

基于 matlab 实现的 CCA 算法

程序说明和测试



姓名： 魏一鸣

日期： 2019 年 8 月 16 日

一、 CCA 算法说明及测试目的

假设有两组多维度的数据，CCA 算法可以对两组数据分别进行线性变换后，得到两组数据的最大线性相关系数，并同时得到两个线性变换向量。

二、 CCA 程序说明

1. 输入

输入为两组数据 X 和 Y，两组数据分别包含在两个文件夹中——“XFile.txt”和“YFile.txt”。在文件中，数据的行向量代表一个数据样本，列向量代表同一维度下各个数据样本的值。程序要求数据样本的维度大于 1 维，且 X 和 Y 的数据组数相等，数据的维度可以不相等。

2. 输出

程序会计算得到 n (n 为 X 和 Y 数据维度的较小值) 个线性变换方式。程序输出会按照每个线性变换后线性相关系数从大到小的顺序输出该线性变换、该线性变化下线性相关系数和正则化系数

(正则化系数详见读书报告和程序注释)，并且程序会为每个线性变化后的两组数据样本绘制离散点图，顺序与各自线性变换一致。

三、 结果验证

首先说明，本实验数据集来自百度文库相同问题，但是采用的不同的算法过程实现，本实验对 CCA 算法实现采用奇异值分解实现，数据集来源实验采用特征分解和拉格朗日函数求解。

1. 问题描述

欲研究儿童形态与肺通气功能的关系，测得某小学 20 名 8 到 12 岁的健康儿童（身高 X1，体重 X2，胸围 X3）与肺通气通能（肺活量 Y1，静息通气 Y2，每分钟最大通气量 Y3），分析儿童形态和肺通气指标的相关性。

2. 输入数据

X1	X2	X3	Y1	Y2	Y3
140.6	43.7	77	2.6	7	108
135.7	39.5	63	2	7	91
140.2	48	75	2.6	6.1	101
152.1	52	88	2.8	10.1	112
132.2	36	62	2.1	7.4	97
147.1	45	78	2.8	9.25	92
147.5	47	76	3.1	8.78	95
130.6	38	61	2	5.31	77
154.9	48	87	2.9	10.6	80
142.4	42	74	2.33	11.1	76
136.5	38	69	1.98	7.77	49
162	58	95	3.29	3.35	58
148.9	42	80	2.7	10.1	82
136.3	33	68	2.4	7.8	76
159.5	49	87	2.98	11.77	88
165.9	55	93	3.1	13.14	110
134.5	41	61	2.25	8.75	75
152.5	53	83	2.96	6.6	71

138.2	35.5	66	2.13	6.62	105
144.2	42	76	2.52	5.59	82

3. 测试结果

试验测试结果与原数据集所属实验结果相同，得到相同的线性变换和相同的散点图。

1) 线性变换 1

该线性变换计算所得线性相关系数最大，是实验最终结果。

对于标准化且转置后的 X 的数据矩阵的线性变换 a 为

$$a = [-0.1590 \quad -0.0348 \quad -0.0405]^T$$

对于标准化且转置后的 Y 的数据矩阵的线性变换 b 为

$$b = [-0.2181 \quad -0.0407 \quad -0.0184]^T$$

此时线性相关系数为

$$\rho = 0.9282$$

该线性变换下的两组数据的散点图为

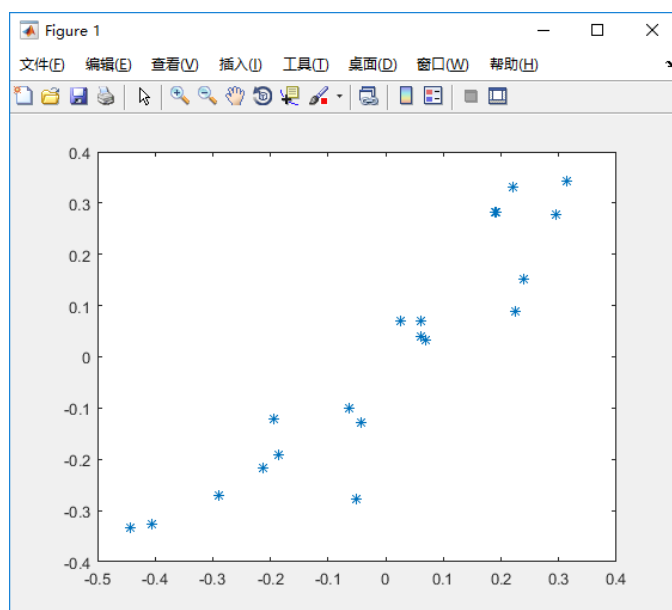


图 1 线性变换 1 下儿童形态与肺通气功能关系散点图

实验数据输出结果截图为

```
=====第1个线性变换=====
线性变换为:
a =
    -0.1590
    -0.0348
    -0.0405

b =
    -0.2181
    -0.0407
     0.0184

rou =
     0.9282

gama =
     0
```

图 2 线性变化 1 实验数据输出

2) 线性变换 2

对于标准化且转置后的 X 的数据矩阵的线性变换 a 为

$$a = [0.6589 \quad -0.3730 \quad -0.3228]^T$$

对于标准化且转置后的 Y 的数据矩阵的线性变换 b 为

$$b = [-0.0909 \quad 0.2445 \quad -0.0753]^T$$

此时线性相关系数为

$$\rho = 0.5302$$

该线性变换下的两组数据的散点图为

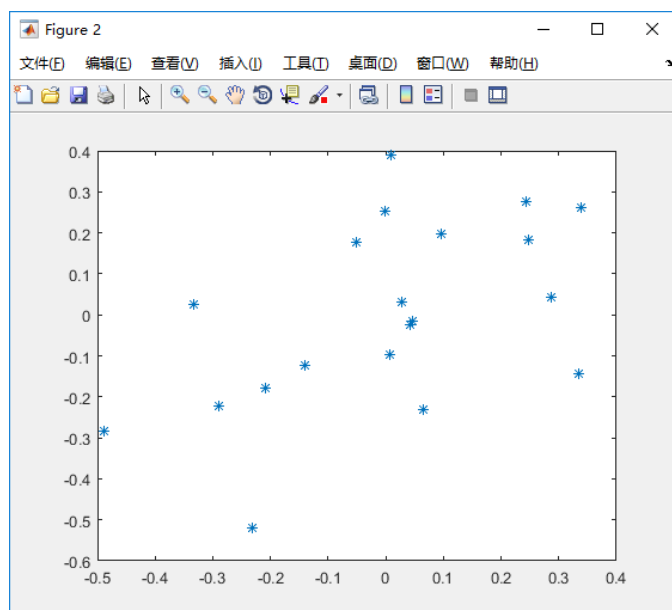


图 3 线性变换 2 下儿童形态与肺通气功能关系散点图

实验数据输出结果截图为

```
=====第2个线性变换=====
线性变换为:
a =
    0.6589
   -0.3730
   -0.3228

b =
   -0.0909
    0.2445
   -0.0753

rou =
    0.5302

gama =
    0
```

图 4 线性变化 2 实验数据输出

3) 线性变换 3

对于标准化且转置后的 X 的数据矩阵的线性变换 a 为

$$a = [-0.5290 \quad -0.3272 \quad 0.8433]^T$$

对于标准化且转置后的 Y 的数据矩阵的线性变换 b 为

$$b = [0.0249 \quad -0.0075 \quad -0.2289]^T$$

此时线性相关系数为

$$\rho = 0.0081$$

该线性变换下的两组数据的散点图为

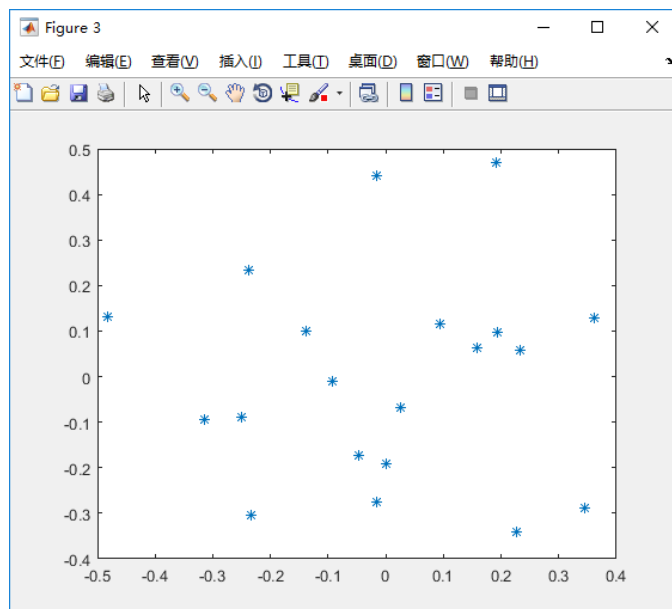


图 5 线性变换 3 下儿童形态与肺通气功能关系散点图

实验数据输出结果截图为

```
=====第3个线性变换=====
线性变换为:
a =
    -0.5290
    -0.3272
     0.8433

b =
     0.0249
    -0.0075
    -0.2289

rou =
     0.0081

gama =
     0
```

图 6 线性变化 2 实验数据输出

参考文献

匿名 典型相关分析(CCA)附算法应用及程序 • 百度文库。

<https://wenku.baidu.com/view/f110b66017fc700abb68a98271fe910ef12daedc.html>