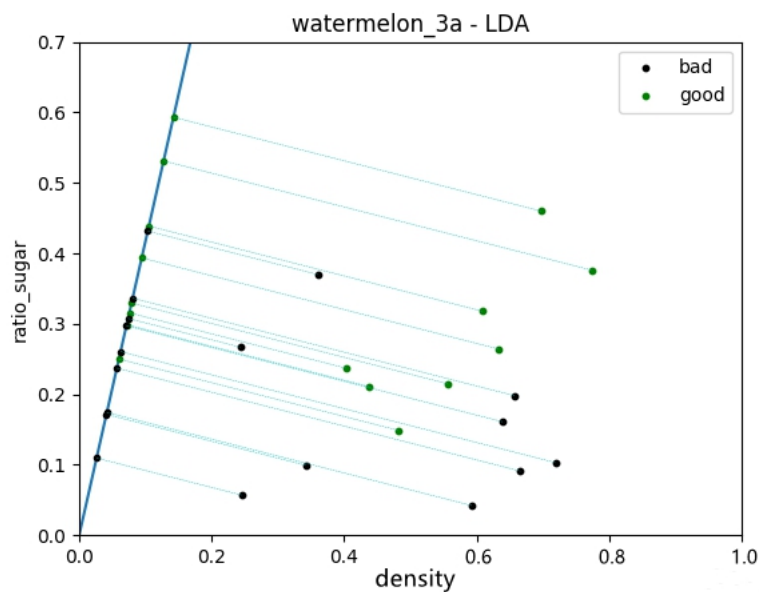


编辑实现线性判别分析，并给出西瓜数据集 3.0a 上的结果

首先绘制散点图。然后用 `sklearn` 实现一下 LDA。要注意模型参数的选择，对于小数据一般选择 `lsqr`，也可以在官方的 [reference^{\[1\]}](#) 进行查找。

接着是 LDA 实现的具体步骤， \mathbf{u} 是均值向量，对于二分类问题，每个个体有两个属性值，因此均值向量求出来是 2×2 的矩阵。根据西瓜书 P61 的公式 3.33 求出类内散度矩阵，注意列向量是乘以自己的转置，最后的类内散度矩阵 (within-class scatter matrix) 是 2×2 的，记为 \mathbf{S}_w 。这里根据公式 3.39，我们求出 \mathbf{S}_w 的逆矩阵就可以了，然而这里我们不用 `np.linalg.inv()` 来求逆矩阵，而是要考虑到数值的稳定性，采用先用奇异值分解 (SVD)，再用分解出的矩阵得到一个类似原 \mathbf{S}_w 的逆矩阵的东西。然后根据公式 3.39 就能得到 \mathbf{w} ，即一个方向向量 (x, y) ，然后数据垂直映射即可。

最后得到了以下结果：



参考资料：

[1]http://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html#sklearn.discriminant_analysis.LinearDiscriminantAnalysis

[2]《机器学习》 [中]周志华 P60-63

[3]https://github.com/PY131/Machine-Learning_ZhouZhihua/blob/master/ch3_linear_model/3.5_LDA/src/LDA.py