

総当たり法による線形判別分析

青木繁伸

2020年3月17日

1 目的

総当たり法による線形判別分析を行う。

2 使用法

```
import sys
sys.path.append("statlib")
from multi import APSS_disc
APSS_disc(dat, sort_by="Correct rate")
```

2.1 引数

<code>dat</code>	群を表す変数が最右列にあるようなデータフレーム
<code>sort_by</code>	デフォルト ("Correct rate") では正判別率の大きい順に並べ替える。 "Wilks' lambda" を指定すると、Wilks' の λ の小さい順に並べ替える。 なお、並べ替えに使う Wilks' の λ は第1線形判別解のものである。

2.2 戻り値

結果の総括表を返す。

3 使用例

```
import pandas as pd
import sys
sys.path.append("statlib")
from multi import APSS_disc

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

以下のようなデータフレームにおいて、sl, sw, pl, pwの4変数を用いてspを予測(判別)する。

```
data.head()
```

```
   sl  sw  pl  pw   sp
0  5.1  3.5  1.4  0.2 setosa
1  4.9  3.0  1.4  0.2 setosa
2  4.7  3.2  1.3  0.2 setosa
3  4.6  3.1  1.5  0.2 setosa
4  5.0  3.6  1.4  0.2 setosa
```

結果の表示

```
APSS_disc(data)
```

	Correct rate	Wilks' lambda p	variables
6	98.000000	5.036981e-113	[sw, pl, pw]
14	98.000000	1.365006e-112	[sl, sw, pl, pw]
10	97.333333	1.144793e-106	[sl, pl, pw]
4	96.666667	4.570512e-102	[sw, pw]
9	96.666667	8.415775e-101	[sl, pl]
13	96.666667	1.085030e-105	[sl, sw, pl]
0	96.000000	4.169446e-85	[pw]
2	96.000000	7.532352e-98	[pl, pw]
8	96.000000	1.568188e-82	[sl, pw]
12	96.000000	2.084625e-102	[sl, sw, pw]
5	95.333333	2.848870e-103	[sw, pl]
1	94.666667	2.856777e-91	[pl]
11	80.000000	1.297560e-55	[sl, sw]
7	74.666667	1.669669e-31	[sl]
3	55.333333	4.492017e-17	[sw]

```
APSS_disc(data, sort_by="Wilks' lambda p")
```

	Correct rate	Wilks' lambda p	variables
6	98.000000	5.036981e-113	[sw, pl, pw]
14	98.000000	1.365006e-112	[sl, sw, pl, pw]
10	97.333333	1.144793e-106	[sl, pl, pw]
13	96.666667	1.085030e-105	[sl, sw, pl]
5	95.333333	2.848870e-103	[sw, pl]
12	96.000000	2.084625e-102	[sl, sw, pw]
4	96.666667	4.570512e-102	[sw, pw]
9	96.666667	8.415775e-101	[sl, pl]
2	96.000000	7.532352e-98	[pl, pw]
1	94.666667	2.856777e-91	[pl]

0	96.000000	4.169446e-85	[pw]
8	96.000000	1.568188e-82	[sl, pw]
11	80.000000	1.297560e-55	[sl, sw]
7	74.666667	1.669669e-31	[sl]
3	55.333333	4.492017e-17	[sw]