

ROC křivka

ROC (Receiver Operating Characteristic) křivka je nástroj pro hodnocení a optimalizaci binárního klasifikačního systému (testu), který ukazuje vztah mezi specificitou a senzitivitou daného testu nebo detektoru pro všechny přípustné hodnoty prahu.

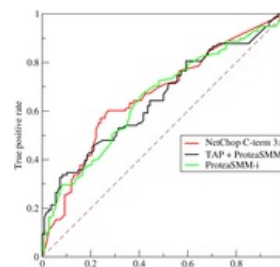
Křivka ROC byla poprvé použita během druhé světové války. Po útoku na Pearl Harbor v roce 1941 hledala armáda Spojených států nové metody zpracování radarových signálů umožňujících odlišit skutečné odrazy od letadel od šumu. ^[1]

Příklad

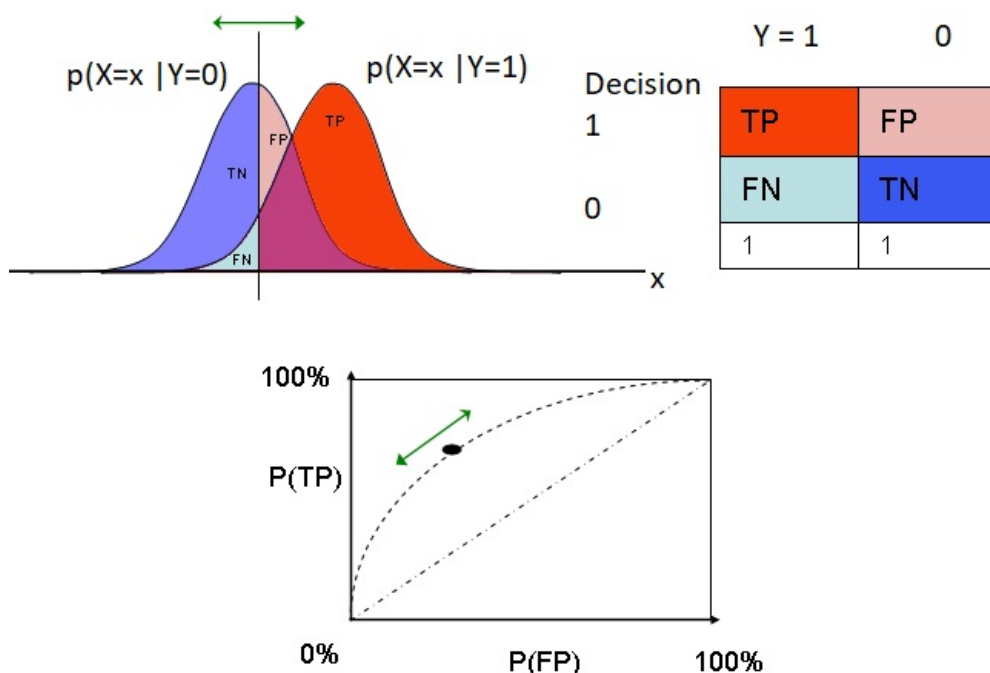
Ilustrujme použití ROC křivky na příkladu z obrázku uvedeného níže. Představme si, že v době chřipkové epidemie máme dvě skupiny obyvatel – zdravé a nakažené. Můžeme skupiny otestovat pomocí změření jejich tělesné teploty. Levá křivka představuje distribuci teplot u zdravé části sledovaných osob, pravá distribuci teplot u nakažených.

Jak se dalo předpokládat, teploty naměřené v obou subpopulacích se částečně překrývají. Separační kritérium (svíslá černá linka představující např. 38,6 °C) bude proto vždy část subpopulací klasifikovat špatně. Část pozorování bude falešně pozitivní (FP, růžová oblast), část falešně negativní (FN, světle modrá oblast). Jejich podíl se bude měnit v závislosti na volbě hodnoty separačního kritéria. Posuneme-li separační kritérium vlevo jak naznačuje zelená šipka, počet FN se zmenší, a počet FP se zvětší a opačně.

Tabulka vpravo ukazuje, že podíl skutečně pozitivních (TP = správně klasifikováno jako pozitivní) a falešně negativních (FN = chybně klasifikováno jako negativní) dá dohromady 1 nebo též 100 %, což je plocha pod pravou zvonovou křivkou. Stejně tak platí: Součet podílů skutečně negativních (TN) a podílů falešně pozitivních (FP) výsledků je 100 %, což odpovídá oblasti pod levou křivkou.



Ukázka ROC křivek tří testů



Ve spodní části obrázku je příslušná ROC křivka.

- Na svislé ose grafu je relativní četnost skutečně pozitivních případů TP, tedy pravděpodobnost, že jako správný bude vyhodnocen pozitivní případ
- Na vodorovné ose je relativní četnost falešně pozitivních případů FP, tedy pravděpodobnost, že jako správný bude vyhodnocen negativní případ

Bod na ROC křivce odpovídá hodnotě dělicího kritéria. Pokud bychom na obrázku nahoře posunuli dělicí kritérium doleva (směrem k nižším měřeným teplotám) pak mírně vzroste podíl TP a výrazně vzroste podíl FP. Bod na ROC křivce se posune poněkud doprava nahoru.

Pokud bychom šli do extrému a za dělicí kritérium vzali teplotu nižší, než má libovolná osoba za sledované skupiny, pak podíl TP bude 1, podíl FP rovněž 1 a bod ROC křivky bude v pravém horním rohu grafu.

Když naopak vezmeme za dělicí kritérium teplotu vyšší než má kdokoli ze sledovaných, podíl falešně pozitivních FP i skutečně pozitivních TP bude nula a ROC křivka protne levý dolní roh grafu.

Můžete si vyzkoušet ilustraci pohybu bodu na ROC křivce při změně prahové hodnoty testu: <oflash file="https://www.wikiskripta.eu/images/2/28/ROC.swf" width="550" height="350" />

Vztah k senzitivitě a specifitě

Vraťme se ještě jednou k obrázku distribučních křivek. Zkusme ve shodě s obrázkem vyjádřit veličiny na osách grafu ROC křivky. Víme, že na ose Y je relativní četnost skutečně pozitivních **TP**. Z obrázku distribučních křivek snadno odečteme, že tato relativní četnost se dá vyjádřit jako $P(TP) = TP / (TP + FN)$. Tento zlomek vyjadřuje jak velký je podíl skutečně pozitivních pozorování zachycených naším testem. Je to tedy **senzitivita** testu.

Zkusme nyní podle obrázku zapsat relativní četnost falešně pozitivních **FP**: $P(FP) = FP / (FP + TN)$.

Pro zjištění jak přesně náš test klasifikuje skutečně negativní případy (**jakou má specifitu**) by nás zajímala relativní četnost TN, kterou je možné vyjádřit zlomkem $TN / (TN + FP)$. Vidíme, že relativní četnost falešně pozitivních měření $P(FP)$, tedy údaj na X ové ose grafu ROC, je možné vyjádřit též pomocí specifity $P(FP) = 1 - \text{specifita}$.

Kvalita testu

Test, detektor, nebo diagnostická metoda jsou tím užitečnější, čím vyšší je jejich senzitivita a specifita. Zjednodušeně lze říci, že nastavováním různých prahových hodnot hledáme na ROC křivce kompromis mezi množstvím falešně pozitivních a falešně negativních pozorování, nebo jinými slovy mezi senzitivitou a specifitou. Ideální ROC křivka nejprve stoupá téměř svisle vzhůru (úspěšnost se blíží 100 %, zatímco míra chyb zpočátku zůstává blízko 0 %), teprve pak se zvyšuje míra falešné positivity. Naproti tomu, když ROC křivka stoupá po úhlopříčce, znamená to, že každé zlepšení senzitivity je zapláceno stejně významným zhoršením specifity a test není dobře navržen. Takový případ by mohl nastat např. pokud by se distribuce teplot pro obě subpopulace kryly (tj. infikovaní by neměli zvýšenou teplotu).

Na grafu ROC se bude nejlepší diagnostický test vyznačovat ROC křivkou s největší plochou pod křivkou (Area Under Curve – AUC). Je-li plocha rovná 1, je test ideální a má 100% senzitivitu i selektivitu. Pokud je plocha pod křivkou 0,5 pak není test lepší než házení mincí. V praxi bude diagnostický test někde mezi těmito extrémy.

Přibližné zhodnocení kvality testu podle plochy pod křivkou

- 0,50 až 0,75 = oprávněný
- 0,75 až 0,92 = dobrý
- 0,92 až 0,97 = velmi dobrý
- 0,97 až 1,00 = vynikající

Terminologie

- Kvalitu testu můžeme popsat pomocí senzitivity a specifity
- Senzitivita je poměr správně pozitivních pozorování ku všem pozitivním případům (jaký podíl skutečných případů test zachytí)
- Specifita je poměr správně negativních pozorování ku všem negativním případům (jaký podíl negativních případů test správně vyloučí)
- Prevalence – výskyt sledovaného jevu v populaci
- Prediktivní hodnota pozitivního testu je pravděpodobnost, že případ je skutečně pozitivní, když ho test jako pozitivní vyhodnotil
- Prediktivní hodnota negativního testu je pravděpodobnost, že případ je skutečně negativní, když jej test jako negativní vyhodnotil

Zatímco senzitivita a specifita jsou přímo vlastnosti testu, pozitivní a negativní prediktivní hodnota testu závisí na prevalenci (rozšíření v populaci).

Ilustrace k terminologii

Představme si, že provádíme např. test na okultní krvácení ve stolici (FOB) u 2030 osob ke zjištění chorobných změn v dolní části zažívacího traktu. Pak můžeme popsat možné stavy pomocí 2x2 kontingenční tabulky (Důležitá poznámka: v tomto příkladu budou v kontingenční tabulce přímo počty osob a pacientů, nikoli relativní četnosti jako výše):

| SKUTEČNÁ HODNOTA | | |
|--|--|---|
| (např. počet pacientů s rakovinou tlustého střeva diagnostikovanou endoskopicky) | | |
| | Pozitivní (p) | Negativní (n) |
| test pozitivní | Skutečně Pozitivní (TP) = 20 | Falešně Pozitivní (Chyba I typu) (FP) = 180 |
| | | → prediktivní hodnota pozitivního testu PPV = TP / (TP + FP) = 20 / (20 + 180) = 10 % |

ZMĚŘENÁ HODNOTA

(např. testem krve ve stolici)

test
negativní

| | Falešně Negativní (Chyba II typu) (FN) = 10 | Skutečně Negativní (TN) = 1820 |
|--|--|--|

→ prediktivní hodnota
negativního testu NPV
 $= TN / (FN + TN)$
 $= 1820 / (10 + 1820) \approx$
99,5 %

$$\begin{aligned} &P \\ &\downarrow \\ &\text{Senzitivita} \\ &= TP / (TP + FN) \\ &= 20 / (20 + 10) \\ &\approx 66,67 \% \end{aligned}$$

$$\begin{aligned} &N \\ &\downarrow \\ &\text{Specifická} \\ &= TN / (FP + TN) \\ &= 1820 / (10 + 1820) \\ &= 91 \% \end{aligned}$$

Nízká prediktivní hodnota pozitivního testu (PPV=10 %) nám ukazuje, že mnoho pozitivních výsledků testu bude falešně pozitivní (planý poplach). Bude tedy nutné všechny pozitivně vytestované případy vyšetřit spolehlivějším testem pro ověření přítomnosti rakoviny tlustého střeva. Přesto tento typ testu může být užitečný, pokud je levný a nenáročný.

Terminologie a vztahy mezi veličinami

- Skutečně Pozitivní (TP) – zjištěné případy výskytu znaku odpovídají skutečnosti
- Skutečně Negativní (TN) – zjištěné případy nepřítomnosti znaku odpovídají skutečnosti – správné zamítnutí
- Falešně Pozitivní (FP) – odpovídá falešnému poplachu – Chyba I typu
- Falešně Negativní (FN) – odpovídá přehlédnutému výskytu – Chyba II typu
- Senzitivita (SE) – neboli míra Skutečně Pozitivních P(TP), (True Positive Rate), odpovídá úspěšnosti detekce
 $P(TP) = TP/P = TP/(TP + FN)$
- Míra Falešně Pozitivních – P(FP), jinak též False positive rate (FPR) odpovídá
 $P(PR) = FP/N = FP/(FP + TN)$
- Přesnost (ACC) – $ACC = (TP + TN)/(P + N)$
- Specifická (SP) – neboli míra Skutečně Negativních (True Negative Rate)
 $SPC = TN/N = TN/(FP + TN) = 1 - FPR$
- Prevalence – výskyt sledovaného jevu v populaci lze vyjádřit $(TP + FN)/(TP + FN + FP + TN)$
- Prediktivní hodnota pozitivního testu (Positive predictive value PPV) – určuje pravděpodobnost, že jev je skutečně pozitivní, když test vyšel pozitivně. $PPV = TP/(TP + FP)$
- Prediktivní hodnota negativního testu (Negative predictive value NPV) – určuje pravděpodobnost, že jev je skutečně negativní, když test vyšel negativně. $NPV = TN/(FN + TN)$

Pokud jsou prevalence, senzitivita a specifická známe, můžeme prediktivní hodnoty vypočítat podle následujících rovnic (Bayesův teorém)

$$PPV = \frac{TN}{(FN + TN)} = \frac{(\text{senzitivita}) * (\text{prevalence})}{(\text{senzitivita}) * (\text{prevalence}) + (1 - \text{specifická}) * (1 - \text{prevalence})}$$

$$NPV = \frac{TP}{(TP + FP)} = \frac{(\text{specifická}) * (1 - \text{prevalence})}{(\text{specifická}) * (1 - \text{prevalence}) + (1 - \text{senzitivita}) * (\text{prevalence})}$$

Ekvivalenci obou vyjádření si můžete ověřit dosazením.

A zároveň závilost prevalence zhodnotit i v přenosti testu

$$ACC = \frac{(\text{senzitivita}) * (\text{prevalence})}{(\text{specifická}) * (1 - \text{prevalence})}$$

Historie

ROC křivka (Receiver Operating Characteristic – "provozní charakteristika přijímače") byla zavedena pro potřeby hodnocení kvality radaru za II. světové války. Po útoku na Pearl Harbor v roce 1941 začala armáda Spojených států intenzivně zkoumat, jak z radarových signálů přesněji vyhodnotit skutečná japonská letadla a odlišit je od falešných poplachů.

Odkazy

Související články

- Senzitivita testu
- Specifická testu

Reference

1. Green, David M.; Swets, John A. (1966). Signal detection theory and psychophysics. New York, NY: John Wiley and Sons Inc.. ISBN 0-471-32420-5

Použitá literatura

- WIKIPEDIA. *Receiver operating characteristic* [online]. ©2003. Poslední revize 2011-07-06, [cit. 2011-07-10]. <https://en.wikipedia.org/wiki/ROC_curve>.
- WIKIPEDIA. *Positive predictive value* [online]. ©2005. Poslední revize 2011-04-02, [cit. 2011-07-15]. <https://en.wikipedia.org/wiki/Positive_predictive_value>.
- ČVUT FEL. *ROC křivka* [online]. ©2006. Poslední revize 2006-03-10, [cit. 2011-07-10]. <<http://cyber.felk.cvut.cz/gerstner/teaching/zbd/biostat3.pdf>>.
- www.anaesthetist.com. *ROC* [online]. ©2001-09-21. Poslední revize 2011-07-28, [cit. 2012-07-23]. <<http://www.anaesthetist.com/mnm/stats/roc/Findex.htm>>.

Externí odkazy

- ZVÁROVÁ, Jana. *Základy statistiky pro biomedicínské obory*. 1. vydání. Praha : Karolinum, 1998. ISBN 80-7184-786-0.
- HOPLEY, Lara a Jo VAN SCHALKWYK. *The magnificent ROC* [online]. ©2001. Poslední revize 2007/03/25, [cit. 2011-07-11]. <<http://www.anaesthetist.com/mnm/stats/roc/Findex.htm>>.