

## מספרה מסומנת ומספרה 11 הרצאה Performance Evaluation

| True class ↓  | Predicted class |               |
|---------------|-----------------|---------------|
|               | positive        | negative      |
| positive (#P) | #TP             | #P - #TP (FN) |
| negative (#N) | #FP             | #N - #FP (TN) |

מסווגים = מסווג מתאים אובייקט לסט מוגדר מראש של קטגוריות או קלאסים, בהתבסס על סט של ערכי פיזיקליים שצפינו בהם. בהרצאה זו נתמקד בשיעור הביצועים של מסווגים שפותרים את המטרה של שני קלאסים: "חיובי" / "שלילי".

במצב של שתי מחלקות יש מטריצה ה-confusion. נניח כי יש לנו 100 דגימות כאשר הערך האמיתי הוא 70 הם חיוביים ו-30 הם שליליים. אם בקלאס החזוי חזינו שרק 50 הם חיוביים אזי נסמן 50 שהם #TP (true positive) ו-20 מתוך הדאטה האמיתי הם חזויים לא נכון. לסיכום, האלכסון הראשי הוא חיזוי נכון והאלכסון המשני הוא חיזוי שגוי. 4 המספרים במטריצה זו 2x2 אמורים להיסכם ל-100, כמספר הדגימות.

- Reduce the 4 numbers to two rates

$$\text{true positive rate} = \text{TPR} = (\#TP)/(\#P)$$

$$\text{false positive rate} = \text{FPR} = (\#FP)/(\#N)$$

- These rates are independent of class ratio
- We can/should compute confidence intervals for both

• Recall – TP+FN – כל אלו שהם חיוביים באמת, TP הם כל אלו שחזינו שהם חיוביים. המכנה נקרא Real Positive. נקרא גם: **True Positivity Rate**

**נרצה שזה יהיה מספר גדול.**

• Specificity – הסיכוי לחזות שליליים באופן נכון.

• False Positive Rate = 1 – specificity **נרצה שזה יהיה מספר קטן.**

• Error = 1 - accuracy

### מידת הביצועים בסקלארים:

$$\text{Accuracy} = \frac{\text{Correctly Classified}}{\text{All Instances}} = \frac{TP+TN}{TP+FP+TN+FN}$$

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

$$\text{Recall} = \frac{TP}{TP+FN} \quad (\text{aka Sensitivity})$$

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

$$\text{FPR} = 1 - \text{Specificity}$$

$$\text{We often study } \text{TPR} = \frac{TP}{TP+FN} \text{ and } \text{FPR} = \frac{FP}{TN+FP}$$

| True | Predicted |     |
|------|-----------|-----|
|      | pos       | neg |
| pos  | 40        | 60  |
| neg  | 30        | 70  |

**Classifier 1**  
TPR = 0.4  
FPR = 0.3

| True | Predicted |     |
|------|-----------|-----|
|      | pos       | neg |
| pos  | 70        | 30  |
| neg  | 50        | 50  |

**Classifier 2**  
TPR = 0.7  
FPR = 0.5

| True | Predicted |     |
|------|-----------|-----|
|      | pos       | neg |
| pos  | 60        | 40  |
| neg  | 20        | 80  |

**Classifier 3**  
TPR = 0.6  
FPR = 0.2

### דוגמה: 3 מסווגים

נתבונן על שלושה מסווגים שונים שמסווגים 200 דגימות. נשים לב שה-test data אותו הדבר, אבל האלגוריתמים יכולים להיות שונים / הפרמטרים שלהם שונים / האתחול שלהם שונה / סט פיזיקלי שונה. את הירוק למשל בטוח לא ניקח מפני שמול האדום יש לו TPR יותר נמוך וגם FPR יותר גבוה.

נשים לב שתוצאות שהן FN – הן תוצאות שמשמעותן שחזינו שלילי אבל האמת הוא חיובי, נגדיר מדד ביצועים (נקרא לפעמים expected benefit)

$$\pi = \alpha * \text{TPR} - \text{FPR}$$

שתופס את האיזון הזה:

המסווג המועדף יהיה תלוי באלפא (אי שלילי), טווח המשקלים שניתנים לשני סוגי הטעויות. עבור אלפא גדולה מ-1 נעניק יותר חשיבות ל-TPR (למשל במקרה של סרטן, שנרצה למתן תוצאות של FP), ועבור אלפא קטנה מ-1 נעניק יותר חשיבות ל-FPR. עבור אלפא שווה נקבל עדיפות שונה כפי שניתן לראות << (השיפוע הוא אחד חלקי אלפא)

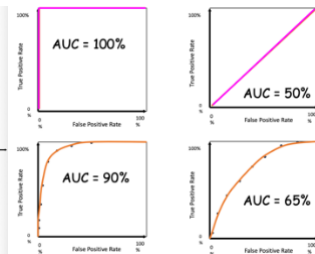
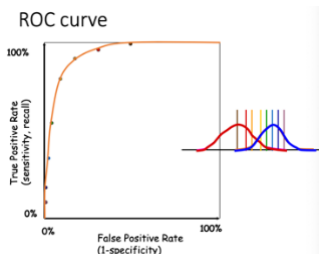
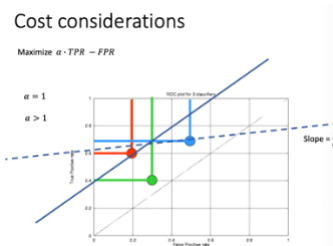
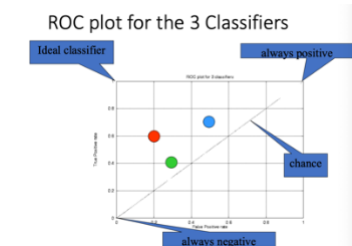
### צורת ROC Curve

- מסווג מניב נקודת ROC יחידה.

- אם למסווג יש פרמטר "רגישות", השינוי שלו יתנו לנו סדרה של נקודות ROC (confusion matrices).

- לחילופין, אם המסווג מונב על ידי אלגוריתם למידה, סדרה של נקודות ROC (עבור ה-test data) יכולה להיווצר על ידי שינוי התכונות של התלחף האימון כמו ההיפר פרמטרים של האלגוריתם.

- כל שהשטח מתחת ל-curve (AUC = area under curve) יותר גדול כך נהיה יותר מרוצים מהבדיקה / הפיצ'ר!

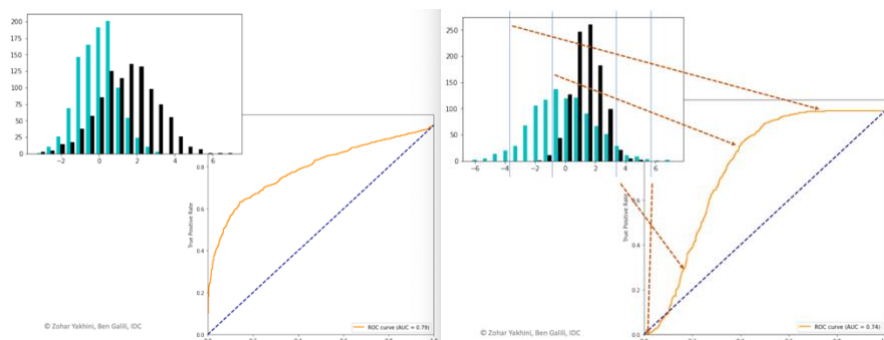


## דוגמה:

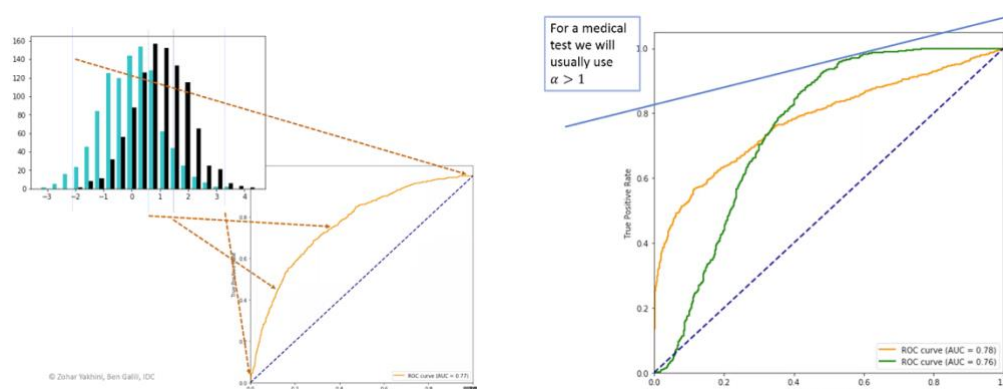
יצרנו שני גאוסיאנים כשהתוחלת של אחד מהם שונות. נשנה את התוחלות ונראה כי ה-AUC משתנה ולעתים נעשה יותר טוב. אלפא יותר גדול מ-1 משמעותי יותר עבור ה-FN (נרצה "להעניש" יותר בדיקות לסרטן למשל)  
**תלוי באלפא איזה AUC מבין שניהם יותר טוב:**

פוזיטיב = השחור

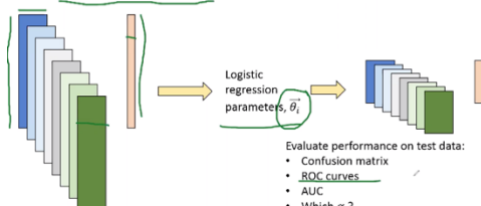
נגטיב = הכחול



נשים לב שבדוגמה זו, כל threshold מייצג קלסיפיקציה וכך נוצר ה-ROC Curve. כך למשל בעמודות השמאליות של הגרף הימני (הכוחות) הבודדות = משמעותן כל מי שבשמאל חולה וכל מי שבימין בריא, נקבל TPR של 100%, כלומר 1 כי הוא מפלג את הדאטה נכון.



Data: different sets of vectors of 4 features (results of a certain measurement technology). 99 instances w labels representing AMI status: H (negative) and M1 (positive)



## דוגמה: בחירת פיצ'רים ו-thresholds עבור קלסיפיקציה LoR

הדאטה שלנו היא קבוצות שונות של וקטורים בעלי 4 פיצ'רים. לכל סט של 4 פיצ'רים נבנה מטריצה. הווקטור הוורוד מייצג את הלייבלים, מריצים גרסיה לוגיסטית, שלומדת וקטור טטה 5-ממדי. ניגש ל-ROC Curve של הלמידה שלנו.

**נרצה AUC הכי טוב ב-test!**

**נשים לב כי בגרפים כאן, הגרף הכתום מתאר את עקומת ה-Curve של ה-training – שלא מעניין אותנו, לעומת הגרף האדום שמתאר את ה-test.**

