自由的百科全書

首頁

分類索引

特色内容

新聞動態

最近更改

隨機條目

維基社群

方針與指引

互助客棧

知識問答

聯絡我們

IRC即時聊天

關於維基百科

資助維基百科

在其他項目中

維基共享資源

打印/導出

下載為PDF

打印頁面

鏈入頁面

相關更改

上傳文件

特殊頁面

固定鏈接

頁面信息

引用本頁

其他語言

Deutsch

English

Español

Français

日本語

한국어

Português

Русский

❖A 還有8種語言

ቖ編輯鏈接

Türkçe

維基數據項

左側跳頂連接

工具

幫助

查看歷史 搜索维基百科

術語

陽性 (P, positive)

陰性 (N, Negative)

真陽性 (TP, true positive)

真陰性 (TN, true negative)

偽陽性 (FP, false positive)

偽陰性 (FN, false negative)

正確的肯定。又稱:命中 (hit)

真陽性率 (TPR, true positive rate)

TPR = TP / P = TP / (TP+FN)

偽陽性率(FPR, false positive rate)

FPR = FP / N = FP / (FP + TN)

即: (真陽性+真陰性) / 總樣本數

又稱:特異度 (SPC, specificity)

SPC = TN / N = TN / (FP + TN) = 1 - FPR

Matthews相關係數 (MCC),即 Phi相關係數

(TP*TN-FP*FN)

 $\sqrt{PNP'N'}$

ACC = (TP + TN) / (P + N)

準確度 (ACC, accuracy)

真陰性率 (TNR)

陽性預測值 (PPV)

陰性預測值 (NPV)

假髮現率 (FDR)

F1評分

PPV = TP / (TP + FP)

NPV = TN / (TN + FN)

FDR = FP / (FP + TP)

F1 = 2TP/(P+P')

Source: Fawcett (2006).

正確的否定。又稱:正確拒絕 (correct rejection)

錯誤的否定,又稱:未命中(miss),第二型錯誤

又稱:命中率 (hit rate)、敏感度(sensitivity)

又稱:錯誤命中率,假警報率 (false alarm rate)

錯誤的肯定,又稱:假警報 (false alarm),第一型錯誤

ROC曲線 [編輯]

維基百科,自由的百科全書

在信號檢測理論中,**接收者操作特徵曲線**(receiver operating characteristic curve,或者叫**ROC曲線**)是一種坐標圖式的分析工具,用於 (1) 選擇最佳的信號偵測模型、捨棄次佳的模型。 (2) 在同一模型中設定最 佳閾值。

中文維基百科條目協作計畫專頁已建立,歡迎報名參與!

在做決策時,ROC分析能不受成本/效益的影響,給出客觀中立的建議。

ROC曲線首先是由二戰中的電子工程師和雷達工程師發明的,用來偵測戰場上的敵軍載具(飛機、船艦),也就是信號檢測理論。之後很快就被引入了心理學來進行信號的知覺檢測。數十年來,ROC分析被用於醫 學、無線電、生物學、犯罪心理學領域中,而且最近在機器學習(machine learning)和數據挖掘(data mining)領域也得到了很好的發展。

目錄 [隐藏] 1 基本概念 2 ROC空間 3 ROC曲線 4 曲線下面積(AUC) 4.1 意義 4.2 計算 4.2.1 梯形法

4.2.2 ROC AUCH法 4.3 潛在問題

5 分析軟體 6 參考文獻 6.1 引用 6.2 來源

 NetChop C-term 3. — TAP + ProteaSMM ProteaSMM-i 3條ROC曲線

基本概念 [編輯]

7 外部鏈接

參見: 第一型及第二型錯誤和靈敏度和特異度

分類模型(又稱分類器,或診斷)是將一個實例映射到一個特定類的過程。ROC分析的是二元分類模型,也就是輸出結果只有兩種類別的模型,例如: (陽性/陰性) (有病/沒病) (垃圾郵件/非垃圾郵件) (敵軍/非敵軍)。

當**訊號偵測**(或**變數測量**)的結果是一個<mark>連續值</mark>時,類與類的邊界必須用一個閾值(英語:threshold)來界定。舉例來說,用血壓值來檢測一個人是否 有高血壓,測出的血壓值是連續的實數(從0~200都有可能),以收縮壓140/舒張壓90為閾值,閾值以上便診斷為有高血壓,閾值未滿者診斷為無高 血壓。二元分類模型的個案預測有四種結局:

1. **真陽性**(TP):**診斷為有,實際上也有**高血壓。 2. **偽陽性**(FP): 診斷為有,實際卻沒有高血壓。 3. **真陰性**(TN): 診斷為沒有,實際上也沒有高血壓。 4. **偽陰性**(FN): 診斷為沒有,實際卻有高血壓。

這四種結局可以畫成2 × 2的混淆矩陣:

		真寶	總	
		p	n	數
預測輸出	p'	真陽性 (TP)	偽陽性 (FP)	Ρ'
	n'	偽陰性 (FN)	真陰性 (TN)	N'
總數		P	N	

ROC空間 [編輯]

ROC空間將偽陽性率(FPR)定義為 X 軸,真陽性率(TPR)定義為 Y 軸。

• TPR: 在所有實際為陽性的樣本中, 被**正確地**判斷為陽性之比率。 TPR = TP/(TP + FN)• FPR: 在所有實際為陰性的樣本中,被錯誤地判斷為陽性之比率。

FPR = FP/(FP+TN)

給定一個二元分類**模型**和它的**閾值**,就能從所有樣本的(陽性 / 陰性)真實值和預測值計算出一個 (X=FPR, Y=TPR) 座標點。 從 (0, 0) 到 (1,1) 的對角線將ROC空間劃分為左上 / 右下兩個區域,在這條線的以上的點代表了一個好的分類結果(勝過隨機分類),而在這條線以下

的點代表了差的分類結果(劣於隨機分類)。 **完美的預測**是一個在左上角的點,在ROC空間座標 (0,1)點,X=0 代表著沒有偽陽性,Y=1 代表著沒有偽陰性(所有的陽性都是真陽性);也就是說, 不管分類器輸出結果是陽性或陰性,都是100%正確。─個**隨機的預測**會得到位於從 (0, 0) 到 (1, 1) **對角線** (也叫無識**別率線**) 上的─個點;最直觀的隨機預測的例子就是拋硬幣。

讓我們來看在實際有100個陽性和100個陰性的案例時,四種預測方法(可能是四種分類器,或是同一分類器的四種閾值設定)的結果差異:

	TP=63	FP=28	91	TP=77	FP=77	154	TP=24	FP=88	112	TP=76	FP=12	88
	FN=37	TN=72	109	FN=23	TN=23	46	FN=76	TN=12	88	FN=24	TN=88	112
	100	100	200	100	100	200	100	100	200	100	100	200
TPR = 0.63			TPR = 0.77			TPR = 0.24			TPR = 0.76			
FPR = 0.28		FPR = 0.77		FPR = 0.88		FPR = 0.12						
ACC = 0.68		ACC = 0.50		ACC = 0.18		ACC = 0.82						
將這4種結果畫在ROC空間裡:												

●點與隨機猜測線的距離,是預測力的指標:離左上角越近的點預測(診斷)準確率越高。離右下角越近的點,預測越不準。

• 在A、B、C三者當中,最好的結果是A方法。 •B方法的結果位於隨機猜測線(對角線)上,在例子中我們可以看到B的準確度(ACC,定義見前面表格)是50%。

•C雖然預測準確度最差,甚至劣於隨機分類,也就是低於0.5(低於對角線)。然而,當將C以(0.5, 0.5)為中點作一個鏡像後,C'的結果甚至要比A還要好。這個作鏡像的方法,簡單說,不管

C(或任何ROC點低於對角線的情況)預測了什麼,就做相反的結論。

ROC曲線 [編輯]

上述ROC空間裡的單點,是給定分類模型且給定閾值後得出的。但同一個二元分類模型的閾值可能設定為高或低,每種閾值的設定會得出不同的FPR和TPR。 •將**同一模型每個閾值**的 (FPR, TPR) 座標都畫在ROC空間裡,就成為**特定模型的ROC曲線**。

例如右圖,人體的血液蛋白濃度是呈正態分佈的連續變數,病人的分佈是紅色,平均值為A g/dL,健康人的分佈是藍色,平均值是C g/dL。健康檢查會測量血液樣本中的某種蛋白質濃度,達到某個 值(閾值,threshold)以上診斷為有疾病徵兆。研究者可以調整閾值的高低(將左上圖的B垂直線往左或右移動),便會得出不同的偽陽性率與真陽性率,總之即得出不同的預測準確率。 1. 由於每個不同的分類器(診斷工具、偵測工具)有各自的測量標準和測量值的單位(標示為:「健康人-病人分佈圖」的橫軸),所以不同分類器的「健康人-病人分佈圖」都長得不一樣。

2. 比較不同分類器時,ROC曲線的實際形狀,便視兩個實際分佈的重疊範圍而定,沒有規律可循。

3. 但**在同一個分類器之内**,閾值的不同設定對ROC曲線的影響,仍有一些規律可循:

•當閾值設定為最高時,亦即所有樣本都被預測為陰性,**沒有樣本被預測為陽性**,此時在偽陽性率 FPR = FP / (FP + TN) 算式中的 FP = 0,所以 FPR = 0%。同時在真陽性率(TPR)算式中, TPR = TP / (TP + FN) 算式中的 TP = 0, 所以 TPR = 0% → 當閾值設定為最高時,必得出ROC座標系左下角的點 (0, 0)。

•當閾值設定為最低時,亦即所有樣本都被預測為陽性,沒有樣本被預測為陰性,此時在偽陽性率FPR = FP / (FP + TN)算式中的 TN = 0,所以 FPR = 100%。同時在真陽性率 TPR = TP / (TP + FN) 算式中的 FN = 0, 所以 **TPR=100%**

→ 當閾值設定為最低時,必得出ROC座標系右上角的點 (1, 1)。

●因為TP、FP、TN、FN都是累積次數,TN和FN隨著閾值調低而減少(或持平),TP和FP隨著閾值調低而增加(或持平),所以FPR和TPR皆必隨著閾值調低而增加(或持平)。 → **隨著閾值調低,ROC點** 往右上(或右/或上)移動,或不動;但**絕不會往左下(或左/或下)移動**。

曲線下面積(AUC) [編輯]

在比較不同的分類模型時,可以將每個模型的ROC曲線都畫出來,比較曲線下面積做為模型優劣的指標。

意義 [編輯]

ROC曲線下方的面積(英語:Area under the Curve of ROC (AUC ROC)),其意義是: • 因為是在1x1的方格里求面積, AUC必在0~1之間。

• 假設閾值以上是陽性,以下是陰性;

ullet 若隨機抽取一個陽性樣本和一個陰性樣本,分類器**正確判斷**陽性樣本的值高於陰性樣本之**機率** $= AUC^{[1]}$ 。 • 簡單說:AUC值越大的分類器,正確率越高。

• AUC = 1, 是完美分類器,採用這個預測模型時,存在至少一個閾值能得出完美預測。絕大多數預測的場合,不存在完美分類器。 • 0.5 < AUC < 1,優於隨機猜測。這個分類器(模型)妥善設定閾值的話,能有預測價值。

從AUC判斷分類器(預測模型)優劣的標準:

● AUC = 0.5, 跟隨機猜測一樣(例: 丟銅板), 模型沒有預測價值。

• AUC < 0.5, 比隨機猜測還差; 但只要總是反預測而行, 就優於隨機猜測。

計算 [編輯]

AUC的計算有兩種方式,都是以逼近法求近似值。 梯形法 [編輯]

梯形法(英語:trapezoid method):簡單地將每個相鄰的點以直線連接,計算連線下方的總面積。因為每一線段下方都是一個梯形,所以叫梯形法。

•優點:簡單,所以常用。 ●缺點:傾向於低估AUC。

ROC AUCH法 [編輯]

潛在問題 [編輯]

AUC of ROC是機器學習的社群最常使用來比較不同模型優劣的方法^[2] 。然而近來這個做法開始受到質疑,因為有些機器學習的研究指出,AUC的雜訊太多,並且很常求不出可信又有效的AUC值(此時便不能保證AUC傳達本節開頭所述之意義),使得 AUC在模型比較時產生的問題比解釋的問題更多[3][4][5]。

所有常用於統計分析的軟體(例:SPSS、SAS、SYSTAT、S-Plus、ROCKIT、RscorePlus)都有依據不同閾值自動計算真陽性和偽陽性比率、並依此繪製ROC曲線的功能。

分析軟體 [編輯]

離散分類器(英語:discrete,或稱「間斷分類器」),如決策樹,產生的是離散的數值或者一個二元標籤。應用到實例中,這樣的分類器最後只會在ROC空間產生單一的點。而一些其他的分類器,如樸素貝葉斯分類器,邏輯回歸或者人工神經網絡,產生

的是實例屬於某一類的可能性,對於這些方法,一個閾值就決定了ROC空間中點的位置。舉例來說,如果可能值低於或者等於0.8這個閾值就將其認為是陽性的類,而其他的值被認為是陰性類。這樣就可以通過畫每一個閾值的ROC點來生成一個生成一條曲 線。MedCalc是較好的ROC曲線分析軟件。 參考文獻 [編輯]

引用 [編輯] 1. ^ Fawcett, Tom (2006); An introduction to ROC analysis,

Pattern Recognition Letters, 27, 861–874.

2. ^ Hanley, James A.; McNeil, Barbara J. A method of comparing the areas under receiver operating characteristic curves derived from the same cases函.

Radiology. 1983-09-01, **148** (3): 839–843 [2008-12-03]. PMID 6878708៤.

John; Bittner, Michael; and Dougherty, Edward R. (2010); Small-sample precision of ROC-related estimates, Bioinformatics 26 (6): 822–830

3. ^ Hanczar, Blaise; Hua, Jianping; Sima, Chao; Weinstein,

performance of predictive distribution models, Global Ecology and Biogeography, 17: 145–151 5. ^ Hand, David J. (2009); *Measuring classifier* performance: A coherent alternative to the area under the ROC curve, Machine Learning, 77: 103–123

4. ^ Lobo, Jorge M.; Jiménez-Valverde, Alberto; and Real,

Raimundo (2008), AUC: a misleading measure of the

• Zou, K.H., O'Malley, A.J., Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, 6;115(5):654–7. • X. H., Zhou. *Statistical Methods in Diagnostic Medicine*. Wiley & Sons. 2002. ISBN 9780471347729.

來源 [編輯]

• Lasko, T.A., J.G. Bhagwat, K.H. Zou and Ohno-Machado, L. (2005). The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics*, 38(5):404–415. • Balakrishnan, N., (1991) *Handbook of the Logistic Distribution*, Marcel Dekker, Inc., ISBN 978-0824785871.

• Gonen M., (2007) *Analyzing Receiver Operating Characteristic Curves Using SAS*, SAS Press, ISBN 978-1-59994-298-1.

• Green, W.H., (2003) *Econometric Analysis*, fifth edition, Prentice Hall, ISBN 0-13-066189-9. • Heagerty, P.J., Lumley, T., Pepe, M. S. (2000) Time-dependent ROC Curves for Censored Survival Data and a Diagnostic Marker *Biometrics*, **56**:337–344

• Hosmer, D.W. and Lemeshow, S., (2000) *Applied Logistic Regression*, 2nd ed., New York; Chichester, Wiley, ISBN 0-471-35632-8. • Brown, C.D., and Davis, H.T. (2006) Receiver operating characteristic curves and related decision measures: a tutorial, *Chemometrics and Intelligent Laboratory Systems*, 80:24–38

• Mason, S.J. and Graham, N.E. (2002) Areas beneath the relative operating levels (ROL) curves: Statistical significance and interpretation. Q.J.R. Meteorol. Soc., 128:2145–2166. • Pepe, M.S. (2003). *The statistical evaluation of medical tests for classification and prediction*. Oxford. ISBN 0198565828. • Carsten, S. Wesseling, S., Schink, T., and Jung, K. (2003) Comparison of Eight Computer Programs for Receiver-Operating Characteristic Analysis. Clinical Chemistry, 49:433–439

• Swets, J.A. (1995). Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers. Lawrence Erlbaum Associates. • Swets, J.A., Dawes, R., and Monahan, J. (2000) Better Decisions through Science. *Scientific American*, October, pages 82–87.

外部鏈接 [編輯]

 An introduction to ROC analysis A more thorough treatment of ROC curves and signal detection theory

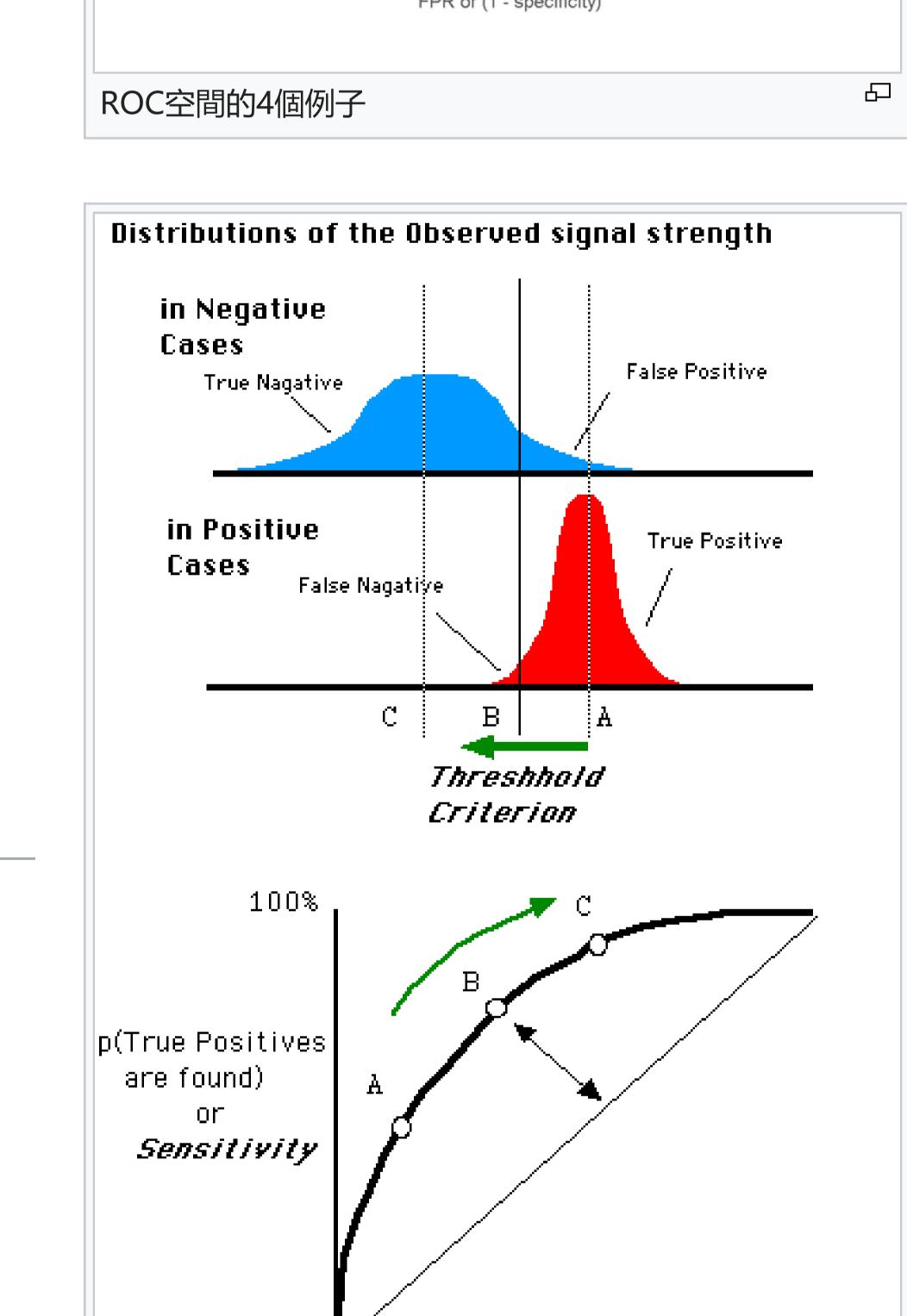
• Tom Fawcett's ROC Convex Hull: tutorial, program and papers 图[永久失效連結]

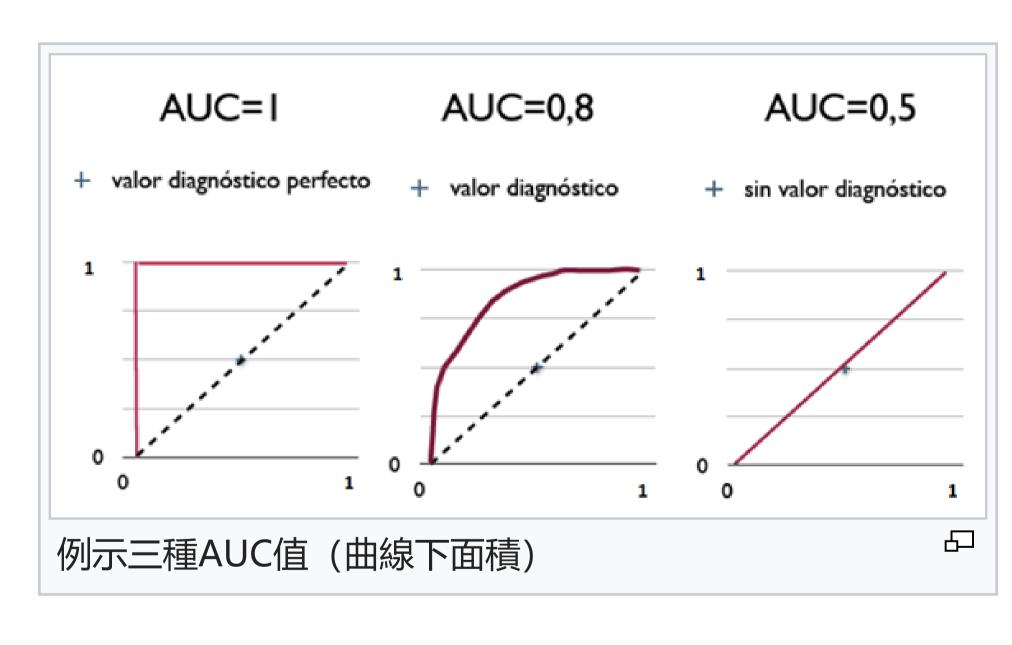
Peter Flach's tutorial on ROC analysis in machine learning

• The magnificent ROC - An explanation and interactive demonstration of the connection of ROCs to archetypal bi-normal test result plots 分類: 檢測論 數據挖掘 社會經濟學 生物統計學 統計分類

本頁面最後修訂於2019年3月14日 (星期四) 09:09。 本站的全部文字在知識共享署名-相同方式共享3.0協議之條款下提供,附加條款亦可能應用。(請參閱使用條款)

Wikipedia®和維基百科標誌是維基媒體基金會的註冊商標;維基™是維基媒體基金會的商標。 維基媒體基金會是按美國國內稅收法501(c)(3)登記的非營利慈善機構。 隱私政策 關於維基百科 免責聲明 開發者 Cookie聲明 手機版視圖





p(False Positives

are found)

100%