The ROC Diagonal is not Layperson's Chance: a New Baseline Shows the Useful Area

Carrington AM, Fieguth PW, Mayr F, James N, Holzinger A, Pickering JW, Aviv RI

IFIP Cross-Domain Conference for Machine Learning and Knowledge Extraction, CD-MAKE, August 23-26 2022 (revised Sept. 3)

Why should you care?

New understandings about:

- evaluating/selecting binary classifiers
- optimizing decision threshold
- explaining models
- optimizing models

(can improve individual outcomes)

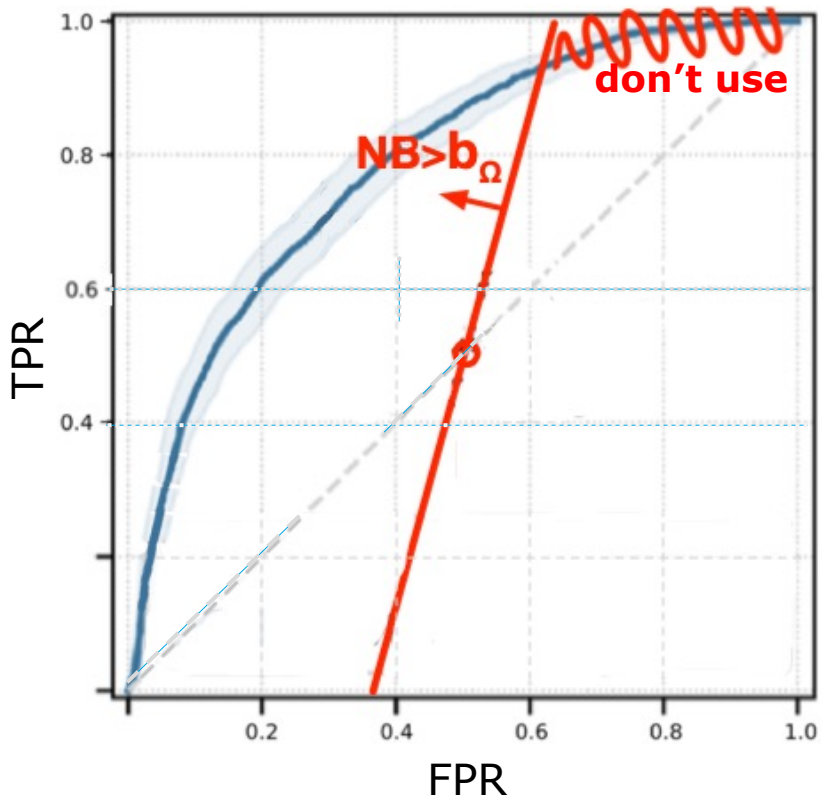
Key points

1. The ROC main diagonal is not chance, it is zero information. information ≠ useful information

2. We define the binary chance baseline.

3. We show the useful parts of: area under the ROC curve.

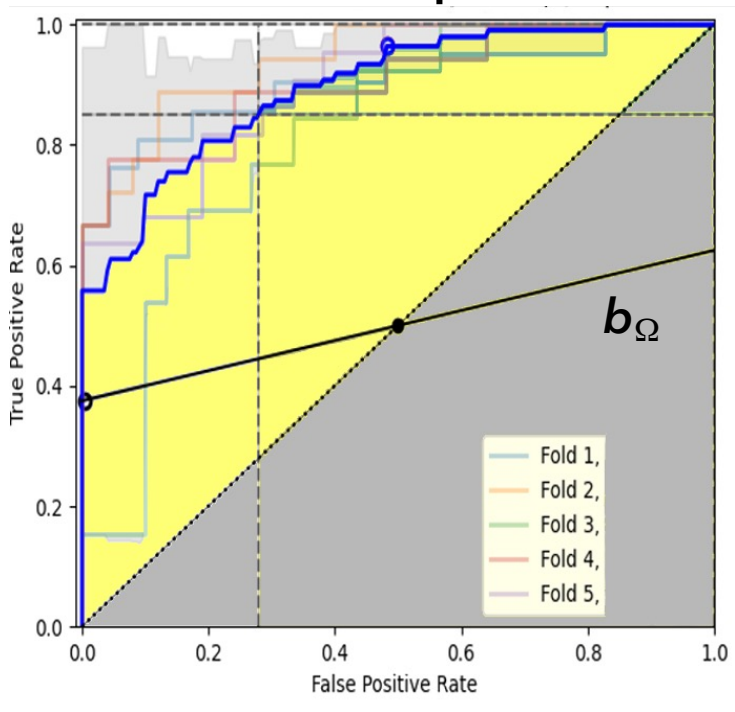
There are thresholds, or ROC points, we should not use, per our baseline b_Ω



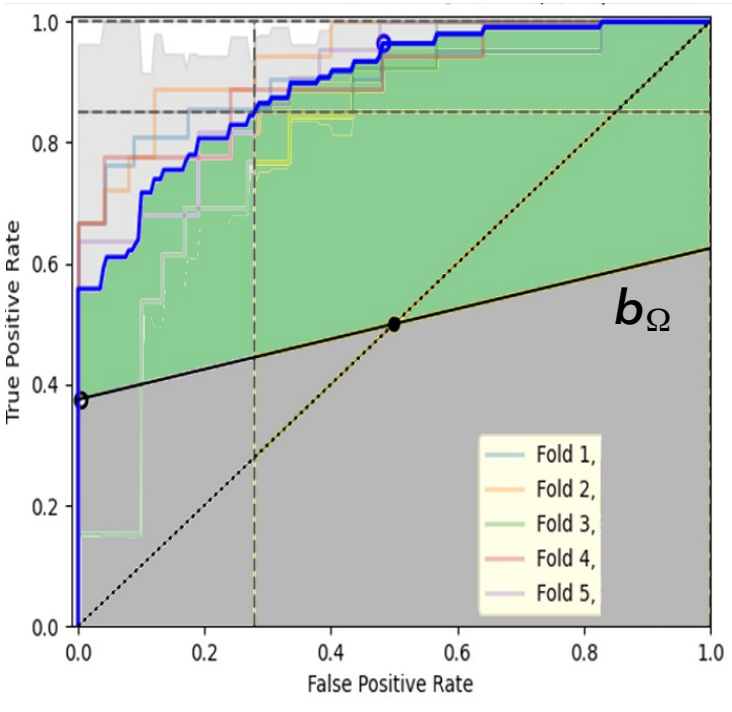
Annotations added to an ROC plot [1].

The useful area is not what the main diagonal (status quo) suggests.

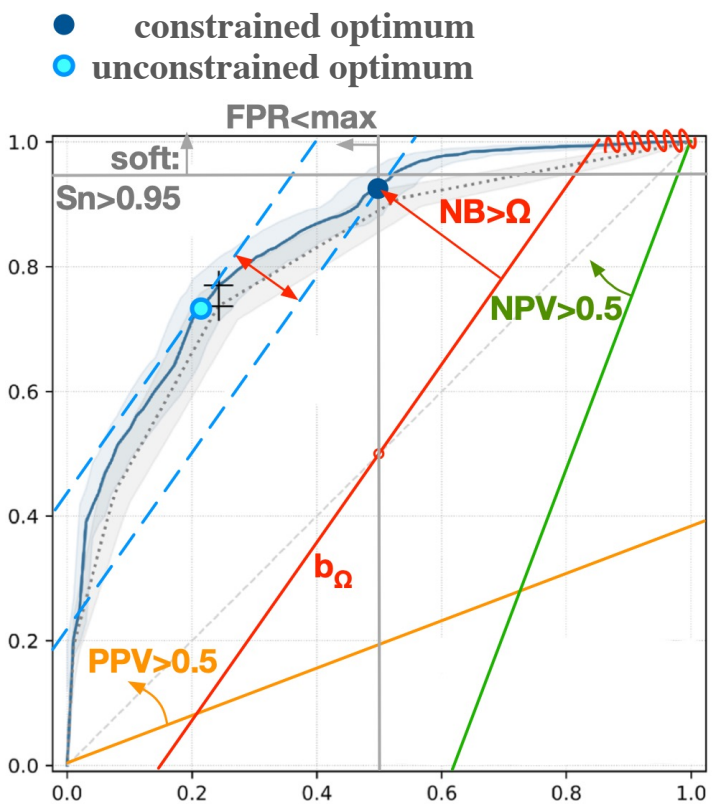
status quo

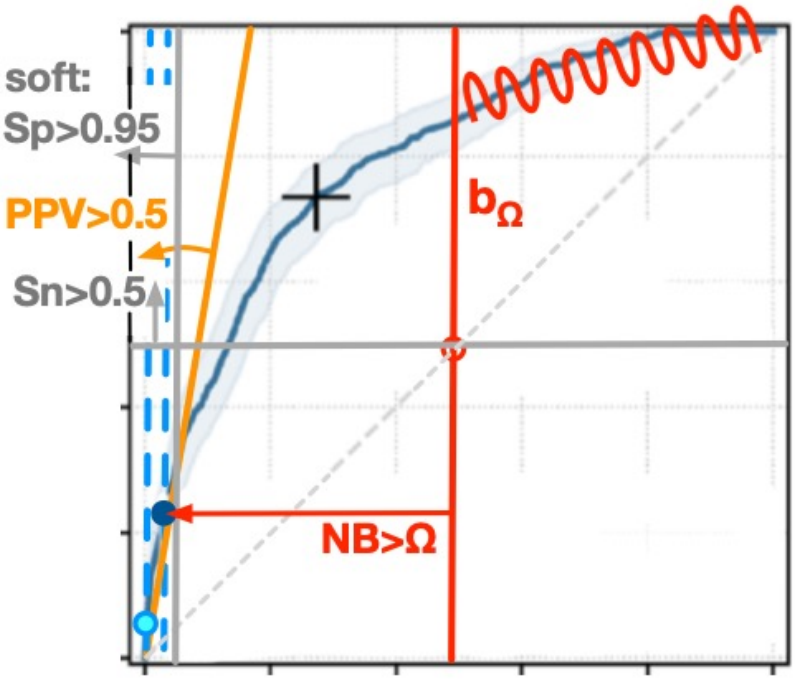


useful area



Our baseline b_Ω helps visualize and explain optimal ROC points





Why does chance matter in ROC plots?

Binary classifiers or tests, **begin to be useful** when they perform **better than chance** in:

accuracy, or

average net benefit

For binary outcomes

What is chance?

What is random?

Three statements in the literature are misleading or myths

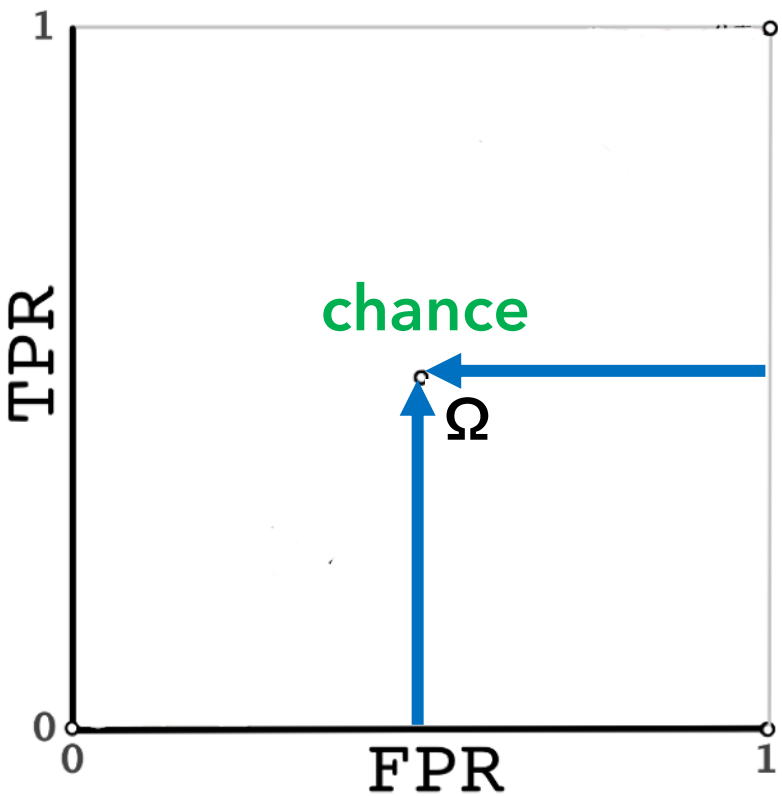
1. "The [ROC] main diagonal represents chance".

2. "An AUC of 0.5 represents a model which is no better than random chance at predicting a specific outcome".

3. ROC points above the main diagonal are useful.

Myth #1a

"The ROC main diagonal represents chance".

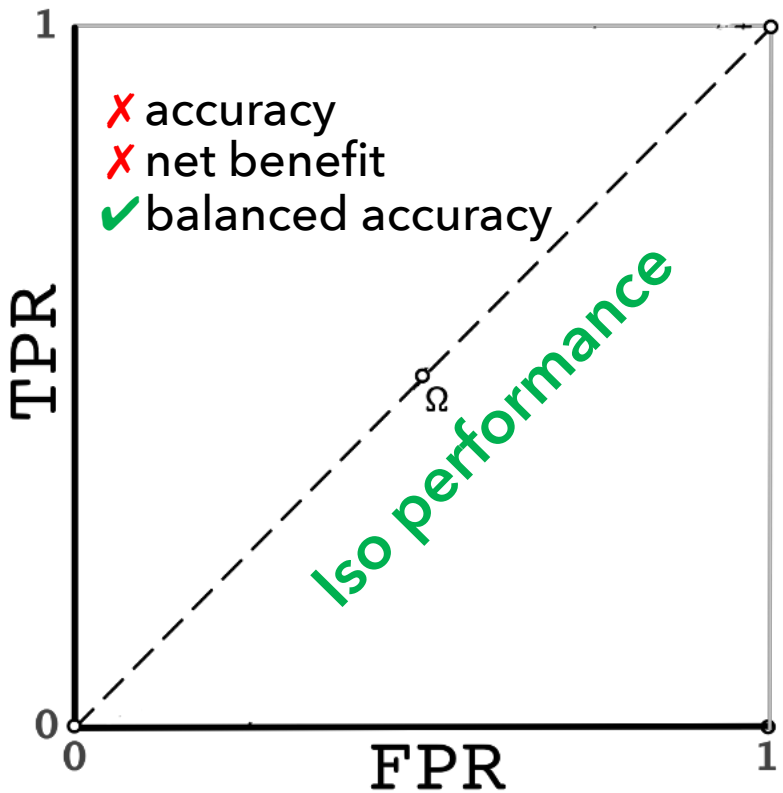


If we are literal/precise:

The center point Ω=(0.5, 0.5) represents chance: a fair coin flip, for binary outcomes.

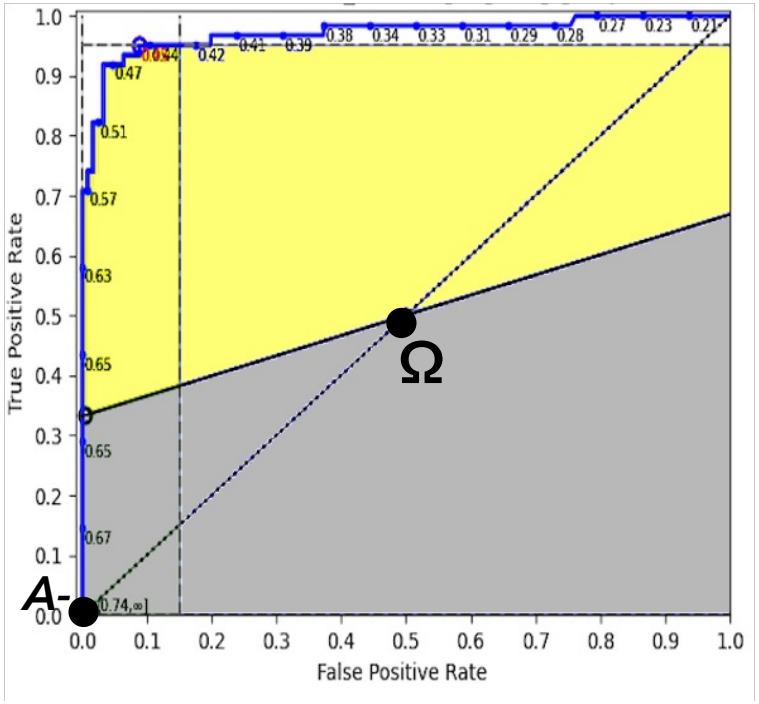
Myth #1b

"The ROC main diagonal represents *[performance equal to]* chance".



Myth #1b

"The ROC main diagonal represents *[performance equal to]* chance".



Point	Acc	BalAcc	NB	CWAcc
A-	62.7%	50%	-1.86	25.2%
Ω	50%	50%	-1.25	50%

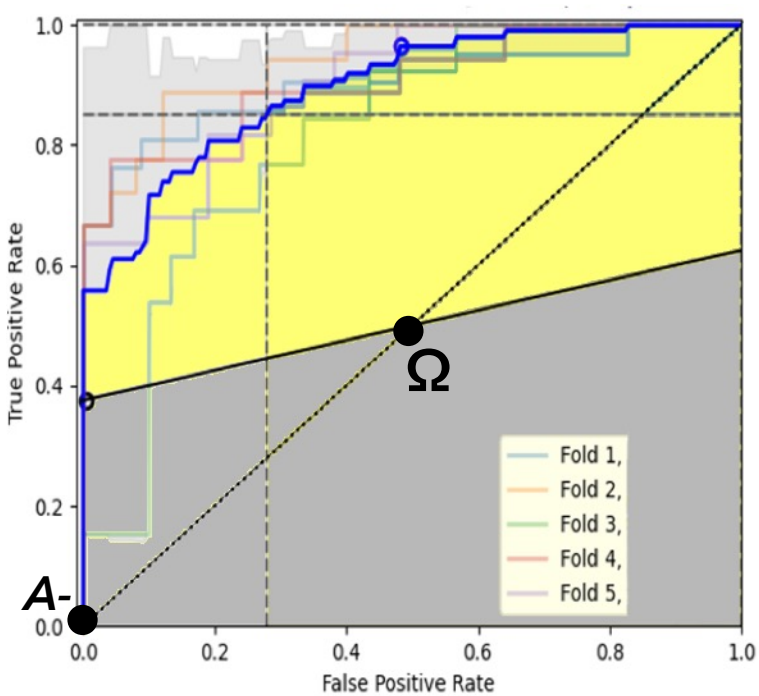
NB ∈ [-2.49, 0]

Class ratio N:P = ~2:1

Costs $C_{FN}:C_{FP}$ = 5:1

Myth #1b

"The ROC main diagonal represents *[performance equal to]* chance".



Point	Acc	BalAcc	NB	CWAcc
A-	55.6%	50%	-2.22	20%
Ω	50%	50%	-1.39	50%

NB ∈ [-2.78, 0]

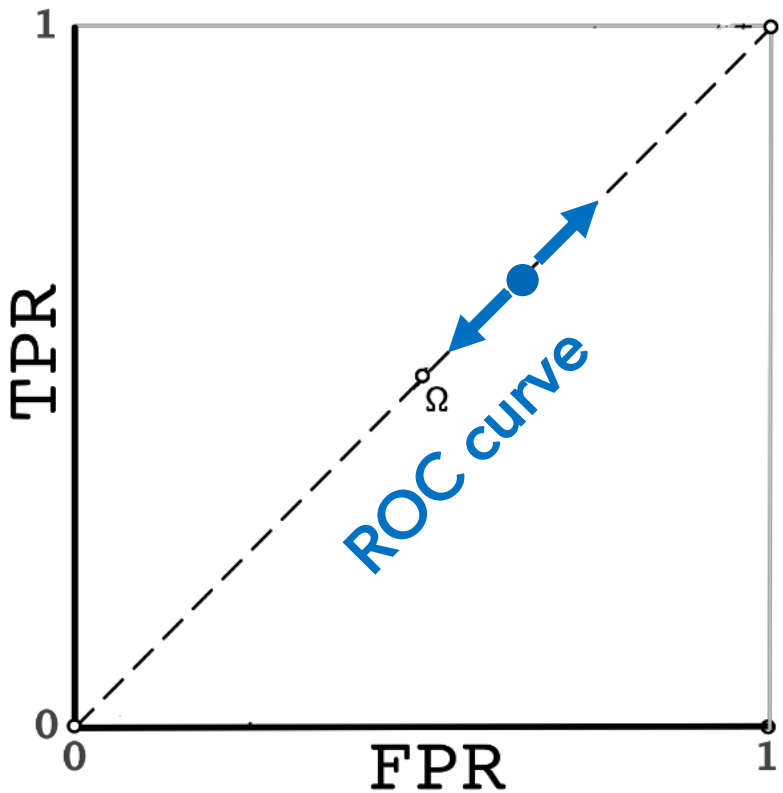
Class ratio N:P = 15:12

Costs $C_{FN}:C_{FP}$ = 5:1

Myth #1c

"The ROC main diagonal represents *[a model performing like]* chance".

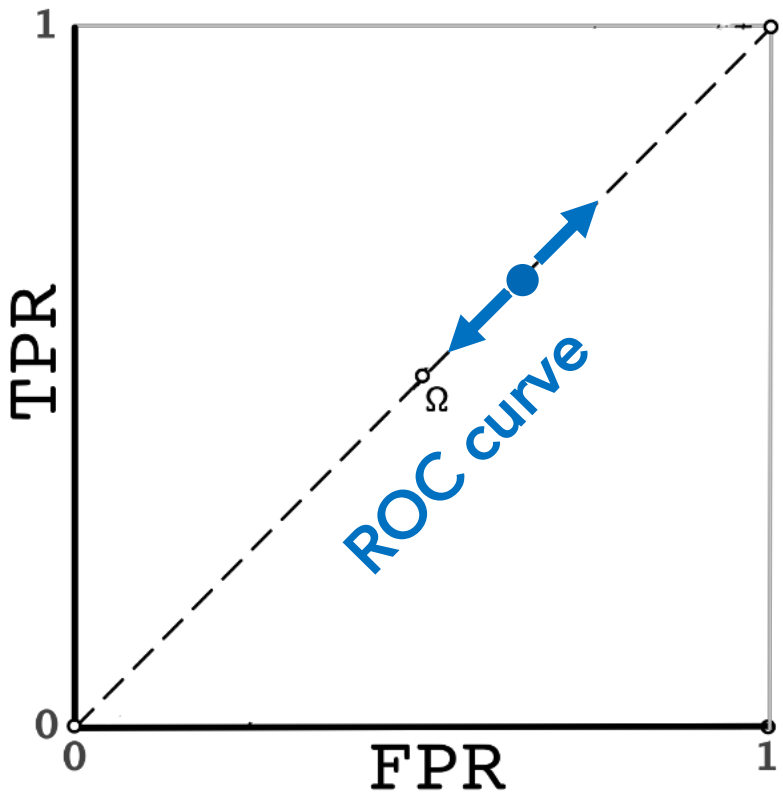
A model with an ROC curve…



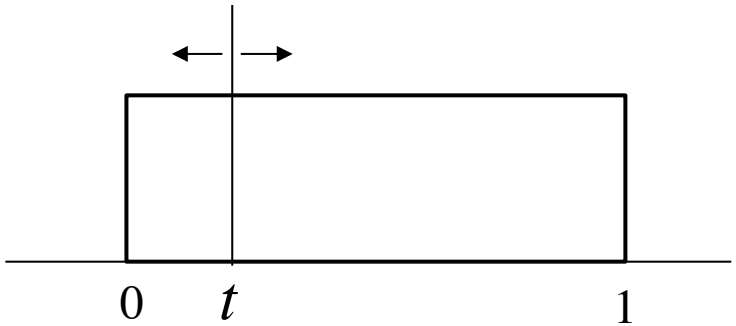
…but a variable threshold, is unlike dictionary or layperson's chance!

Myth #1c

"The ROC main diagonal represents *[a model performing like]* chance".

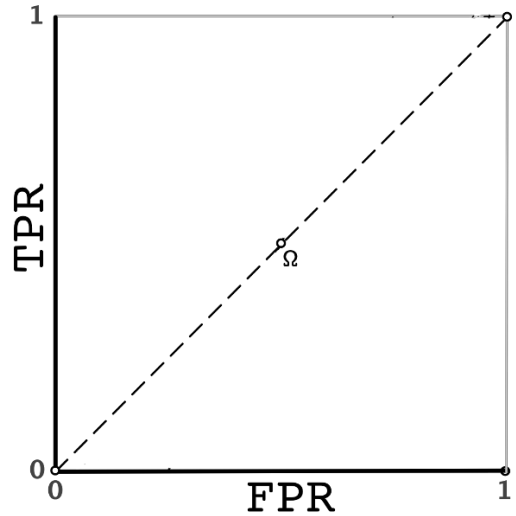


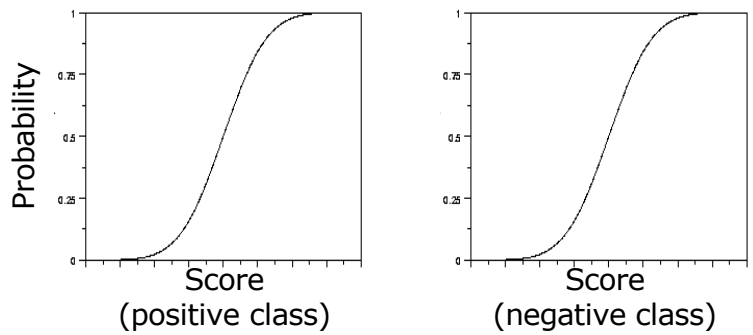
We can form the main diagonal with a weighted coin toss, with variable weight (or threshold, *t*)…



…but this is not layperson's chance!

What is the main diagonal?





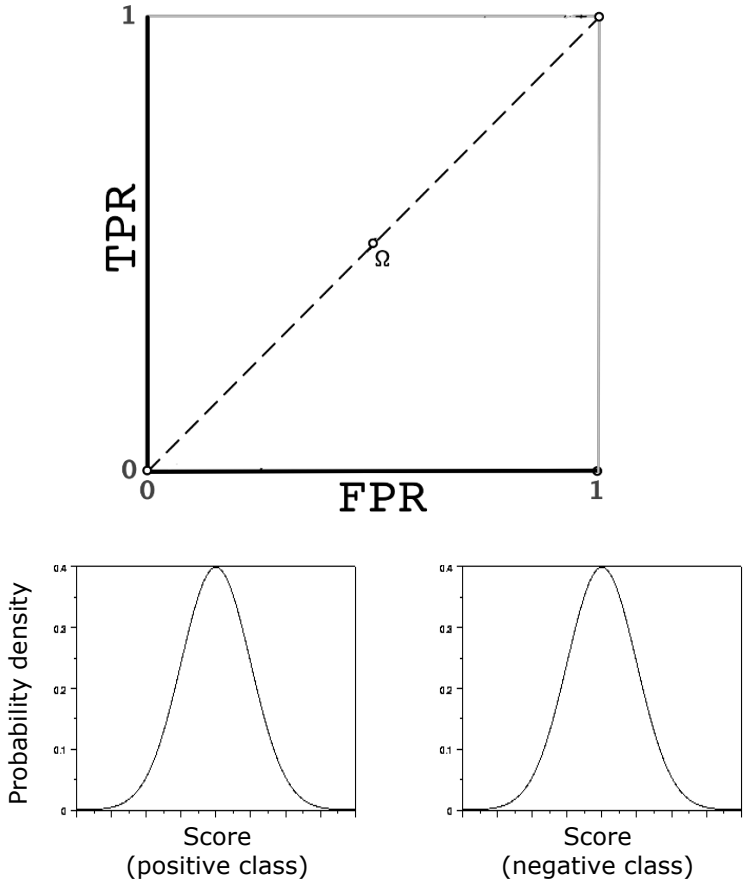
1. An iso performance line per slide 12.

2. A ROC curve per slide 16.

3. The ROC curve for a classifier where any non-zero range of classification scores have an equal chance of either ground truth.

i.e. two same *c.d.f.s*

What is the main diagonal?



4. An iso performance line for:

zero Jensen-Shannon divergence

zero Kullback-Leibler divergence

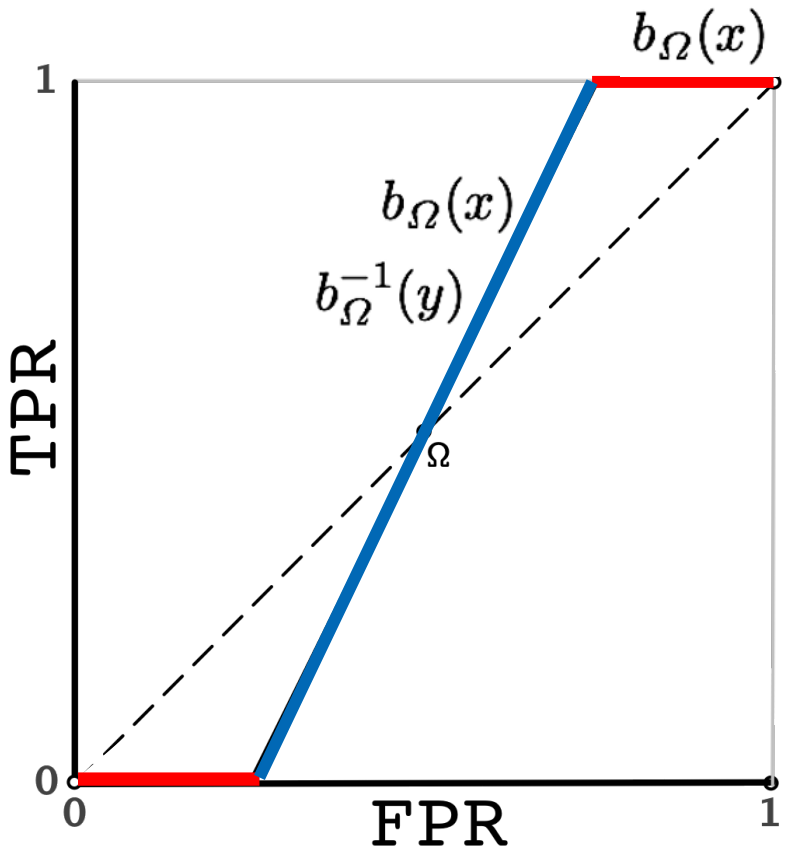
zero Hellinger distance

zero relative (information) entropy

…between the *p.d.f.*s of +/- classes

The binary chance baseline, b_Ω

An iso performance line passing through chance Ω

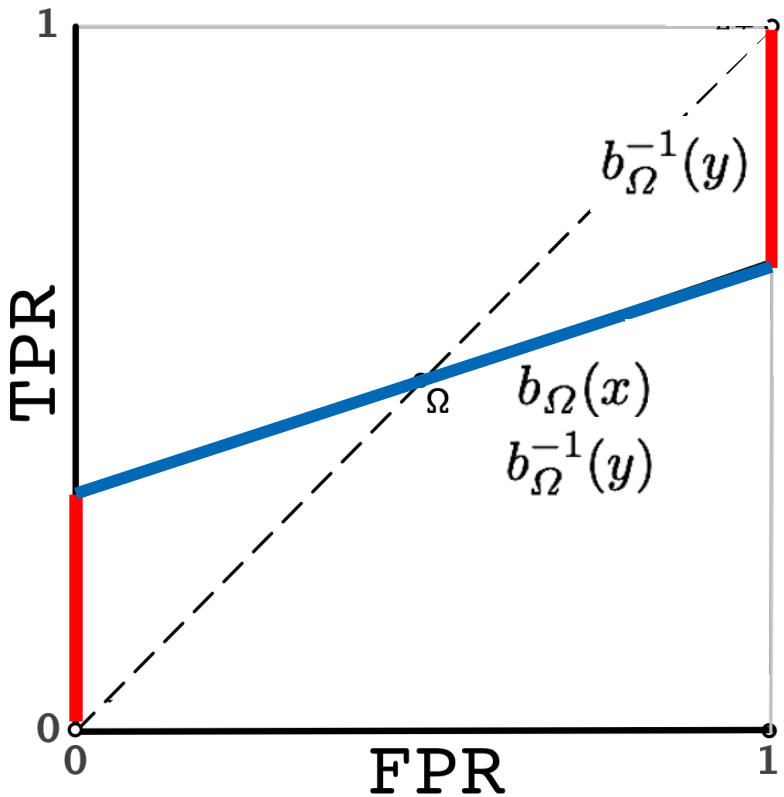


with a slope based on:

- prevalence
- cost of a false positive vs neg
- cost of treatment vs. no treat.

defined within the plot

See the paper for formal definitions



$$b_\Omega(x) := \frac{(1-\pi)}{\pi} \cdot \frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}}(x - 0.5) + 0.5 \text{ , } x \in [0,1]$$

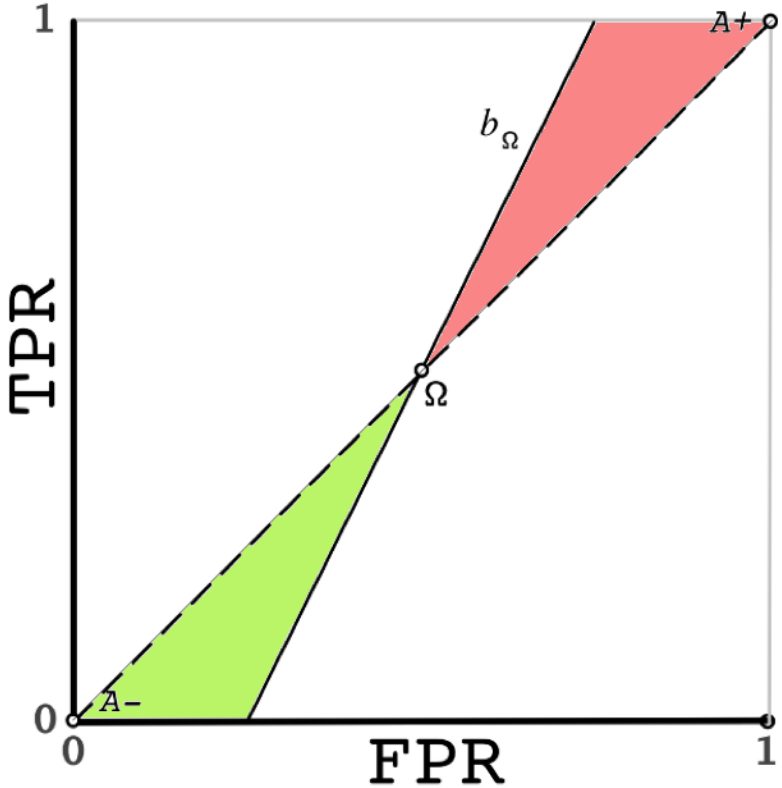
$$b_\Omega^{-1}(y) := \frac{(y - 0.5)}{\frac{(1-\pi)}{\pi} \cdot \frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}}} + 0.5 \text{ , } y \in [0,1]$$

Our baseline can represent…

Measures	Prevalence incl.	Costs incl.
Balanced Accuracy		
Accuracy	✓	
Net Benefit (& cost-weighted accuracy)	✓	✓

Myth #2

"An AUC of 0.5 represents a model which is no better than random chance at predicting a specific outcome."

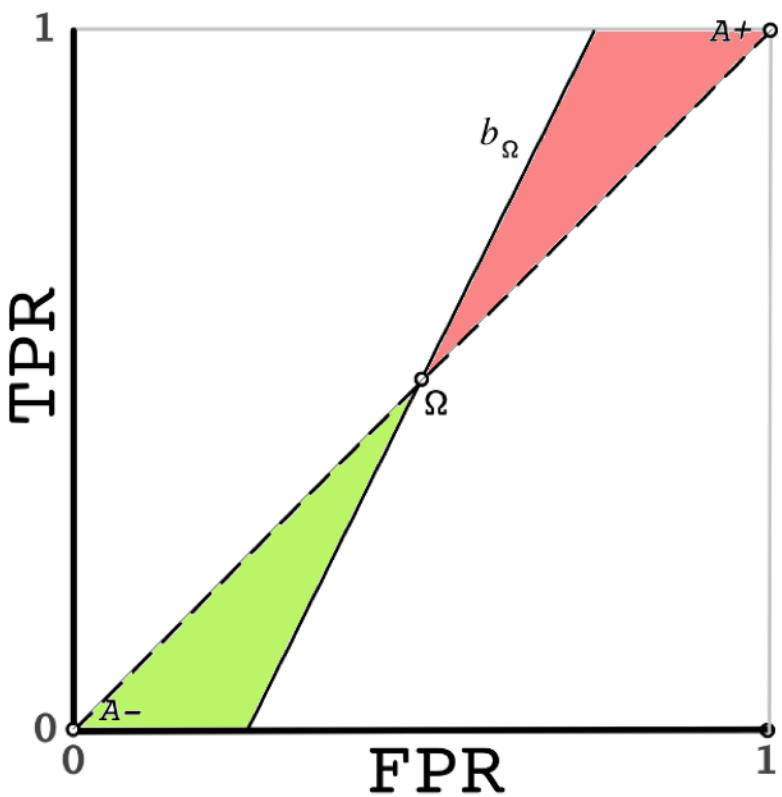


Correction: for <u>low</u> prevalence and balanced costs:

A classifier using a threshold in the bottom half of the main diagonal is better than chance in accuracy & net benefit.

Myth #3

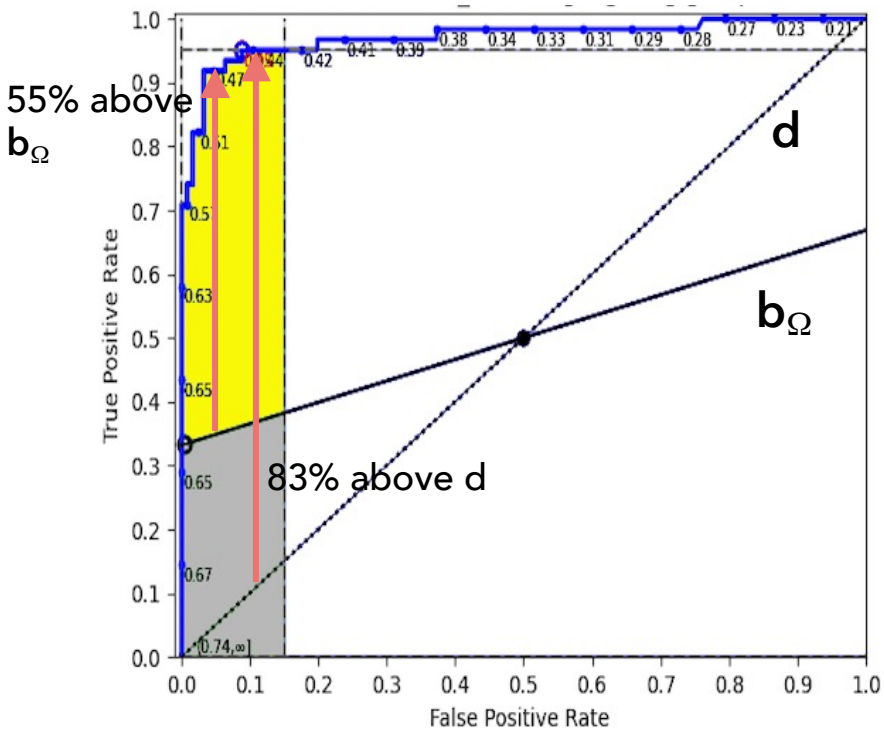
ROC points above the main diagonal are useful.



Clearly, the previous plot also reveals the 3rd statement as a myth.

For partial areas, the difference can be large

The vertical aspect* of useful area (yellow), is much less than the main diagonal suggests.



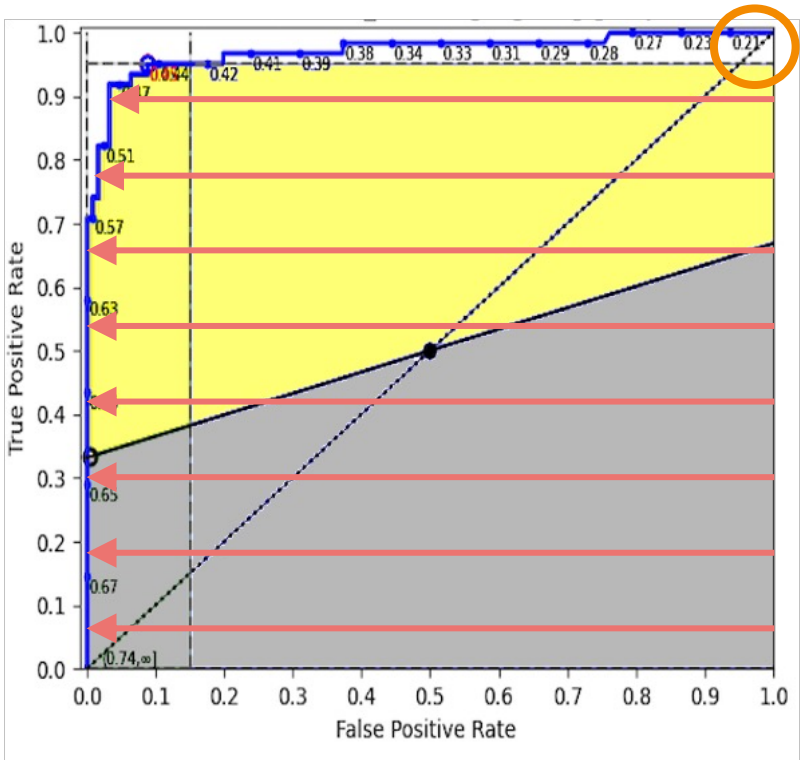
$$[\cdot]_+ := \min(\cdot,\ 0)$$

$$AUC_{\Omega+}(\theta_{xy}) := \frac{1}{2}\int_{x_1}^{x_2} [r(x) - b_\Omega(x)]_+ \, dx + \frac{1}{2}\int_{y_1}^{y_2} \left[\left(1 - r^{-1}(y)\right) - \left(1 - b_\Omega^{-1}(x)\right)\right]_+ \, dy$$

*

The horizontal aspect* is not much different

The horizontal aspect* of useful area (yellow) is slightly less than the main diagonal suggests.



$$[\cdot]_+ := \min(\cdot,\ 0)$$

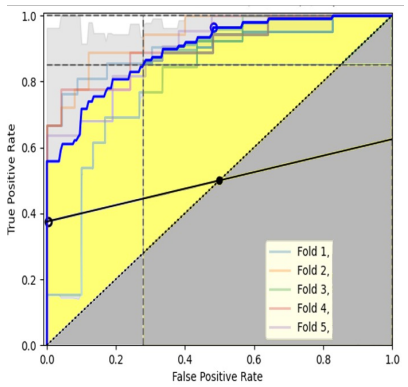
$$AUC_{\Omega+}(\theta_{xy}) := \frac{1}{2}\int_{x_1}^{x_2} [r(x) - b_\Omega(x)]_+ \, dx + \frac{1}{2}\int_{y_1}^{y_2} \left[\left(1 - r^{-1}(y)\right) - \left(1 - b_\Omega^{-1}(x)\right)\right]_+ \, dy$$

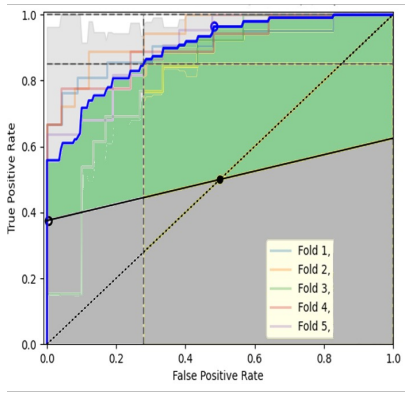
*

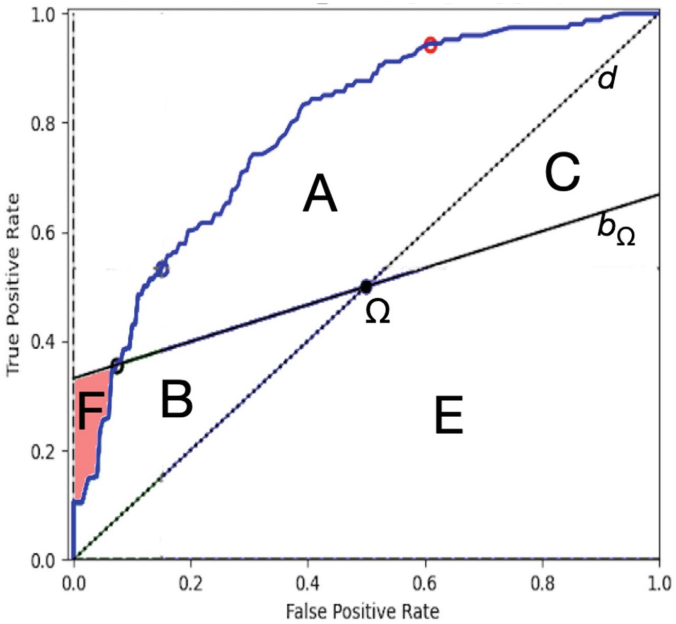
Vertical and horizontal together yields a 2.2% difference in useful area or AUC

Description	AUC in a part	AUC in whole or normalized	Average Sens	Average Spec
ROC in ROI_1 above the main diagonal	AUC_{d1} =28.5%	-	83.0% above	46.8% above
ROC in ROI_1 above the binary chance baseline	$AUC_{\Omega 1}$ =26.3%	-	54.8% above	46.7% above

However, for the whole ROC, the magnitude of area is the same





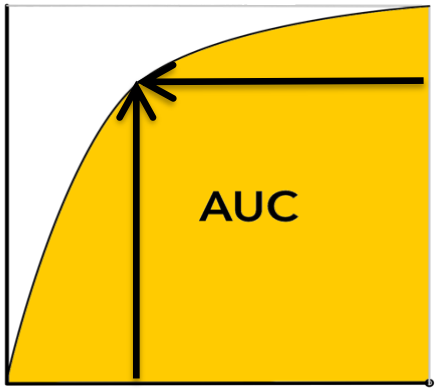


AUC-*d* = A+B
= AUC-(C+E)
= AUC-0.5

AUC-b_Ω = A+C-F
= AUC-(B+E+F)
= AUC-0.5

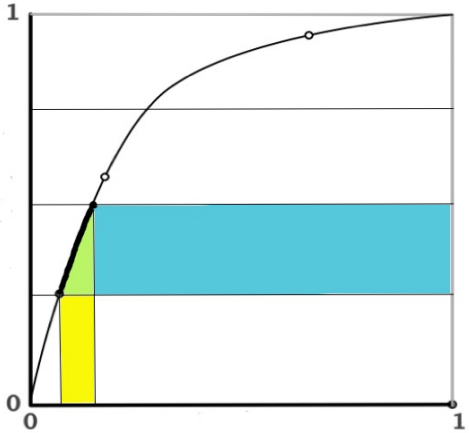
AUC = A+B+C+E

Useful area/AUC combines the vertical and horizontal aspects



$$AUC = \int_0^1 r(x)\,\mathrm{d}x$$

$$AUC = \int_0^1 1 - r^{-1}(y)\,\mathrm{d}y$$



concordant partial AUC$_i$

$$AUC_i(\theta_i) = \frac{1}{2}\int_{x_1}^{x_2} r(x)\,dx + \frac{1}{2}\int_{y_1}^{y_2} 1 - r^{-1}(y)dy$$

In summary

1. The ROC main diagonal is not chance.

2. We define the binary chance baseline.

3. We show the useful ROC areas and points.

Code available

pip install bayesianROC deepROC

Chance ROC Analysis.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Table of contents

+ Code + Text

Compute chance ROC analysis

Show code

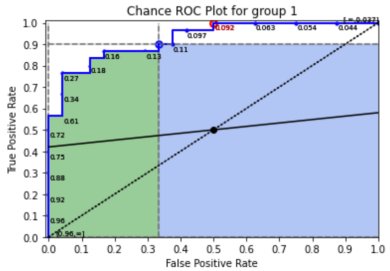
```
Using sample prevalence: 0.556
Set Bayesian prior to binary chance (0.5, 0.5).

Seeking vertical root on interval [0, 1] with range [-0.42, 0.42]
returning vertical root value 1.1e-16 at (0.03,0.43)
```

Chance ROC analysis plots

color fills to be fixed

Show code



Chance ROC analysis numbers

Show code

```
Chance ROC analysis and plots (Carrington, 2022) show performance in subgroups.

Legend: TBD

Description              AUC      AUC or      Average  Average
                         part     normalized  Sens     Spec
group 1: FPR [0.0, 0.3333333333333333]
AUCni                    0.3400   0.7286      0.4200   0.9000
AUCni_d                  0.2190   -           0.2533   0.2000
AUCni_Ω                  0.0381   -           -0.0133  0.0667
group 2: FPR [0.3333333333333333, 0.6666666666666666]
AUCni                    0.1767   0.6625      0.7300   0.5500
AUCni_d                  0.2375   -           0.2300   0.2500
AUCni_Ω                  0.3500   -           0.2300   0.5500
group 3: FPR [0.6666666666666666, 1.0]
AUCni                    0.1633   0.6125      0.8900   0.1500
AUCni_d                  0.0542   -           0.0567   0.0500
AUCni_Ω                  0.2583   -           0.3233   0.1500
group 4: FPR [0, 1]
AUC                      -        0.1633      0.2967   0.0300
AUC_d                    0.1800   -           0.1800   0.1800
AUC_Ω                    0.1800   -           0.1800   0.1800

                              ROC                      Balanced   Avg Net   Cost Weighted
Description                   point          Accuracy  Accuracy   Benefit   Accuracy
Perfect                       (0.000,1.000)  1.0000    1.0000     0.0000     100.0%
Perfectly Wrong               (1.000,0.000)  0.0000    0.0000     -3.0000    0.0%
Binary Chance                 (0.500,0.500)  0.5000    0.5000     -1.5000    50.0%
All Negatives                 (0.000,0.000)  0.5000    0.5000     -2.5000    16.7%
All Positives                 (1.000,1.000)  0.5000    0.5000     -0.5000    83.3%
Intersection point A_Ω        (0.100,0.500)  0.7000    0.7000     -1.3000    56.7%
Whole ROC (group 4):
Unconstrained ROC optimum     (0.900,1.000)  0.5500    0.5500     -0.4500    85.0%
Group 1:
Unconstrained ROI_1 optimum   (0.300,0.600)  0.6500    0.6500     -1.1500    61.7%
Group 2:
Unconstrained ROI_2 optimum   (0.500,0.800)  0.6500    0.6500     -0.7500    75.0%
Group 3:
Unconstrained ROI_3 optimum   (0.900,1.000)  0.5500    0.5500     -0.4500    85.0%
```

Questions?

André M Carrington, Paul W Fieguth, Franz Mayr, Nick D James,
Andreas Holzinger, John W Pickering, Richard I Aviv

acarrington at toh.ca

http://chanceROC.org

The Ottawa Hospital | L'Hôpital d'Ottawa

Inspired by research. Driven by compassion.
Inspiré par la recherche. Guidé par la compassion.

uOttawa
Faculté de médecine
Faculty of Medicine

Department of Radiology,
Radiation Oncology and Medical Physics

Département de radiologie,
radio-oncologie et physique médicale

Glossary

θxy = {x1, x2, y1, y2) the set of parameters defining a ROC region
π prevalence (could be sample or population prevalence, I do not specify)
Ω chance, i.e., binary chance (a horseshoe symbolizes luck, as in chance)
A- the all negative classifier (treat none)
AUC, AUROC: area under the ROC curve
AUC_{d2} AUC minus the diagonal in ROI #2
AUC_i the concordant partial AUC (the proper generalization of AUC to a part; sum of parts = AUC)
$AUCn_i$ the normalized concordant partial AUC (the proper generalization of AUC to a part, comparable to AUC etc)
AUC_Ω useful area relative to chance
$AUC_{\Omega 1}$ useful area relative to chance, AUC minus binary chance in ROI #1
$AUC_{\Omega+}$ positive useful area relative to chance (as opposed to negative)
b_Ω binary chance baseline
$y=b_\Omega(x)$ binary chance baseline, as a function of x
$x=b_\Omega^{-1}(y)$ binary chance baseline, the inverse function, i.e., a function of y
C_{FN} cost of a false negative
C_{FP} cost of a false positive
C_{TN} cost of a true negative
C_{TP} cost of a true positive

c.d.f. cumulative distribution function
d (main) diagonal
FNR false negative rate
FPR false positive rate
NB average net benefit
N negatives
NPV negative predictive value (or inverse precision)
p.d.f. probability density function
P positives
PPV positive predictive value (or precision)
y=r(x) ROC curve as a function of x
$x=r^{-1}(y)$ ROC curve as a function of y
ROC receiver operating characteristic
Sn sensitivity (or TPR, or recall)
Sp specificity (or TNR)
t threshold
TNR true negative rate
TPR true positive rate

References

1. Nawabi J, Kniep H, Elsayed S, Friedrich C, Sporns P, Rusche T, Böhmer M, Morotti A, Schlunk F, Dührsen L, Broocks G. Imaging-based outcome prediction of acute intracerebral hemorrhage. Translational stroke research. 2021 Dec;12(6):958-67.
2. Statlog Heart.
3. Powers, D.M.W. Evaluation: from precision, recall and F-factor to ROC, informedness, markedness & correlation. Tech. Rep. December, Flinders University (2007)
4. Zhou, X.H., McClish, D.K., Obuchowski, N.A.: Statistical methods in diagnostic medicine, vol. 569. John Wiley & Sons (2002)
5. Farrow L, Zhong M, Ashcroft GP, Anderson L, Meek RD. Interpretation and reporting of predictive or diagnostic machine-learning research in Trauma & Orthopaedics. The Bone & Joint Journal. 2021 Dec 1;103(12):1754-8.
6. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. Physics in Medicine & Biology. 2018 Mar 29;63(7):07TR01.
7. Carrington AM, Fieguth PW, Qazi H, Holzinger A, Chen HH, Mayr F, Manuel DG. A new concordant partial AUC and partial C statistic for imbalanced data in the evaluation of machine learning algorithms, BMC Medical Informatics and Decision Making, 2020.
8. Carrington AM, Manuel DG, Fieguth PW, Ramsay T, Osmani V, Wernly B, Bennett C, Hawken S, Magwood O, Sheikh Y, McInnes M, Holzinger A. Deep ROC analysis and AUC as balanced average accuracy for improved classifier selection, audit and explanation. IEEE Transactions on Pattern Analysis and Machine Intelligence, Early Access, January 25, 2022. doi:10.1109/TPAMI.2022.3145392.

References

9. McClish DK. Analyzing a portion of the ROC curve. Medical Decision Making. 1989 Aug;9(3):190-5.
10. Mallett S, Halligan S, Thompson M, Collins GS, Altman DG. Interpreting diagnostic accuracy studies for patient care. British Medical Journal. 2012 Jul 2;345:e3999.
11. Thompson ML, Zucchini W. On the statistical analysis of ROC curves. Statistics in Medicine. 1989 Oct;8(10):1277-90.
12. Dodd LE, Pepe MS. Partial AUC estimation and regression. Biometrics. 2003; 59(3):614–23. https://doi.org/10.1111/1541-0420.00071.
13. McClish DK. Evaluation of the accuracy of medical tests in a region around the optimal point. Academic radiology. 2012 Dec 1;19(12):1484-90.
14. Ma H, Bandos AI, Rockette HE, Gur D. On use of partial area under the ROC curve for evaluation of diagnostic performance. Statistics in medicine. 2013 Sep 10;32(20):3449-58.
15. Vivo JM, Franco M, Vicari D. Rethinking an ROC partial area index for evaluating the classification performance at a high specificity range. Advances in Data Analysis and Classification. 2018 Sep 1;12(3):683-704.
16. Fawcett T. An introduction to ROC analysis. Pattern recognition letters. 2006 Jun 1;27(8):861-74.
17. Candy JV, Breitfeller EF. Receiver operating characteristic (roc) curves: An analysis tool for detection performance. Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States); 2013 Aug 8.

Alternatives have disadvantages

- Kappa omits costs, but includes prevalence
- All positive/negative baselines (treat none/all) can be impractical
- Other plots show few measures vs. ROC

Examples of myth #1

- Figure 1. "**The main diagonal represents chance**" [3]
- "**The (0,0)-(1,1) line segment has an area of 0.5; it is called the chance diagonal**" [4]
- Figure 2.2 "**chance diagonal**" [4]
- "**chance diagonal**" [6]

Details on myth 2

- "**An AUC of 0.5 represents a model which is no better than random chance** at predicting a specific outcome" [5]

- "**The practical lower bound for the ROC curve area is 0.5. The (0,0)-(1,1) line segment has an area of 0.5; it is called the chance diagonal**" [4]

Details on myth 3

- "**Points above the diagonal represent performance better than chance, those below worse than chance**" [3]

- 'Curves lying above this line are considered a representation of "good" detection performance (i.e., better than a coin toss), while those lying below the line are considered "bad" detection performance (i.e., worse than a coin toss)' [17]

Wisconsin breast cancer detail (slide 12)

Description	AUC part	AUC or normalized	Average Sens	Average Spec
group 1: FPR [0, 0.15]				
AUC_i, AUCn_i	0.5401	0.9806	0.9047	0.9926
AUC_i - diagonal	0.2851	0.5176	0.8297	0.4684
AUC_i — b_Ω	0.2634	0.4782	0.5478	0.4672
group 2: FPR [0.00, 1.00]				
AUC	-	0.9715	0.9715	0.9715
AUC - diagonal	0.4715	-	0.4715	0.4715
AUC - b_Ω	0.4715	-	0.4715	0.4715

NP class ratio: 1.68
prevalence: 0.37
Cost_FN : Cost_FP = 5.0:1.0

Description	ROC point	Accuracy	Balanced Accuracy	Avg Net Benefit	Cost Weighted Accuracy
Perfect	(0.000,1.000)	1.0000	1.0000	0.0000	100.0%
Perfectly Wrong	(1.000,0.000)	0.0000	0.0000	-2.4903	0.0%
Binary Chance	(0.500,0.500)	0.5000	0.5000	-1.2452	50.0%
All Negatives	(0.000,0.000)	0.6274	0.5000	-1.8629	25.2%
All Positives	(1.000,1.000)	0.3726	0.5000	-0.6274	74.8%
Intersection ROC and b_Ω	(0.004,0.333)	0.7491	0.6646	-1.2452	50.0%
Whole ROC (group 2):					
Unconstrained ROC optimum	(0.087,0.952)	0.9272	0.9322	-0.1449	94.2%
Group 1:					
Unconstrained ROI_1 optimum	(0.087,0.952)	0.9272	0.9322	-0.1449	94.2%

Statlog heart detail (slide 13)

Description	AUC part	AUC or normalized	Average Sens	Average Spec
group 1: TPR [0.85, 1]				
AUC_i, AUCn_i	0.3888	0.8926	0.9641	0.5487
AUC_i - diagonal	0.1526	-	0.3247	0.4737
AUC_i – b_Ω	0.1959	-	0.4293	0.5487
group 2: TPR [0.00, 1.00]				
AUC	-	0.8951	0.8951	0.8951
AUC - diagonal	0.3951	-	0.3951	0.3951
AUC - b_Ω	0.3951	-	0.3951	0.3951

NP class ratio: 1.25
prevalence: 0.44
Cost_FN : Cost_FP = 5.0:1.0

Description	ROC point	Accuracy	Balanced Accuracy	Avg Net Benefit	Cost Weighted Accuracy
Perfect	(0.000,1.000)	1.0000	1.0000	0.0000	100.0%
Perfectly Wrong	(1.000,0.000)	0.0000	0.0000	-2.7778	0.0%
Binary Chance	(0.500,0.500)	0.5000	0.5000	-1.3889	50.0%
All Negatives	(0.000,0.000)	0.5556	0.5000	-2.2222	20.0%
All Positives	(1.000,1.000)	0.4444	0.5000	-0.5556	80.0%
Intersection ROC and b_Ω	(0.003,0.376)	0.7207	0.6862	-1.3889	50.0%
Whole ROC (group 2):					
Unconstrained ROC optimum	(0.482,0.964)	0.7160	0.7408	-0.3481	87.5%
Group 1:					
Unconstrained ROI_1 optimum	(0.482,0.964)	0.7160	0.7408	-0.3481	87.5%

Sept 3 revisions

- Glossary added
- Slide 6, left figure, NPV baseline corrected.
- Slide 39, two numbers missing were filled in
- Slide 38 and 39, minor changes to text labels for clarity