

線形判別分析

青木繁伸

1 目的

線形判別分析を行う。

2 使用法

```
from disc import disc
disc(data, verbose=True)
```

2.1 引数

`data` 分類に必要な説明変数と、群を表す変数が最終列（最も右側の列）になるように用意されたデータフレーム
`verbose` 必要最小限のプリント出力をする

2.2 戻り値の名前

"ncase"	サンプルサイズ
"p"	説明変数の個数
"vname"	説明変数の名前
"gVname"	群変数の名前
"ng"	群の数
"gName"	群の名前
"num"	各群のサンプルサイズ
"t"	全体の変動・共変動行列
"w"	群内変動・共変動
"names2"	2群判別の対象とする群の名前
"cFunction"	分類関数の係数
"dFunction"	判別関数の係数
"partialF"	分類関数の係数の検定統計量
"partialP"	分類関数の係数の検定統計量の p 値
"df1"	第1自由度
"df2"	第2自由度
"wilksLambda"	ウィルキスの Λ
"wilksLambdaF"	ウィルキスの Λ の検定統計量
"wilksLambdaP"	ウィルキスの Λ の検定統計量の p 値
"wilksLambdaDf1"	第1自由度

"wilksLambdaDf2"	第2自由度
"distance"	各群の重心からの距離 (χ^2 分布にしたがう)
"df"	自由度 (説明変数の個数に等しい)
"Pvalue"	各群の所属する確率 p
"prediction"	どの群に所属するか判別
"correct"	正しい判別がされたデータ数
"correctTable"	判別表
"correctRate"	正判別率
"discriminantValue"	判別値 (2群の場合)

3 使用例

3.1 3群判別の場合

```
import pandas as pd

data = pd.read_csv("data/iris.csv")

import sys
sys.path.append("statlib")
from disc import disc

a = disc(data)
```

coefficients of discriminant functions

	setosa:versicolor	setosa:virginica	versicolor:virginica	\
sl	-7.845958	-11.098318	-3.252360	
sw	-16.515361	-19.902591	-3.387230	
pl	21.642090	29.197184	7.555094	
pw	23.832640	38.477524	14.644884	
constant	13.455863	-18.059850	-31.515713	

	F	p value
sl	4.721152	1.032884e-02
sw	21.935928	4.831201e-09
pl	35.590175	2.756205e-13
pw	24.904333	5.143154e-10
constant	NaN	NaN

coefficients of classification functions

	setosa	versicolor	virginica
sl	-47.088333	-31.396418	-24.891698
sw	-47.175741	-14.145020	-7.370559
pl	32.861278	-10.422902	-25.533090
pw	34.796822	-12.868458	-42.158226

```
constant 170.419715 143.507990 206.539415
```

```
result of discrimination
```

	setosa	versicolor	virginica
setosa	50	0	0
versicolor	0	48	2
virginica	0	1	49

```
correct rate = 98.0 %
```

3.2 2 群判別の場合

```
data = data.iloc[0:100, :]  
  
a = disc(data)
```

```
coefficients of discriminant functions
```

	setosa:versicolor	F	p value
sl	-3.052770	0.715527	0.399741
sw	-18.022959	25.700452	0.000002
pl	21.766195	24.597600	0.000003
pw	30.844165	10.705452	0.001490
constant	-13.961740	NaN	NaN

```
coefficients of classification functions
```

	setosa	versicolor
sl	-50.078491	-43.972951
sw	-36.233259	-0.187341
pl	7.889082	-35.643307
pw	70.340670	8.652340
constant	173.031449	200.954928

```
result of discrimination
```

	setosa	versicolor
setosa	50	0
versicolor	0	50

```
correct rate = 100.0 %
```