判別分析（二次の判別関数）

青木繁伸

1　目的

二次の判別関数を使った判別分析を行う。

R の **MASS** パッケージに `qda()`, `predict.qda()` がある。

2　使用法

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

2.1　引数

`data`	データフレーム（最後列が群を表す変数）
`verbose`	必要最小限のプリント出力をする

2.2　戻り値の名前

`"group means"`	群ごとの平均値
`"group var.s"`	群ごとの分散・共分散行列
`"group inv-var.s"`	群ごとの分散・共分散行列の逆行列
`"scores"`	判別スコア
`"p values"`	p 値
`"predict"`	判別結果
`"correct table"`	判別結果表
`"correct rate"`	正判別率
`"vname"`	説明変数の名前

3　使用例

```
import pandas as pd

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

import sys
sys.path.append("statlib")
```

```
from quad_disc import quad_disc

a = quad_disc(data)
```

各群ごとの平均値

	sl	sw	pl	pw
sp				
setosa	5.006	3.428	1.462	0.246
versicolor	5.936	2.770	4.260	1.326
virginica	6.588	2.974	5.552	2.026

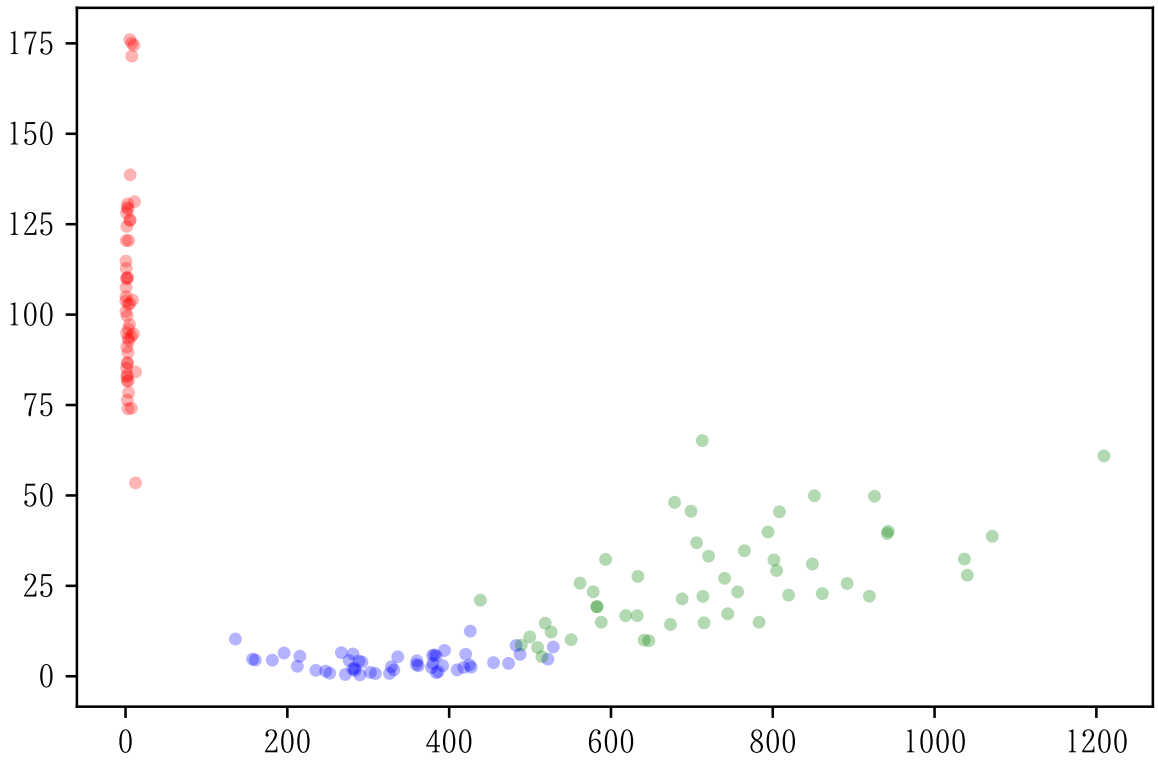
判別結果表

prediction	setosa	versicolor	virginica
sp			
setosa	50	0	0
versicolor	0	47	3
virginica	0	0	50

正判別率 = 98.0 %

```
import scipy as sp
import matplotlib.pyplot as plt

scores = a["scores"]
color = sp.repeat(["red", "blue", "green"], 50)
plt.scatter(scores[:, 0], scores[:, 1], c=color, s=9, alpha=0.3)
plt.show()
```



群ごとの平均値

```
a["group means"]
```

```
array([[5.006, 3.428, 1.462, 0.246],
       [5.936, 2.77 , 4.26 , 1.326],
       [6.588, 2.974, 5.552, 2.026]])
```

群ごとの分散・共分散行列

```
a["group var.s"]
```

```
array([[[0.12424898, 0.09921633, 0.0163551 , 0.01033061],
        [0.09921633, 0.1436898 , 0.01169796, 0.00929796],
        [0.0163551 , 0.01169796, 0.03015918, 0.00606939],
        [0.01033061, 0.00929796, 0.00606939, 0.01110612]],

       [[0.26643265, 0.08518367, 0.18289796, 0.05577959],
        [0.08518367, 0.09846939, 0.08265306, 0.04120408],
        [0.18289796, 0.08265306, 0.22081633, 0.07310204],
        [0.05577959, 0.04120408, 0.07310204, 0.03910612]],

       [[0.40434286, 0.09376327, 0.3032898 , 0.04909388],
        [0.09376327, 0.10400408, 0.07137959, 0.04762857],
        [0.3032898 , 0.07137959, 0.30458776, 0.04882449],
        [0.04909388, 0.04762857, 0.04882449, 0.07543265]]])
```

群ごとの分散・共分散行列の逆行列

```
a["group inv-var.s"]
```

```
array([[[ 18.94343879, -12.40482616,  -4.50020654,  -4.77612733],
        [-12.40482616,  15.57054017,   1.11107914,  -2.10409783],
        [ -4.50020654,   1.11107914,  38.77620413, -17.9350353 ],
        [ -4.77612733,  -2.10409783, -17.9350353 , 106.04590614]],

       [[  9.50276377,  -3.67621661,  -8.63171191,   6.45450344],
        [ -3.67621661,  19.71096643,   2.1160224 , -19.4803247 ],
        [ -8.63171191,   2.1160224 ,  19.80375773, -26.93722701],
        [  6.45450344, -19.4803247 , -26.93722701,  87.24479383]],

       [[ 10.5338668 ,  -3.47972624,  -9.96014594,   1.78815223],
        [ -3.47972624,  15.87544225,   1.1026887 ,  -8.47285053],
        [ -9.96014594,   1.1026887 ,  13.40582053,  -2.89091847],
        [  1.78815223,  -8.47285053,  -2.89091847,  19.31405035]]])
```

判別スコア

```
a["scores"][0:10, :]
```

```
array([[  0.44911379, 114.80448926, 182.9359087 ],
       [  2.08109416,  83.31536333, 153.97494977],
```

```
       [  1.28433511,  94.92042508, 160.49414815],
       [  1.70620698,  82.77879889, 140.64139404],
       [  0.76168538, 120.48102441, 184.03694564],
       [  3.71264736, 120.4800689 , 183.29808569],
       [  3.42419613,  96.0034728 , 154.01866346],
       [  0.34343924, 103.81486506, 167.44412555],
       [  2.9964765 ,  73.9469893 , 132.76555812],
       [  3.20008592,  93.49589445, 156.73861306]])
```

p 値

```
a["p values"][0:10, :]
```

```
array([[9.78261918e-01, 6.86991632e-24, 1.74567505e-38],
       [7.20846451e-01, 3.45379036e-17, 2.86278901e-32],
       [8.64027652e-01, 1.18488690e-19, 1.14537123e-33],
       [7.89589812e-01, 4.48817183e-17, 2.05742920e-29],
       [9.43509606e-01, 4.21615702e-25, 1.01263801e-38],
       [4.46289110e-01, 4.21813888e-25, 1.45937653e-38],
       [4.89497690e-01, 6.97143576e-20, 2.80168131e-32],
       [9.86840168e-01, 1.51497396e-21, 3.69799777e-35],
       [5.58415223e-01, 3.32732990e-15, 9.97411287e-28],
       [5.24917069e-01, 2.38001221e-19, 7.31636239e-33]])
```

判別結果

```
a["predict"][0:10]
```

```
['setosa',
 'setosa',
 'setosa',
 'setosa',
 'setosa',
 'setosa',
 'setosa',
 'setosa',
 'setosa',
 'setosa']
```

判別結果表

```
a["correct table"]
```

prediction	setosa	versicolor	virginica
sp			
setosa	50	0	0
versicolor	0	47	3
virginica	0	0	50

正判別率

```
a["correct rate"]
```

98.0