

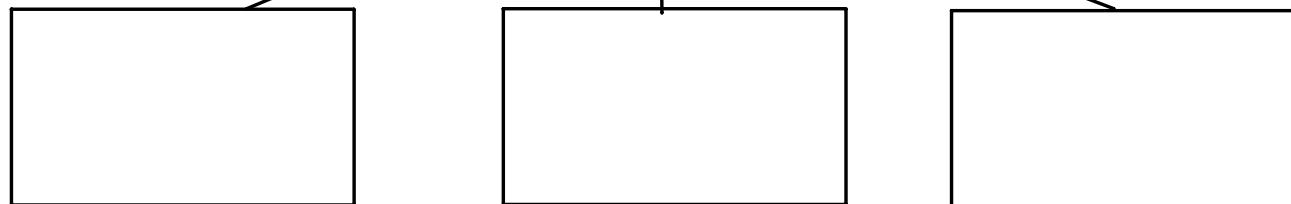
Confusion Matrixes, True Possitive

7:45 AM Tuesday, January 19, 2021

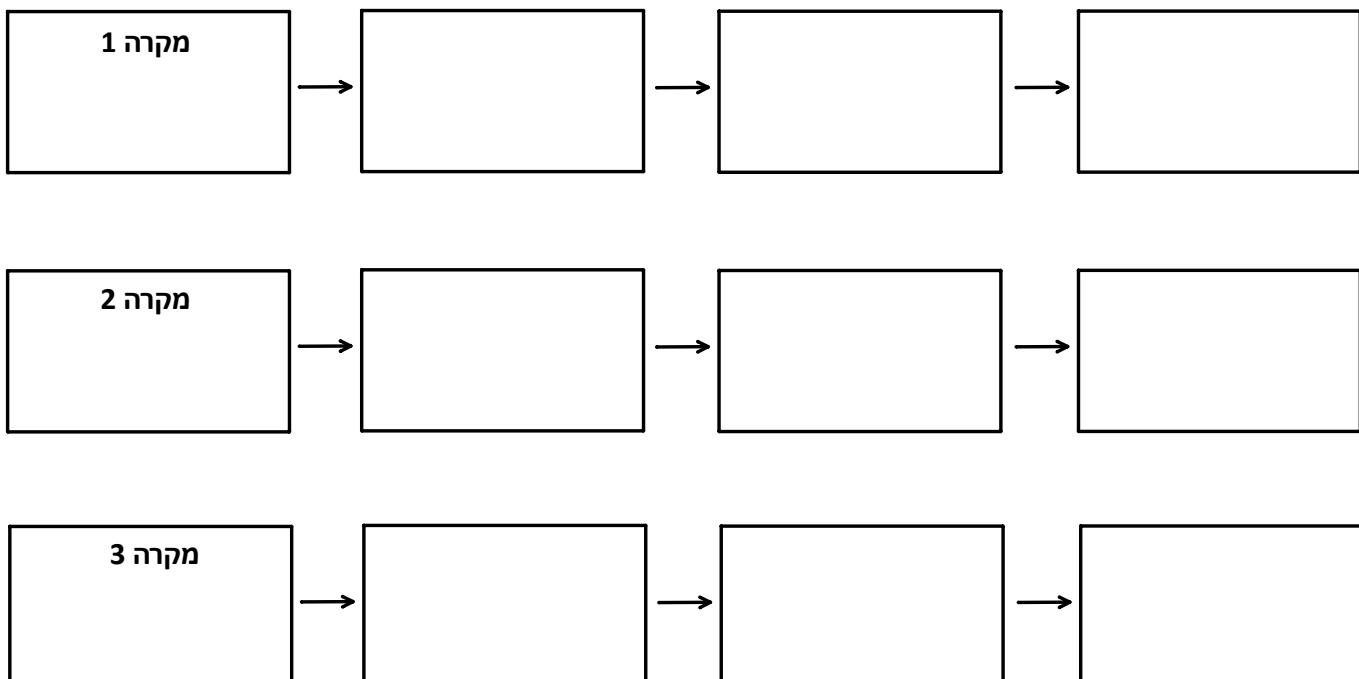
Lec6

2:02:00

KEY - Words
True Positive
True Negative
False Positive
False Negative



איך נפתרו?



ביצד מעריבים הצלחה בקלאסיפיקציה

פונקציות עלות שלמדו לנו כמו MSE, Ceontropy, ועם פונקציית אקטיבציה נcona

אבל: הן לא אינטואטיביות לבני אדם. בפרק זה נלמד מדרדים שנועדו להנגיש את המידע לאנשים:

$$Accuracy = \frac{TruePreds}{AllPreds} = \frac{TN + TP}{AllSamples}$$

כמה סיגרים שגויים? Misclassifications. המשלים ל-1: Accuracy.

שגיאות מסוג I: אזעקות שווה (בראים שסוגו חולמים):

نمודן:

* הסתברות לאזעקות שווה (מתוך המיקרים השליליים) PFA

* לחילופין, ספציפיות (סגוליות): ההסתברות לזייה נכון של בריא = כמה לסמור על אבחנה שלילית מתוך כל הבראים?

= כמה ספציפיות היא אבחנה חיובית

نمודן:

לחלופין, Precision: כמה לסמור על הבדיקה חיובית מתוך אלו שאоторו כחולים נcona

לחלופין: False Positive Rate (FPR)

* להילוףין: ההיסטברות לFP מתוך כל הבראים

אם רציתם להעלות את FP

אנחנו צריכים למזער את FP

$$Specificity = \frac{TN}{TN + FP}$$

Specifity ↑ ↔ ↓ FalsePositive

שגיאות מסוג II: "פיסופים" (חולים שסוגו כבראים)

نمודן:

* רגישות: TPR, Recall, Sensitivity, Detection Rate, prob of detection

כמה סיגרנו מתוך החולים, = ההסתברות לא-אייזור כל הרלוונטיים? כמה לסמור על אבחנה שלילית?

מתוך כל החיוביים - כמה זיהינו בחביבים נcona.

Recall / Sensitivity: אם אנחנו רציתם להעלות את FN

אנחנו צריכים למזער את FN (שגיאות מסוג 2)

$$recall = \frac{TP}{TP + FN} = sensitivity$$

Recall ↑ ↔ ↓ FalseNegative

KPIs in Confusion Matrix

למשל בדיקת חלבון PSA בدم

לזהוי חсад לסרטן הערמוני.

СПЦИФИЧНОСТЬ НЕМНОГА: рабба маобхним

אין חולים בסרטן הערמוני. אך זו

בדיקה קלה וולה. משתמשים בה

כדי לזהות חсад ואז עשוים ביפויסה

(בדיקה יותר יקרה ומסוכנת).

גם ה-SENSITIVITY Precision הוא נמוך ב

PSA אך הדיקוק נמדד בהתאם הבאמת

חולים מתוך המאובחנים בחולים

(רוצים למזער טיעות באבחנת חוליה),

שגיאה מתוך אלו שאובחנו בחולים),

בעוד שההספциficity נמדדת בהתאם

הבדיקות הנכונות של לא חולים

מתוך הלא חולים באמת (רוצים

למזער טיעות באבחנת חוליה שגיה

בקרב לא חולים)

Pg.2

Confusion Matrix		Predicted Value		% אבחנות שפיקחת שגיאות מתוך "החולים" בפועל" - מיעור FN
		חיוני חיובי	חיוני שלילי	
Actual True Value	באמת 1	True Positive – TP (correctly predicted as positives)	False Negative – FN (incorrectly predicted as negatives) (Type II error)	Recall =Sensitivity =TPR predicted positives out of all actual positives: $\frac{TP}{TP + FN} = \frac{TP}{positives}$
	באמת 0	False Positive – FP (incorrectly predicted as positives) (Type I error)	True Negative – TN (correctly predicted as negatives)	Specificity =TNR predicted as negatives (out of all actual negatives): $\frac{TN}{FP + TN} = \frac{TN}{Negatives}$
Precision		אבחנות "בראים" שגיאות מקרוב ה"בראים" באמת – מדדית FP		Accuracy % total correct predictions: $\frac{TP + TN}{TP + TN + FP + FN} = \frac{TP + TN}{All Examples}$
"חולים" באמת מתוך החזויים חולים – (מדדית (EP)		% true malignants out of all predicted as malignant: $\frac{TP}{TP + FP} = \frac{TP}{PredictedPositives}$		

Recall = Sensitivity = TPR
Predicted Positives out of all actual positives

Specificity=TNR
Predicted as negative, out of all actual negatives

Precision
% of true sick, out of all predicted as sick

		הסיווג החצוי (y)		
		חיזוי חיובי – קטגוריה 1	חיזוי שלילי – קטגוריה 0	
הסיווג בפועל (t)	באמת 1	True Positive – TP 890	False Negative – FN (Type II error) 20	910
	באמת 0 שלילי	False Positive – FP (Type I error) 10	True Negative – TN 80	90
		900	100	

דוגמה

בדיקת קורונה ל-1000 איש
התגלו בבריאים, מתוכם 10 חולמים באמת
התגלו בחולים, מתוכם 20 בריאים באמת

$$\text{Accuracy} = \frac{TN+TP}{AllSamples} = \frac{890+80}{1000}$$

$$\text{recall} = \frac{TP}{TP + FN} = \frac{890}{890 + 20}$$

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{890}{890 + 10} = \frac{89}{90}$$

$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{80}{80 + 10} = \frac{8}{9}$$

$$FPR = 1 - spec = \frac{1}{9} = \frac{FP}{RealNeg} = \frac{10}{90}$$

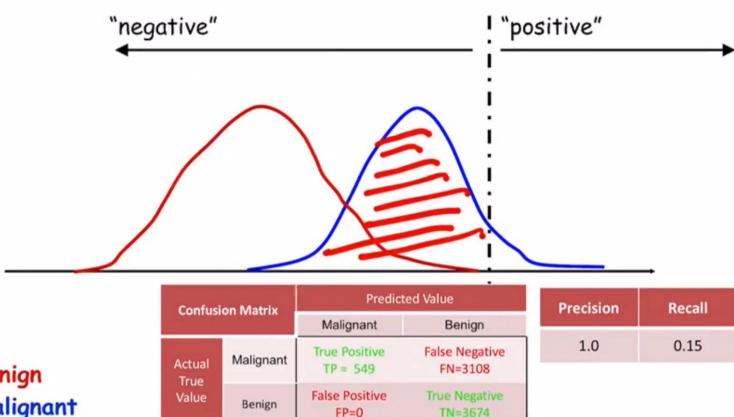
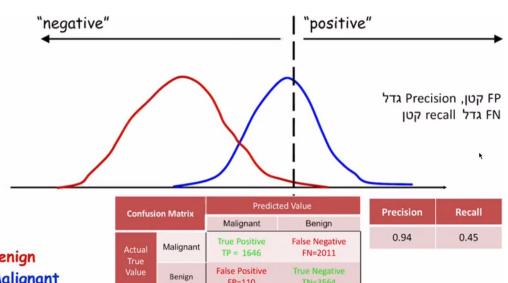
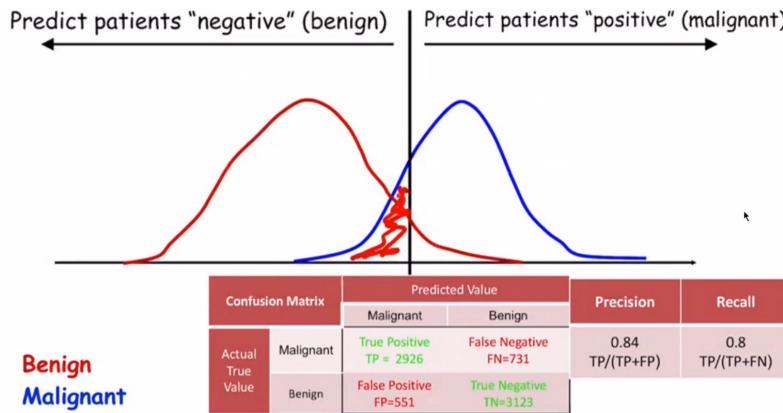
Trade-off between the Precision & Recall

הזהות הסף ימינה, אנחנו צריכים למצוא את כל החולים
[ולא אכפת לנו להגדיל על מישחה בריא שהוא חולה]

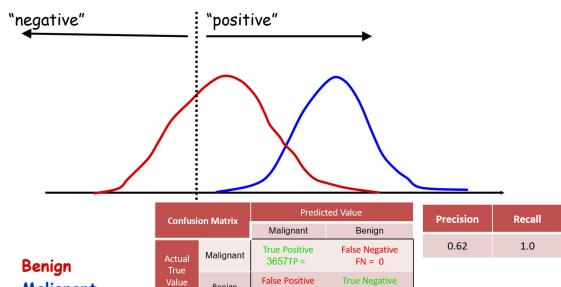
$$\text{Recall} \uparrow \leftrightarrow \downarrow \text{FalseNegative}$$

$$\text{Specificity} \uparrow \leftrightarrow \downarrow \text{FalsePositive}$$

הזהה ימינה (3 שקפים)

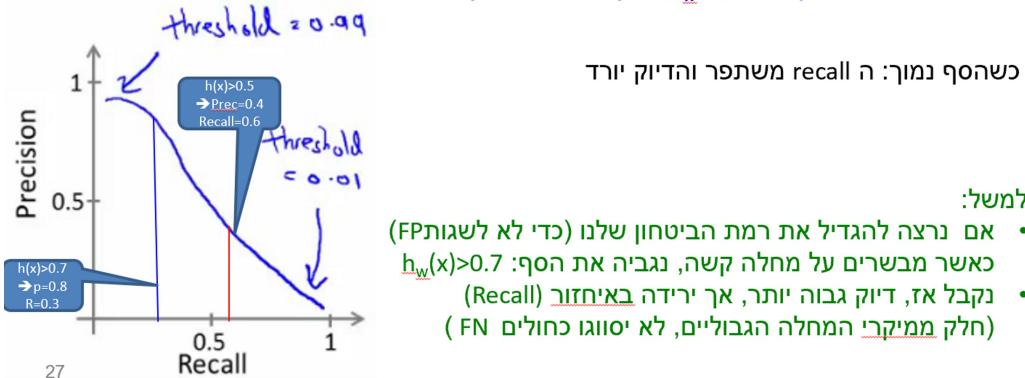


Moving the Threshold: left



Trading off Precision and Recall

- בקולסיפקציה מחשבים הסתברות אך החלטה כיצד לסוג תליה בסוף.
למשל: במיקרום רבים מחליטים לסוג לקטgorיה, אם הסיכוי גדול מ 0.5.
- ונכל לשנות את הסוף וכן לשנות את הפרופורציות של Precision, Recall בהתאם להשפעת הכלכליות
- ונכל ליצור גרפ שראה כיצד משתנים הדיק (precision) והאיחזור (recall) כפונקציה של הסוף.
- אם נשימוש בסוף $h_w(x) > 0.5$, נקבל מדי דיק ואיחזור מסוימים.



באשר חלק מהקטגוריות נדירות

- נניח שעשינו רגרסיה לוגיסטייה לחיזוי מחלת וקיבלנו $\text{accuracy} = 0.99$
האם זה טוב?
- אם ידוע לנו שהסיכוי ה prior למחלת באוכלוסייה הוא רק 0.005? (חצי אחוז מהבדיקות באמת חולים).

	$y=1$	$y=0$	
T=1	TP=0	FN=5	Pos=5
T=0	FP=0	TN=995	Neg=995
0	1000	Tot=1000	

- מה תהיה הaccuracy של אלגוריתם "טיפש" שמסווג, תמיד "בריא"?
- 0.995
- מה יהיה ה Precision, Recall של האלגוריתם ה"טיפש"?
- $p=TP/(TP+FP)=(0/0)=1$, $\text{Recall}=TP/(TP+FN)=(0/5)=0$
- במקרים בהם אחת הקטגוריות נדירה, מדידת ה accuracy בלבד לא תעזר:
- אנחנו מעוניינים בשני הממדדים P, R יהיו גבוהים.

F1-Measure - Combine Precision and Recall

Makes sure that P and R both would be high

$$F1 = 2 \cdot \frac{P \cdot R}{P + R}$$

- כאשר P או R הם 0, המדד נותן את המינימום 0
- כאשר שניהם 1, המדד מקבל את ערכו המקסימלי 1

ממוצע הרמוני ל m מספרים

$$harmonicAvg = \frac{m}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_m}}$$

$$\frac{2}{B + A} = \frac{2AB}{B + A} \leftarrow m = 2$$

מתן משקל לפרמטר שיתור חשוב לנו:

$$F - Measure = \frac{1}{\alpha \cdot \frac{1}{Precision} + (1 - \alpha) \cdot \frac{1}{Recall}}$$

$\alpha > 0.5$ – prefers Precision (less FP)

$\alpha < 0.5$ – prefers Recall (less FN)

Balanced F-Measure ($\alpha = 0.5$)

$$F1 = \frac{2PR}{P + R} \leftarrow NormalF1$$

ממוצע נוסף שמתאים בעיקר לריבוי קטגוריות =

Balanced Accuracy

שיטה נוספת לתחת מספר אחד המשערך ביצועי קלסיפייר באשר הטסט איןנו מאוזן

- ממוצע (רגיל) של ה Recall ים של כל קטgorיה.
- רלוונטי ל multiclass וגם ל binary
- כאשר הטסט מאוזן, קיבל את ה Accuracy
- בקלאסיפיקציה ביןארית, זהה הממוצע בין הרגישות (recall) לסגוליות (specificity)

למשל: בקטgorיה A : מתוך 100 סוגנו נכון 90 – 90%

ובקטgorיה B מתוך 200 סוגנו נכון 160 80%

חשב את ה accuracy, accuracy Balanced

$$Accuracy = 200 * 0.8 + 100 * 0.9 / (300) = 250 / 300 = 0.833$$

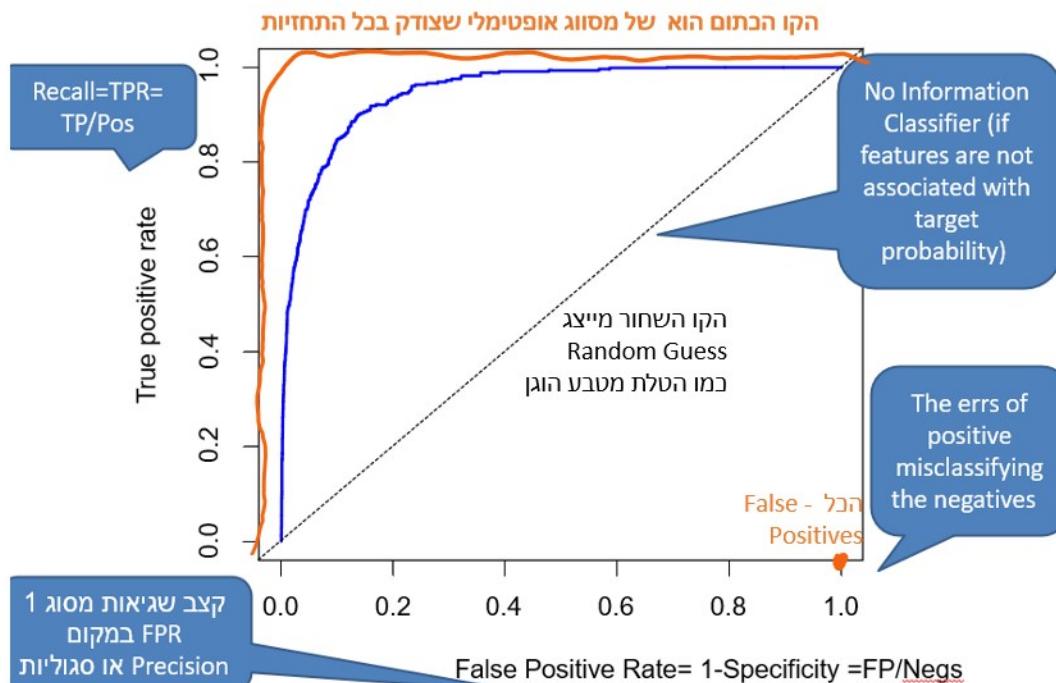
$$B. Accuracy = (0.8 + 0.9) / 2 = 0.85$$

multiclass - ממוצע של הרגישויות זה ה BA ל Sensitivities



ROC Curve (Receiver Operating Characteristics)

Traces Recall vs FPR=1-specificity,
as we vary the threshold for the posterior $P(+|x)$



דירוג נוסף על הגרף הוא בעדרת השטח מתחת לגרף

השיטה שמדרגת לפי AUC מדרגת את האלגוריתם (יחס Recall - Specificity) באופן כללי. לעומת זאת F1 שנותן יחס מוקמי בין Recall & Precision

באשר ש 0.5 זה השטח מתחת לגרף של משטנה מקרי (כמו הטלת מטבע) - הקוו השחור שבגרף
ובכל שזה יותר מתקרב ל 1 כבה אנחנו יותר מדויקים (בזיהוי מושלם)

