

ROC 曲線と ROC 曲線下面積

青木繁伸

2020年3月17日

1 目的

原データまたは度数分布表データに基づいて、ROC 曲線を描く。また、ROC 曲線下面積を計算する。

2 使用法

疾病群，健康群の原データについて

```
import sys
sys.path.append("statlib")
from plot import ROC_curve0
ROC0(disease, normal, lowest=NULL, width=NULL, verbose=True)
```

度数分布表の形でまとめられているデータについて

```
import sys
sys.path.append("statlib")
from plot import ROC_curve
ROC(x, disease, normal, verbose=True)
```

2.1 引数

disease	疾病群の測定値ベクトル
normal	健康者群の測定値ベクトル
lowest	最も小さい値のキリのよい数値
width	階級幅（計算精度）のキリのよい数値
	lowest か width がデフォルト値なら，データから適当に決める
x	分割表の下限値のベクトル（例題参照）
verbose	必要最小限のプリント出力をする

2.2 戻り値の名前

"result"	結果表
"cindex"	ROC 曲線下面積

3 使用例

3.1 度数分布表を与える場合

```
x = [100, 220, 230, 240, 250, 260, 270, 280, 290, 300, 320, 340, 360,
     400]
disease = [3, 2, 1, 4, 7, 4, 16, 5, 3, 9, 10, 5, 10, 21]
normal = [25, 7, 19, 17, 7, 8, 7, 6, 2, 2, 0, 0, 0, 0]

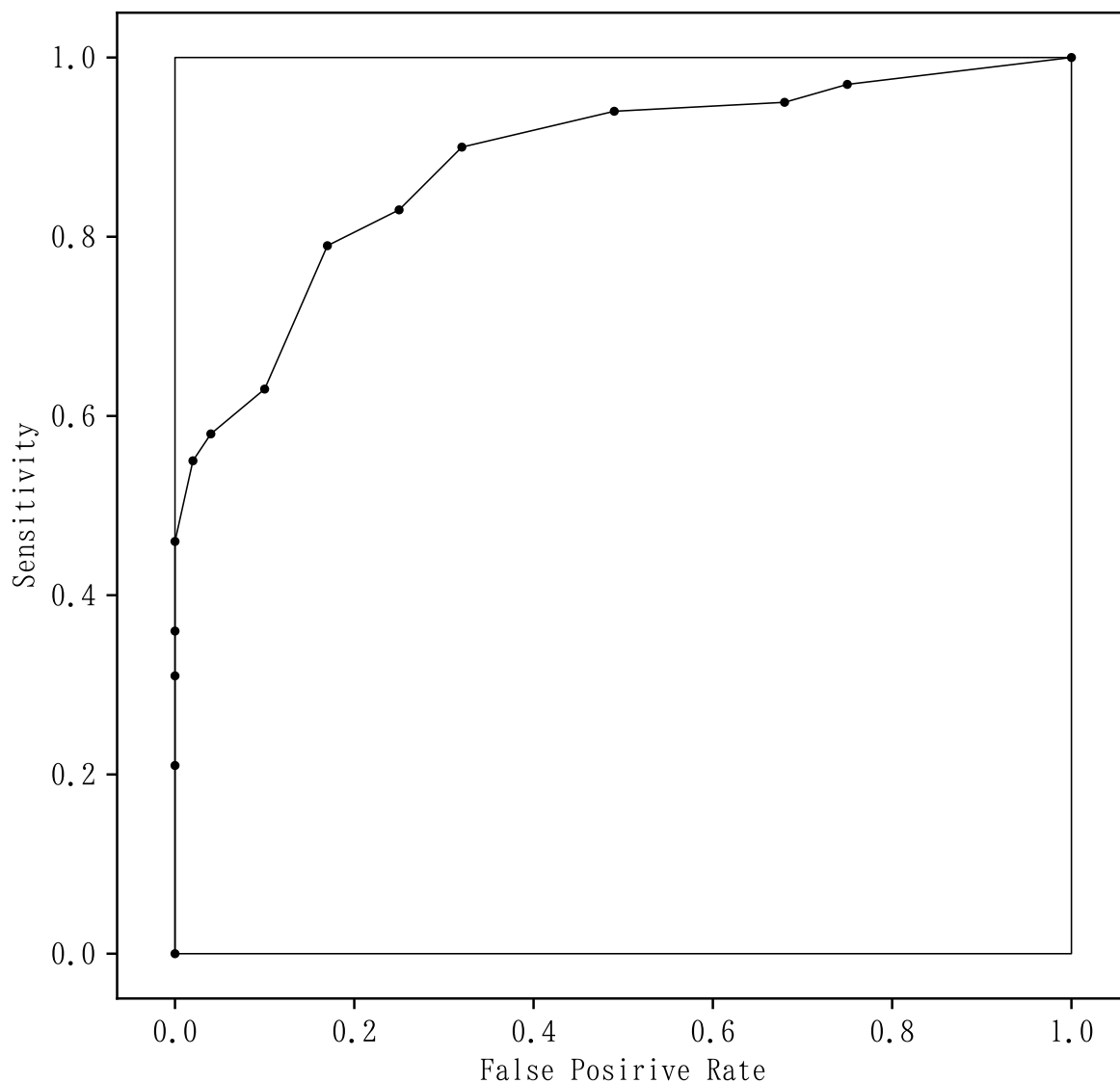
import sys
sys.path.append("statlib")
from plot import ROC_curve

a = ROC_curve(x, disease, normal)
```

ROC 曲線下面積

c index = 0.88215

	Value	Disease	Normal	Sensitivity	Specificity	F. P. rate
0	100	3	25	1.00	0.00	1.00
1	220	2	7	0.97	0.25	0.75
2	230	1	19	0.95	0.32	0.68
3	240	4	17	0.94	0.51	0.49
4	250	7	7	0.90	0.68	0.32
5	260	4	8	0.83	0.75	0.25
6	270	16	7	0.79	0.83	0.17
7	280	5	6	0.63	0.90	0.10
8	290	3	2	0.58	0.96	0.04
9	300	9	2	0.55	0.98	0.02
10	320	10	0	0.46	1.00	0.00
11	340	5	0	0.36	1.00	0.00
12	360	10	0	0.31	1.00	0.00
13	400	21	0	0.21	1.00	0.00



3.2 原データを与える場合

```
disease = [95, 95, 109, 103, 109, 107, 116, 108, 90, 114, 93, 134,
           123,
           105, 108, 91, 95, 109, 104, 109, 132, 120, 123, 128, 104, 98,
           111, 100, 128, 95, 121, 100, 98, 113, 90, 108, 109, 108, 115,
           113, 102, 121, 121, 94, 103, 104, 110, 110, 116, 89, 110, 125,
           116, 123, 111, 118, 117, 114, 108, 110, 109, 106, 100, 94, 114,
           110, 113, 89, 113, 111, 115, 107, 118, 94, 112, 120, 99, 113]

normal = [93, 91, 92, 100, 85, 83, 83, 100, 88, 97, 108, 103, 94, 76,
          92, 89, 99, 87, 87, 118, 96, 100, 97, 100, 90, 112, 100, 106,
          113, 100, 100, 93, 94, 105, 109, 92, 92, 100, 101, 93, 112, 89,
          85, 91, 101, 104, 109, 95, 85, 112, 93, 115, 95, 104, 109, 104,
          103, 106, 93, 103, 108, 120, 112, 95, 95, 103, 87, 100, 99, 88,
```

```
123, 90, 100, 82, 91, 79, 104, 105, 101, 97, 104, 91, 118, 102,
93, 108, 94, 106, 100, 89, 98, 92, 119, 107, 109, 83, 111, 98]
```

```
import sys
sys.path.append("statlib")
from plot import ROC_curve0

a = ROC_curve0(disease, normal)
```

最小値 x = 76

最大値 x = 134

最小値より小さいキリのよい数値 = 76

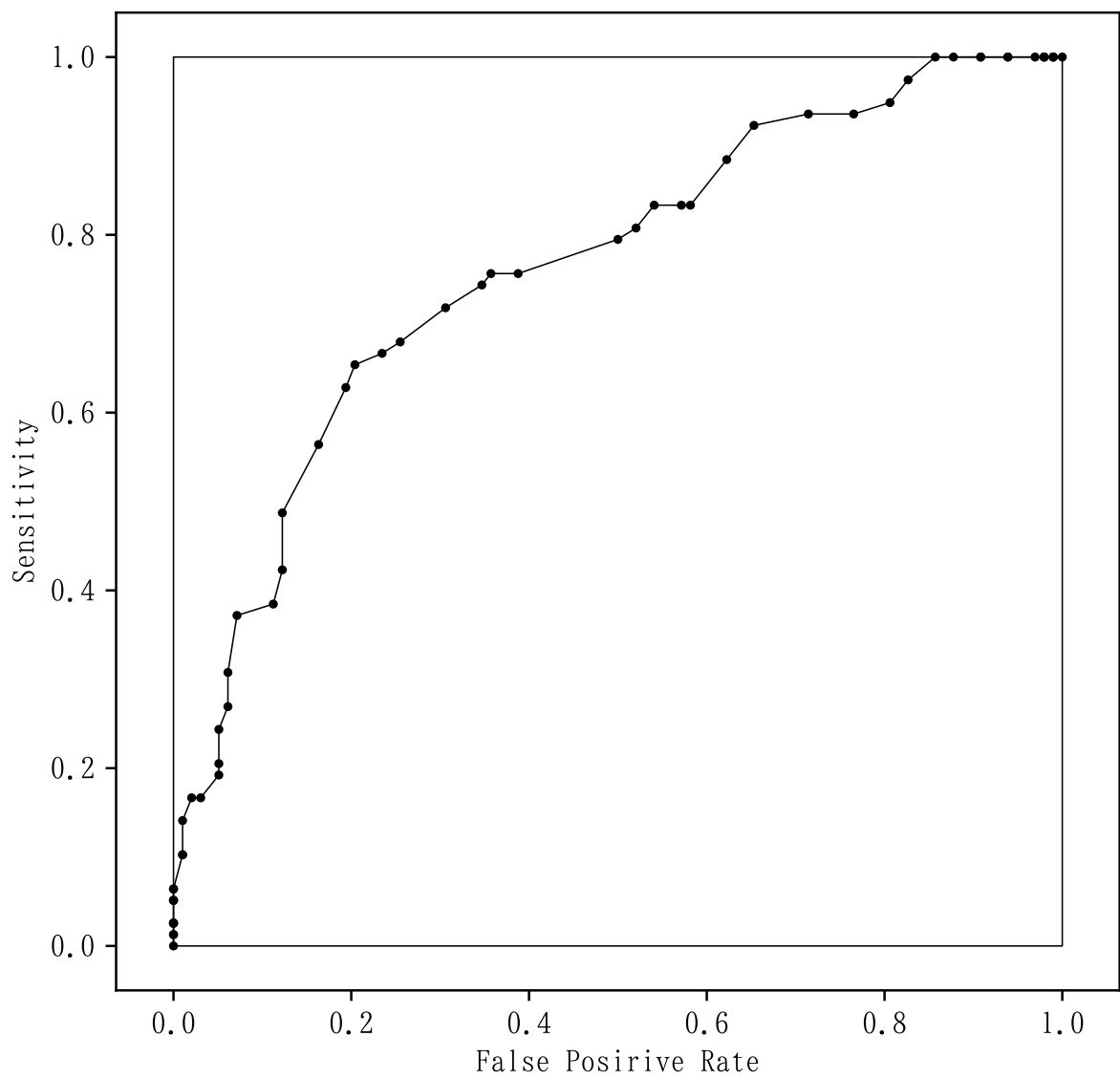
度数分布を作成する階級幅の切りのよい数値 = 1

ROC 曲線下面積

c index = 0.76262

	Value	Disease	Normal	Sensitivity	Specificity	F. P. rate
0	76.0	0.0	1.0	1.000000	0.000000	1.000000
1	77.0	0.0	0.0	1.000000	0.010204	0.989796
2	78.0	0.0	0.0	1.000000	0.010204	0.989796
3	79.0	0.0	1.0	1.000000	0.010204	0.989796
4	80.0	0.0	0.0	1.000000	0.020408	0.979592
5	81.0	0.0	0.0	1.000000	0.020408	0.979592
6	82.0	0.0	1.0	1.000000	0.020408	0.979592
7	83.0	0.0	3.0	1.000000	0.030612	0.969388
8	84.0	0.0	0.0	1.000000	0.061224	0.938776
9	85.0	0.0	3.0	1.000000	0.061224	0.938776
10	86.0	0.0	0.0	1.000000	0.091837	0.908163
11	87.0	0.0	3.0	1.000000	0.091837	0.908163
12	88.0	0.0	2.0	1.000000	0.122449	0.877551
13	89.0	2.0	3.0	1.000000	0.142857	0.857143
14	90.0	2.0	2.0	0.974359	0.173469	0.826531
15	91.0	1.0	4.0	0.948718	0.193878	0.806122
16	92.0	0.0	5.0	0.935897	0.234694	0.765306
17	93.0	1.0	6.0	0.935897	0.285714	0.714286
18	94.0	3.0	3.0	0.923077	0.346939	0.653061
19	95.0	4.0	4.0	0.884615	0.377551	0.622449
20	96.0	0.0	1.0	0.833333	0.418367	0.581633
21	97.0	0.0	3.0	0.833333	0.428571	0.571429
22	98.0	2.0	2.0	0.833333	0.459184	0.540816
23	99.0	1.0	2.0	0.807692	0.479592	0.520408
24	100.0	3.0	11.0	0.794872	0.500000	0.500000
25	101.0	0.0	3.0	0.756410	0.612245	0.387755
26	102.0	1.0	1.0	0.756410	0.642857	0.357143

27	103.0	2.0	4.0	0.743590	0.653061	0.346939
28	104.0	3.0	5.0	0.717949	0.693878	0.306122
29	105.0	1.0	2.0	0.679487	0.744898	0.255102
30	106.0	1.0	3.0	0.666667	0.765306	0.234694
31	107.0	2.0	1.0	0.653846	0.795918	0.204082
32	108.0	5.0	3.0	0.628205	0.806122	0.193878
33	109.0	6.0	4.0	0.564103	0.836735	0.163265
34	110.0	5.0	0.0	0.487179	0.877551	0.122449
35	111.0	3.0	1.0	0.423077	0.877551	0.122449
36	112.0	1.0	4.0	0.384615	0.887755	0.112245
37	113.0	5.0	1.0	0.371795	0.928571	0.071429
38	114.0	3.0	0.0	0.307692	0.938776	0.061224
39	115.0	2.0	1.0	0.269231	0.938776	0.061224
40	116.0	3.0	0.0	0.243590	0.948980	0.051020
41	117.0	1.0	0.0	0.205128	0.948980	0.051020
42	118.0	2.0	2.0	0.192308	0.948980	0.051020
43	119.0	0.0	1.0	0.166667	0.969388	0.030612
44	120.0	2.0	1.0	0.166667	0.979592	0.020408
45	121.0	3.0	0.0	0.141026	0.989796	0.010204
46	122.0	0.0	0.0	0.102564	0.989796	0.010204
47	123.0	3.0	1.0	0.102564	0.989796	0.010204
48	124.0	0.0	0.0	0.064103	1.000000	0.000000
49	125.0	1.0	0.0	0.064103	1.000000	0.000000
50	126.0	0.0	0.0	0.051282	1.000000	0.000000
51	127.0	0.0	0.0	0.051282	1.000000	0.000000
52	128.0	2.0	0.0	0.051282	1.000000	0.000000
53	129.0	0.0	0.0	0.025641	1.000000	0.000000
54	130.0	0.0	0.0	0.025641	1.000000	0.000000
55	131.0	0.0	0.0	0.025641	1.000000	0.000000
56	132.0	1.0	0.0	0.025641	1.000000	0.000000
57	133.0	0.0	0.0	0.012821	1.000000	0.000000
58	134.0	1.0	0.0	0.012821	1.000000	0.000000



```
a = ROC_curve0(disease, normal, lowest=75, width=5)
```

最小値 x = 76

最大值 x = 134

ROC 曲線下面積

c index = 0.75687

	Value	Disease	Normal	Sensitivity	Specificity	F. P. rate
0	75	0.0	2.0	1.000000	0.000000	1.000000
1	80	0.0	4.0	1.000000	0.020408	0.979592
2	85	2.0	11.0	1.000000	0.061224	0.938776
3	90	7.0	20.0	0.974359	0.173469	0.826531
4	95	7.0	12.0	0.884615	0.377551	0.622449
5	100	9.0	24.0	0.794872	0.500000	0.500000
6	105	15.0	13.0	0.679487	0.744898	0.255102

7	110	17.0	6.0	0.487179	0.877551	0.122449
8	115	8.0	4.0	0.269231	0.938776	0.061224
9	120	8.0	2.0	0.166667	0.979592	0.020408
10	125	3.0	0.0	0.064103	1.000000	0.000000
11	130	2.0	0.0	0.025641	1.000000	0.000000
12	135	0.0	0.0	0.000000	1.000000	0.000000

