

מונח: $\hat{y} = f_{\theta}(x)$
מחיר סיום ממוצע

שאלה: איך נבדוק, האם המודל הוא טוב?
איך אנו מודדים?

\bar{X}

מטרה: סיום ביצוע
מטרה: $y \in \{0, 1\}$
י-מספר שורה $i = 1, \dots, M$

מדידה

- '1' - positive group or result
- '0' - negative group or result
- Y - actual class
- \hat{Y} - predicted class

Confusion matrix

* כל מדידה = סה"כ תוצאות מודל

ציון: מדידה \leftarrow מדידה או ציון
negative positive

יש מ מדידות מדידה

היה ציון

ציון מדידה: $\hat{Y} = 1$
גורמים מדידה מדידה
מדידה של מדידה

Covid 19
239
מדידה

		Predicted	
		Yes	No
Actual	Yes	141	67
	No	0	31

FP
שגיאה
מדידה מדידה

$TP = 141$ - מדידה מדידה

$FN = 67$ - היה מדידה

מדידה מדידה

Performance Metrics

Accuracy

מדידה הכי טובה

$$\text{Accuracy} = \frac{\text{correct predictions}}{\text{total predictions}}$$

$$= \frac{TP + TN}{TP + NT + FP + FN}$$

$$\text{Accuracy} = \frac{141 + 31}{239} = 0.7196652 \approx 72.0\%$$

מדידה: כמה פעמים
מדידה ציון
(באופן חסר)
מדידה: כמה פעמים
מדידה יצאה
מדידה

		Predicted	
		Yes	No
Actual	Yes	141	67
	No	0	31

Precision

$$\text{Precision} = \frac{TP}{FP + TP} = \frac{\text{Correctly predicted 1's}}{\text{All predicted 1's}}$$

$$\Pr(Y = 1 | \hat{Y} = 1)$$

Actual \ Predicted	Yes	No
Yes	141	67
No	0	31

$$\text{Precision} = \frac{141}{0 + 141} = 1 = 100\%$$

מכ"מ: כמה מלפוסים
יש המוק מה שזה
כמלפוס
בזיקה: כמה חיוביים המוק
מה שזה כחיובי

Recall (sensitivity)

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{\text{Correctly predicted 1's}}{\text{Actual 1's}}$$

$$\Pr(\hat{Y} = 1 | Y = 1)$$

Actual \ Predicted	Yes	No
Yes	141	67
No	0	31

$$\text{Recall} = \frac{141}{141 + 67} = 0.678 = 67.8\%$$

מכ"מ: כמה מלפוסים זה
למק סה"כ המלפוסים?
בזיקה: כמה חיוביים למק
סה"כ החיוביים זהו?

Specificity

$$\text{Specificity} = \frac{TN}{FP + TN} = \frac{\text{Correctly predicted 0's}}{\text{Actual 0's}}$$

Actual \ Predicted	Yes	No
Yes	141	67
No	0	31

$$\text{Specificity} = \frac{31}{0 + 31} = 1 = 100\%$$

א חיובי
הזעזע

מכ"מ: כמה ציבור
זוהו למק סה"כ
הציבורים?
בזיקה: למק כל שליליים
כמה זוהו
כשליליים

F₁-score

The harmonic mean between precision and recall,

$$F_1 = \frac{2}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = \frac{2TP}{2TP + FP + FN} \quad (8.13)$$

Example 8.1: Back to the previous example,

$$F_1 = \frac{2 \cdot 141}{2 \cdot 141 + 0 + 67} = 0.808 = 80.8\% \quad (8.14)$$

Imbalanced Dataset

יטן/בסיס נתונים: 990 ערכים של 0, 10 ערכים של 1
 ניקה מסווג $\hat{y}=0$ ללא טעות
 990 ערכים של 0, 10 ערכים של 1

		Predicted	
		Yes	No
Actual	Yes	0	10
	No	0	990

and the resulting quantities are

$$\text{Accuracy} = \frac{990}{1000} = 0.99 = 99\%$$

$$\text{Precision} = \frac{TP}{FP + TP} = \frac{0}{0 + 0} = \text{Undefined}$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{0}{0 + 10} = 0$$

$$\text{Specificity} = \frac{TN}{FN + TN} = \frac{990}{1000} = 0.99 = 99\%$$

$$F_1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} = \frac{0}{\dots} = 0$$

בסיס נתונים עם מספרים לא מאוזנים: הגישה לא מוצלחת!

Receiver Operating Characteristics (RoC)

thr ברירה מתאים $\frac{1}{2}$

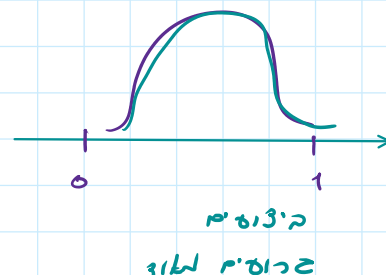
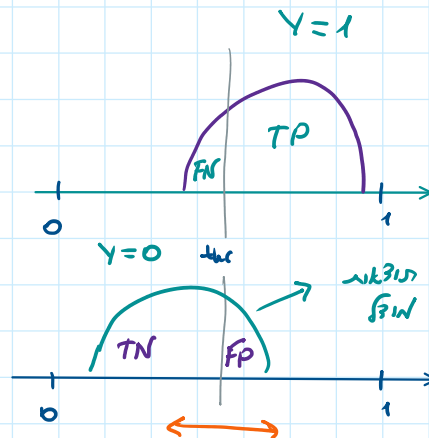
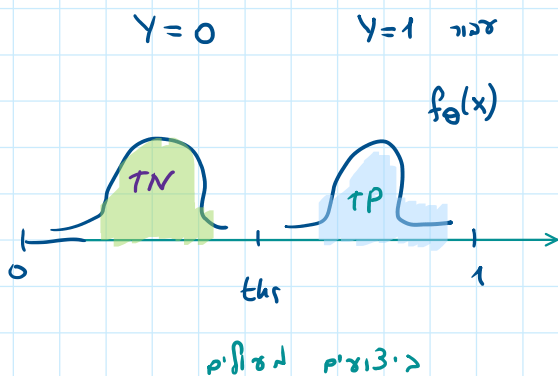
ישנם מוצגים עם מוצג הסתברותי

$$\Pr(\hat{y} = 1) = f_{\theta}(x)$$

$$\hat{y} = \begin{cases} 1 & f_{\theta}(x) \geq \text{thr} \\ 0 & f_{\theta}(x) < \text{thr} \end{cases}$$

conf. matrix

יש $\leq \text{thr}$ יש



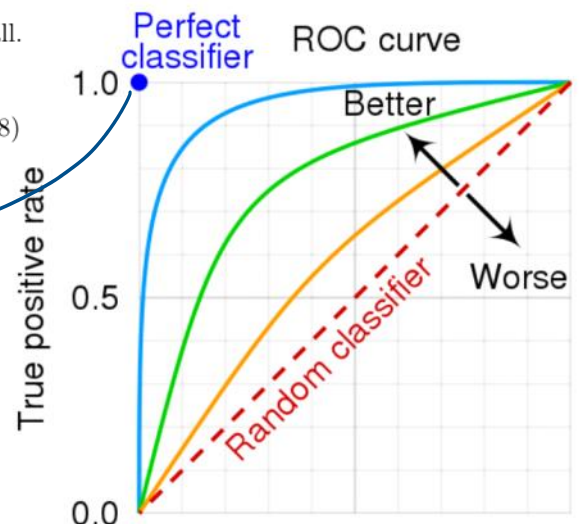
- True Positive Rate (TPR) is a synonym for recall.
- False Positive Rate (FPR) is defined by

$$FPR = \frac{FP}{FP + TN} = 1 - \text{specificity} \quad (8.18)$$

$$\text{recall} = 100\% \\ \text{specificity} = 100\%$$

Area under curve (AUC)

RoC מתאר =
 thr - > ברירה מתאים



הער - > מילון יגל מ'813'2 =

$$\frac{1}{2} \leq AUC \leq 1$$

מ'25
8204

234
863'k

