医学統計学演習：資料 11

芳賀昭弘 *

1　ROC 解析と推論の評価

「健康・病気」などのような２つの値（二値）をとる事象を考えます。そしてこれを検診の「白血球数」から予測することを考えてください。白血球数がある基準範囲内であれば健康、外れるようであれば病気と予測するとしましょう。検査結果で病気が疑われるケースを**陽性**、疑われないケースを**陰性**と言います。この陽性と陰性を分ける点（今の例では白血球の数[*1]）を**カットオフ値**（**閾値**）と呼ばれます。この値を極端にとれば、被検者の多くを陽性とすることもできますし、逆にその多くを陰性とすることもできます。被検者の多くを陽性とするようなカットオフ値では、実際に病気でない人を病気と判定することで（これを**偽陽性**という[*2]）さらなる検査を要請することになり資源の無駄や被検者の負担を増やしてしまいます。逆に陰性に偏るようなカットオフ値では、偽陽性は少なくなりますが、本当は病気なのに誤って健康と判定する（これを**偽陰性**という[*3]）ことを避けることができなくなります。適切なカットオフ値を決めることは重要であることがわかります。また、「白血球数」に加えて「血小板数」も病気の予測に用いた場合、予測の精度は増すでしょうか？このように二値推論の複数のモデルを比較することでより良い予測モデルを定量的に評価したいことがあります。カットオフ値による予測性能の変化の振る舞いや推論モデルの比較に使われるのが Receiver Operating Characteristic (ROC) curve（ROC 曲線）です。実際起こった現象と予測の結果を比較する場合、表 2 の 2×2 分割表（**混同行列**）を使うとわかりやすくなります。ROC 曲線は、閾値を変えながら混同行列を作成し、偽陽性率を横軸に陽性率をプロットすると描くことができます。まずは表 2 に示されている混同行列の形と指標の定義を頭に入れましょう[*4]。

1.1　ROC 曲線の描画

横軸を偽陽性率、縦軸を陽性率として閾値を変えてプロットしたものが ROC 曲線です。それでは今回も例題を通して ROC 曲線の描画にチャレンジしてみましょう。
表 3 の「模試（入試 3 ヶ月前）の点数 vs. 入試結果」　と「模試（入試 1 ヶ月前）の点数 vs. 入試結果」 を使って２つの ROC 曲線を描いてください。

解答を見ておきましょう。図 1 の赤線が入試 3 ヶ月前、青線が入試 1 ヶ月前の点数に基づいた ROC 曲線です。これを描くために、この図の右側に示してある表のように、模試の点数に閾値を設定し、その閾値によって真の合格者のうち何名を検知できたか（真陽性率=感度）と真の不合格者のうち何名を誤って合格者として

* Electronic address: haga@tokushima-u.ac.jp

*1 もう少し詳しく述べると、検査による白血球数を x、日本人成年男性の中央値を m として $|x-m|<c$ を基準範囲とすると、この c が日本人成年男性のカットオフ値となります

*2 第一種の過誤

*3 第二種の過誤

*4 表 2 の「実際」と「予測」が入れ替わった形で書かれている本もあります。その場合偽陽性と偽陰性の位置が入れ替わりますので注意してください。

表 2：混同行列

		実際	
		真	偽
予測	真	真陽性	偽陽性
	偽	偽陰性	真陰性

- **感度** Sensitivity：真陽性/(真陽性 + 偽陰性)（実際が真の中で、予測が真の割合）
- **特異度** Specificity：真陰性/(偽陽性 + 真陰性)（実際が偽の中で、予測が偽の割合）
- **陽性的中率** Positive Predictive Value：真陽性/(真陽性 + 偽陽性)（予測が真のうち、実際に真の割合）
- **陰性的中率** Negative Predictive Value：真陰性/(偽陰性 + 真陰性)（予測が偽のうち、実際に偽の割合）
- **オッズ比** Odds Ratio： (真陽性/偽陰性)/(偽陽性/真陰性)
- **相対危険度** Risk Ratio：真陽性/(真陽性 + 偽陽性) / 偽陰性/(偽陰性 + 真陰性)

しまうか（偽陽性率=1-特異度）を算出します。あとは縦軸の真陽性率を横軸の偽陽性率に対してプロットすることでここで示された ROC 曲線を得ることができます。

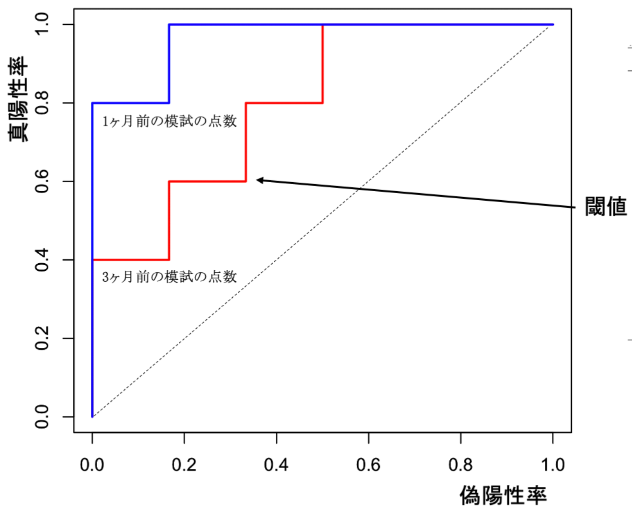
表 3：模試の点数による実際の試験の合格者の予測

模試（入試 3 ヶ月前）の点数	入試結果	模試（入試 1 ヶ月前）の点数	入試結果
100	合格	100	合格
90	合格	90	合格
80	不合格	80	合格
70	合格	70	合格
60	不合格	60	不合格
50	合格	50	合格
40	不合格	40	不合格
30	合格	30	不合格
20	不合格	20	不合格
10	不合格	10	不合格
0	不合格	0	不合格

模試の点数によって実際の試験の合格者はどのくらい正確に予測できたのか、ということをこの ROC 曲線から評価してみましょう。まず、閾値を幾つに設定すると合格・不合格をよりよく予測できるでしょうか？これに対しては、Balanced Error Rate (BER) という指標が時々使われます。予測の誤りは偽陽性率と偽陰性率の２通りがありますが、BER はその平均（(偽陽性率 + 偽陰性率）/2）で与えられます。これは**誤分類率**を意味しており、BER を最小にする閾値が間違いの少ない閾値であるという考え方に則っています。[*5]

ROC 曲線を積分したもの（曲線の下の領域の面積）を **Area Under the Curve (AUC)** といい、モデ

[*5] 実際に合格した人を不合格と予測すること（偽陰性）と、実際に不合格だった人を合格と予測すること（偽陽性）ではどちらの間違いが深刻でしょう？これは今回は偽陽性の方が罪深いと言えそうです。これは合格を陽性と定義したからであり、もし不合格を陽性とした場合は偽陰性がより深刻な間違いであるということもできます。何れにしても、間違い方には２通り（第一種の過誤と第二種の過誤）あって、どちらかがより深刻な間違いであることが多いです。これをもっと実感してもらうために “がん”ではない人を検査で “がん”と判定してしまうこと（偽陽性）と、“がん”である人を検査で “がん”ではないと判定してしまうこと（偽陰性）で考えてみてください。偽陽性は、より詳細な検査に移ってから間違いに気づくことになるので（少ないに越したことはないですが）まだ良いです。偽陰性の方は、そのまま放置されることで早期治療のチャンスを逃すため、生死に影響を及ぼすほど深刻です。このような場合には閾値の設定には BER ではなく、偽陽性率と偽陰性率に重み（例えば偽陽性率:偽陰性率=1:1000 など）を掛けて加える重み付き平均をとる損失関数を用意する手法がよいと考えられます。



模試の点数	合格 or 不合格	真陽性率*	偽陽性率**	偽陰性率	真陰性率
		0/5	0/6	5/5	6/6
100	合格	1/5	0/6	4/5	6/6
90	合格	2/5	0/6	3/5	6/6
80	不合格	2/5	1/6	3/5	5/6
70	合格	3/5	1/6	2/5	5/6
60	不合格	3/5	2/6	2/5	4/6
50	合格	4/5	2/6	1/5	4/6
40	不合格	4/5	3/6	1/5	3/6
30	合格	5/5	3/6	0/5	3/6
20	不合格	5/5	4/6	0/5	2/6
10	不合格	5/5	5/6	0/5	1/6
0	不合格	5/5	6/6	0/5	0/6

* 陽性率：模試の点数に閾値を設定したとき、真の合格者（今の例では5名）のうち何名を検知できたか

** 偽陽性率：模試の点数に閾値を設定したとき、真の不合格者（今の例では6名）のうち何名を誤って合格者としてしまうか

図１　表３のデータから得られる ROC 曲線（左）とその描画方法（右）

ルの予測性能を評価を定量するときに便利です。今の例では、模試（入試３ヶ月前）の点数と模試（入試１ヶ月前）の点数が予測モデルであり、AUC はそれぞれ 0.8 と 0.966･･･ です。AUC は最大が１であり、１に近いほど予測性能が良いと言えることになります。今回の例では、模試（入試１ヶ月前）の点数で入試の合否を予測するモデルの方が優れていると言えます。

1.2　確信度

上記の例で、模試の点数が高い場合には実際の試験では合格する確率は高くなるだろうことが期待されると思います。したがって、この模試の点数が試験に合格する「確信度」を与えていると考えることができます。別の例で、肺野の X 線レントゲン画像からがんをスクリーニングする過程において、がんであるかどうかを人が見て推定する際に 10 段階で評価したとします。0 が正常で、10 はがんであると最も確信して推定しているとすると、この 10 段階評価が確信度を表していると言えます。ただし、がんと思っても正常だったり（偽陽性）、正常と思ったが実際にはがんであったり（偽陰性）と、間違えることはあるでしょう。実際に正常症例を集めて、それで評価した確信度の分布と、異常症例を集めて同じく確信度を分布かしたものが図２の左に示されています。あるところに閾値を設定し、その閾値以上ががん、それ未満が正常というようにすると、図のように真陰性率 TNR、偽陽性率 FPR、真陽性率 TPR、偽陰性率 FNR が図に示されている部分（積分領域）になることがわかります。ROC 曲線はこの２つの曲線を同時にプロットし、閾値を設定して作成されていることに気づくと思います（図２右）。また、互いに分布が離れていれば ROC 曲線の AUC は大きく、試験の合否やがんと正常の識別が容易であり、逆に分布が重なっていると、それらの識別が困難になってくることもわかります。図３には、正常症例と異常症例の分布の形状（特に標準偏差）の違いによる ROC 曲線の形状変化を示しています。これらの ROC 曲線の振る舞いは、左側の分布と閾値の関係から導くことができるので、ROC 曲線の作成過程をしっかりと身につけておきましょう。

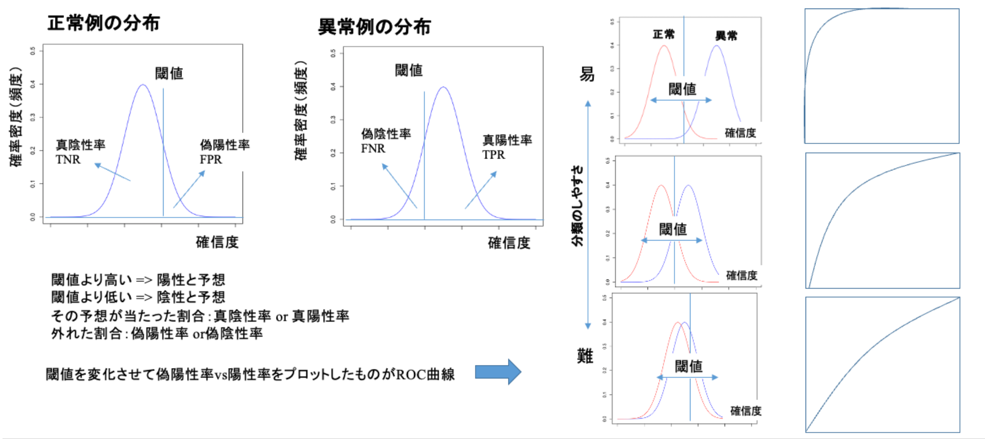


図 2　左：正常例、異常例における真陰性と偽陽性、真陽性と偽陰性の閾値との関係。右：正常例、異常例の分布の重なりと ROC 曲線との関係。

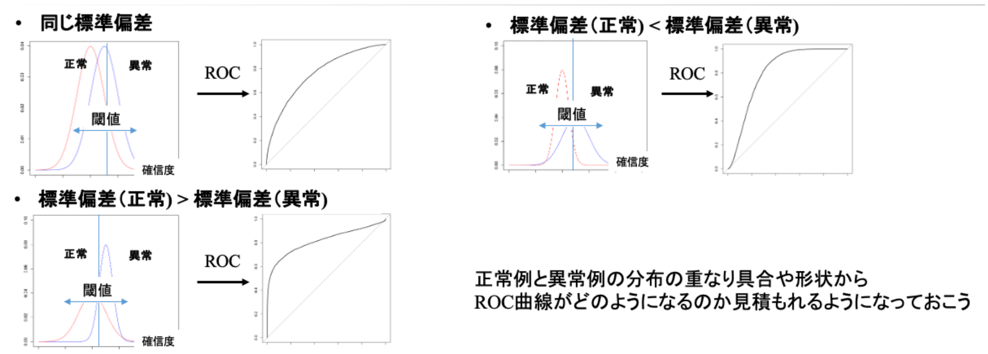


図 3　正常と異常の分布の違い（標準偏差の違い）による ROC 曲線の形状の変化。

1.3　pROC による ROC 解析

最後に pROC もしくは ROCR というライブラリを使った R による ROC 解析の例を示しておきます。pROC ライブラリのインストールは

```
> install.packages(“pROC”)
```

とします（既にインストールされているようであれば行う必要はありません)。pROC を使うには次のライブラリ関数で pROC ライブラリを読み込みます。

```
> library(pROC)
```

この library(pROC) は ROC 解析をする時に毎回宣言しなければならないことに注意しておいてください。それでは、Point に模試（入試 3 ヶ月前）の点数を、Class に合否（合格を 1, 不合格を 0)[*6]を登録し、ROCR ライブラリの関数 prediction や performance を使って ROC 曲線を書かせてみましょう。

[*6] ここで不合格を 1、合格を 0 とすると異なる結果が得られます。通常、予想に使う値が大きい → 陽性という対応で考えますが、今回は模試の点数が大きい → 陰性という対応で考えるため、本来の陽性（不合格）と陰性（合格）を反転させています。この例からもわかるように、入力するデータと結果に整合性があるかどうかを確かめながら解析を行う必要があります。これは ROC 解析だけでなく一般的な統計解析で言えることです。

```
> Point = c(0,10,20,30,40,50,60,70,80,90,100)
> Class = c(0,0,0,1,0,1,0,1,0,1,1)
> ROC1 = roc(Class, predictor=Point)
> plot(ROC1,col=”Red”)
```

図 1 の赤線を書かせることができたと思います。4 行目で ROC 解析を実施し、5 行目で縦軸を Sensitivity（tpr)、横軸を Specificity（fpr）としてプロットしています。AUC は以下で確認できます。

```
> ROC1
```

なお、2 つの ROC 曲線に差があるかどうかを検定する方法として delong 検定というものがあり、pROC では関数が用意されている。模試（入試 1 ヶ月前）の点数についての ROC 解析を実施し、それを ROC2 として保存しているとしよう。ROC1 と ROC2 の両者に差があるかどうかを検定するには、

```
> roc.test(ROC1, ROC2, method=“delong”, alternative=“two.sided”)
```

とすればよい。

1.4 ROCR による ROC 解析

ROCR ライブラリのインストールは

```
> install.packages(“ROCR”)
```

とします。インストールしたら、次のライブラリ関数で ROCR ライブラリを読み込みます。

```
> library(ROCR)
```

この library(ROCR) は ROC 解析をする時に初めに宣言しなければならないことに注意しておいてください。それでは、Point に模試（入試 3 ヶ月前）の点数を、Class に合否（合格を 1, 不合格を 0)[*7]を登録し、ROCR ライブラリの関数 prediction や performance を使って ROC 曲線を書かせてみましょう。

```
> Point = c(0,10,20,30,40,50,60,70,80,90,100)
> Class = c(0,0,0,1,0,1,0,1,0,1,1)
> pred = prediction(Point,Class)
> perf = performance(pred,”tpr”,”fpr”)
> plot(perf,col=”Red”,lwd = 2)
```

同じく図 1 の赤線を書かせることができたと思います。4 行目の”tpr”と”fpr” は真陽性率（true positive rate) と偽陽性率（false positive rate) のことです。5 行目で縦軸を tpr、横軸を fpr としてプロットしています。AUC は次のようにして計算することができます。

```
> auc.tmp = performance(pred,”auc”)
> as.numeric(auc.tmp@y.values)
```

ここでは模試の点数による合否の予測という例で ROC 曲線とその解析について説明しました。模試の点数をもっと複雑なモデルにすることで人工知能っぽくすることができます。例えば、模試の点数も 1 つの模試だけではなく過去に受けた模試の点数の時系列データや性別・年齢・居住地域・親の年収・学生時代の部活動等々を入力にして合格確率を出力するモデルを作ったとします。このモデルの性能は、モデルの合格確率と実際の合否で ROC 曲線を描かせることで模試の点数単独のモデルと同じように評価・比較することができ

[*7] ここで不合格を 1、合格を 0 とすると異なる結果が得られます。通常、予想に使う値が大きい → 陽性という対応で考えますが、今回は模試の点数が大きい → 陰性という対応で考えるため、本来の陽性（不合格）と陰性（合格）を反転させています。この例からもわかるように、入力するデータと結果に整合性があるかどうかを確かめながら解析を行う必要があります。これは ROC 解析だけでなく一般的な統計解析で言えることです。

ます。

2 演習

1. E_11_ROC.csv をダウンロードし、Histology に対す項目 Energy の ROC 解析を実施し、その AUC を求めよ。
2. 上記の解析において、Cutoff=0.5 及び BER を最小とした時の感度と特異度を求めよ。
3. 最も AUC の値が大きくなる項目を探査せよ（シェルスクリプトで for 文を使って一気に計算するスクリプトを作成し、manaba に貼り付けること)。