

医学統計学演習：資料 7

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に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??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は則っています。^{*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。

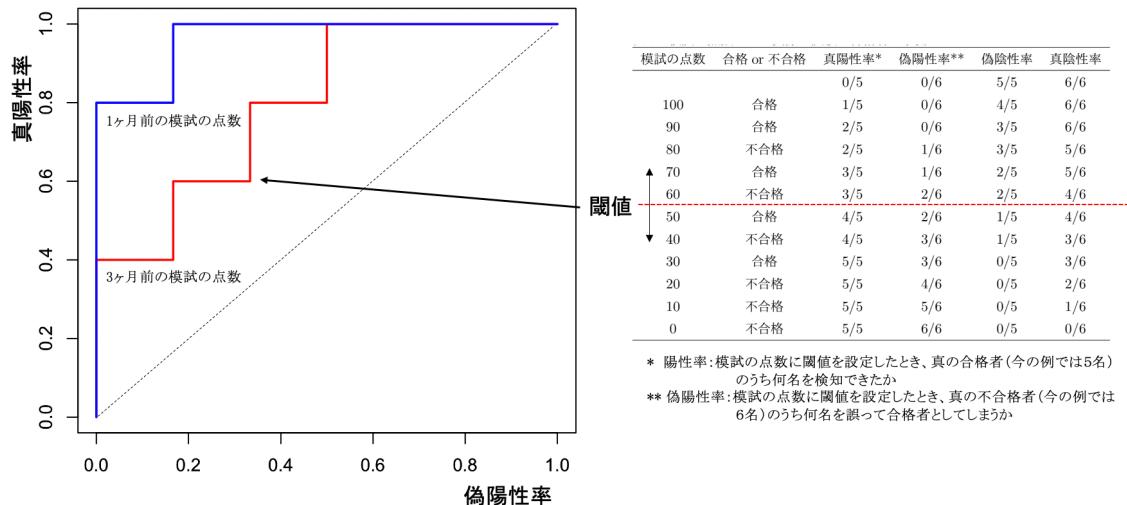


図1 表3のデータから得られる ROC 曲線（左）とその描画方法（右）

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。

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。

1.2 pROC による ROC 解析

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。

```
> 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??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
```

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。

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。

1.3 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??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ます。

2 演習

1. E_11_ROC.csvをダウンロードし、Histologyに対する項目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。

^{*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。

- 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）。