一、

已知





假设







对于样例

由





得









由





得





由



得



由



得





由









得





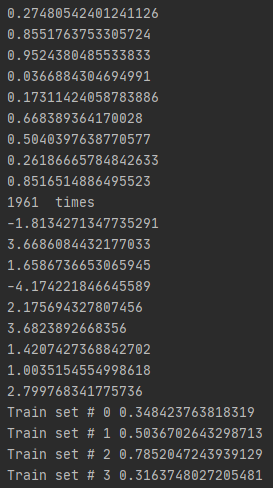




对于样例

对于样例

对于样例



二、

(1) 1.

线性判别的思想在于，让同类样例投影点的协方差尽可能小，即令类内散度矩阵尽可能小。同时，让不同类中心之间的距离尽可能大，即类间散度矩阵尽可能大。

但该问题难以直接求解，因而引入了正交投影矩阵，将该问题降维，同时引入广义瑞利商的概念，将问题转化为最大化

因而线性判别的结果，也即对于最大化的问题而言，其结果只与正交投影矩阵的投影方向有关，而与比例因子无关。

1. 2.

应用拉格朗日乘子法求解极值的必要条件是目标函数与约束函数相切。

为了一般化最大化的求解，通常引入的约束条件是，因而

随机变量的分布律为



故似然函数为



而





故



三、

最大信息增益算法

数据预处理

import pandas as pd

from sklearn.datasets import load\_iris

iris = load\_iris()

data = pd.DataFrame(iris['data'])

data[4] = iris['target']

data: pd.DataFrame = data.sample(frac=1.0)

split = int(150 \* 0.75)

data\_train: pd.DataFrame = data.iloc[0: split, :]

data\_train = data\_train.reset\_index(drop=True)

data\_test: pd.DataFrame = data.iloc[split: 150, :]

data\_test = data\_test.reset\_index(drop=True)

data\_train.to\_csv('train.csv')

data\_test.to\_csv('test.csv')