

Machine learning

Decision Trees

Lecture III

פיתוח:
ד"ר יהונתן שלר
משה פרידמן

מוטיבציית קורונה ...

- המשק מנסה לחזור בהדרגה לשגרה.
- נניח שיש לנו נתונים על חזרת עובדים לעבודה מחל"ת מסין, ספרד ואיטליה – עם נתונים דמוגרפיים (גיל, מקום מגורים, מגדר, עבודה וכו').
- אנחנו רוצים לבנות מודל עבור האוכלוסייה בישראל – כדי לעזור לנבא האם אדם מסוים יוחזר לעבודה מחל"ת ...

המשחק "נחש מי?"



נניח שאנחנו משחקים
"נחש מי?"

כיצד נוכל לבנות
אסטרטגיה מנצחת?

המשחק "נחש מי?" – בניית אסטרטגיה



❖ נוכל לבנות מעין גרף

❖ בכל שלב נבחר שאלה

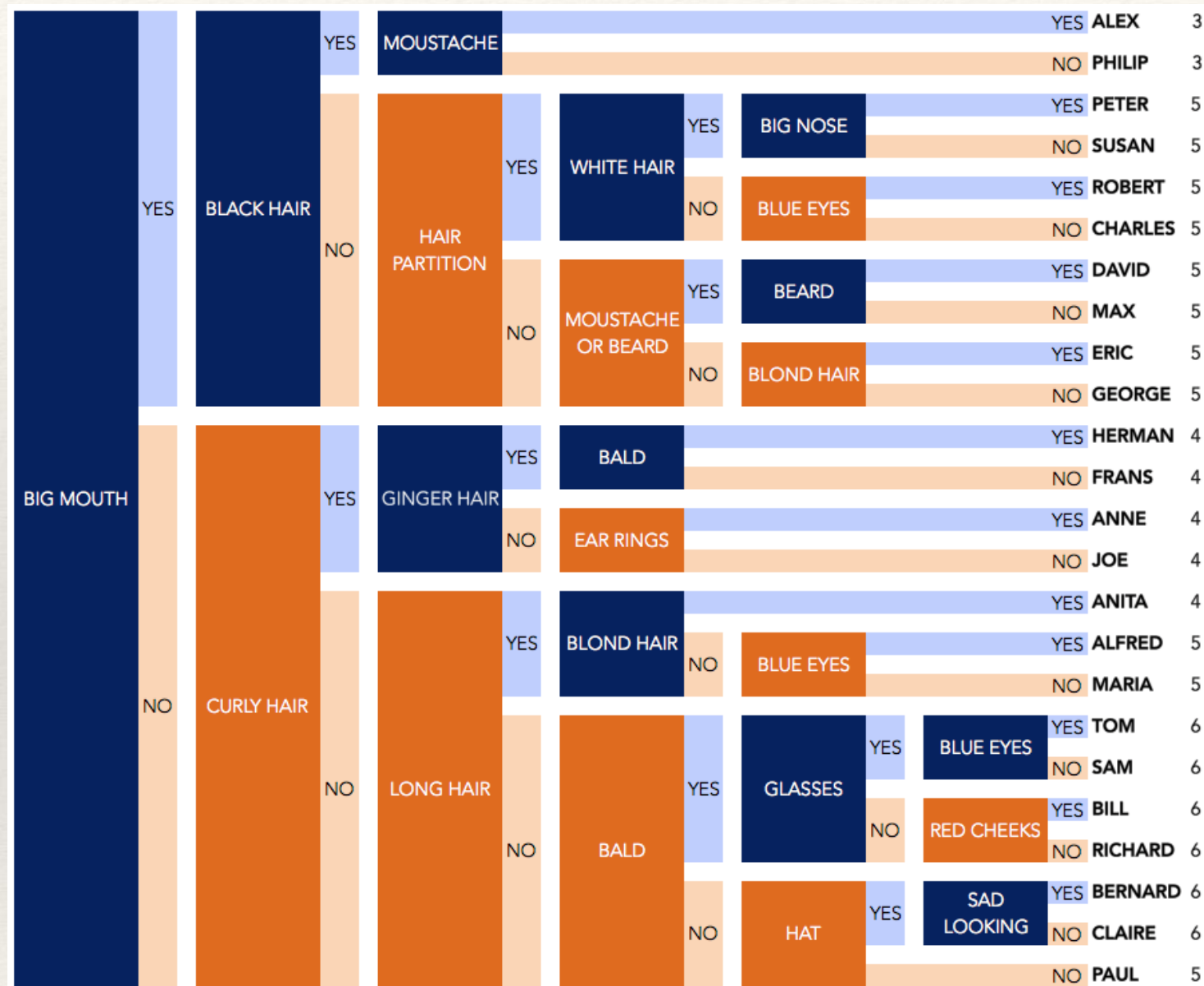
❖ עד שנגיע לתוצאה



המשחק "נחש מי?" – בניית אסטרטגיה

איך נבחר את השאלות הנכונות?

❖ נרצה להגיע לתשובה לאחר מספר שאלות מינימלי



עצי החלטה - נושאים חלק א'

❖ מהם עצי החלטה

❖ אלגוריתם בסיסי

❖ Training Set Error

❖ קריאת עץ החלטה

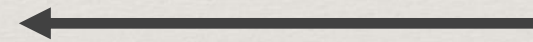
❖ בחירת צומת בעץ

❖ תורת האינפורמציה – אינפורמציה עצמית, אנטרופיה, information gain

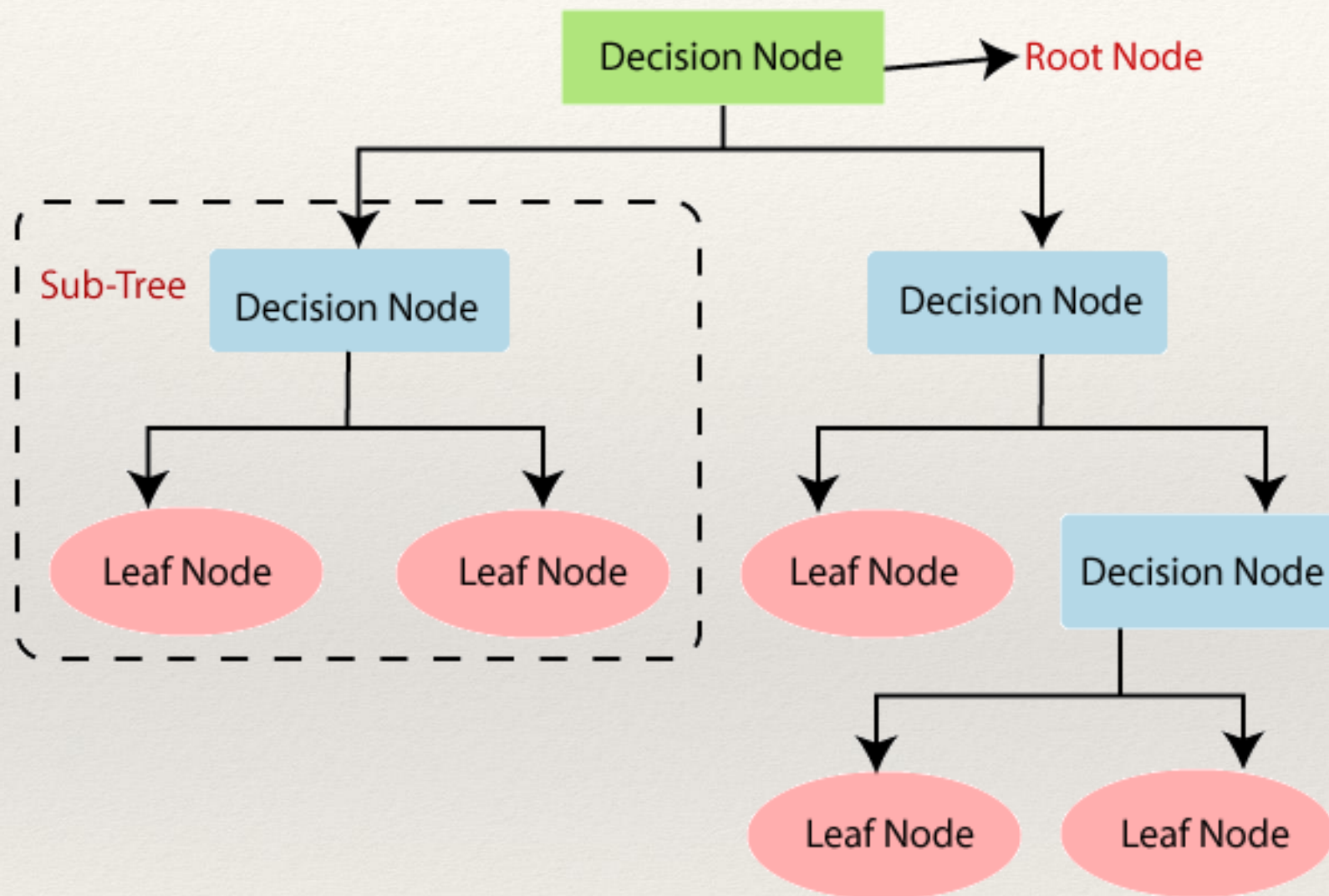
עצי החלטה

עץ החלטה: סדרת
השאלות שמביאה אותנו
להחלטה

מקור השם



עצי החלטה



עץ החלטה: סדרת
השאלות שמביאה אותנו
להחלטה

צומת שורש: השאלה
הראשונה בעץ ההחלטה

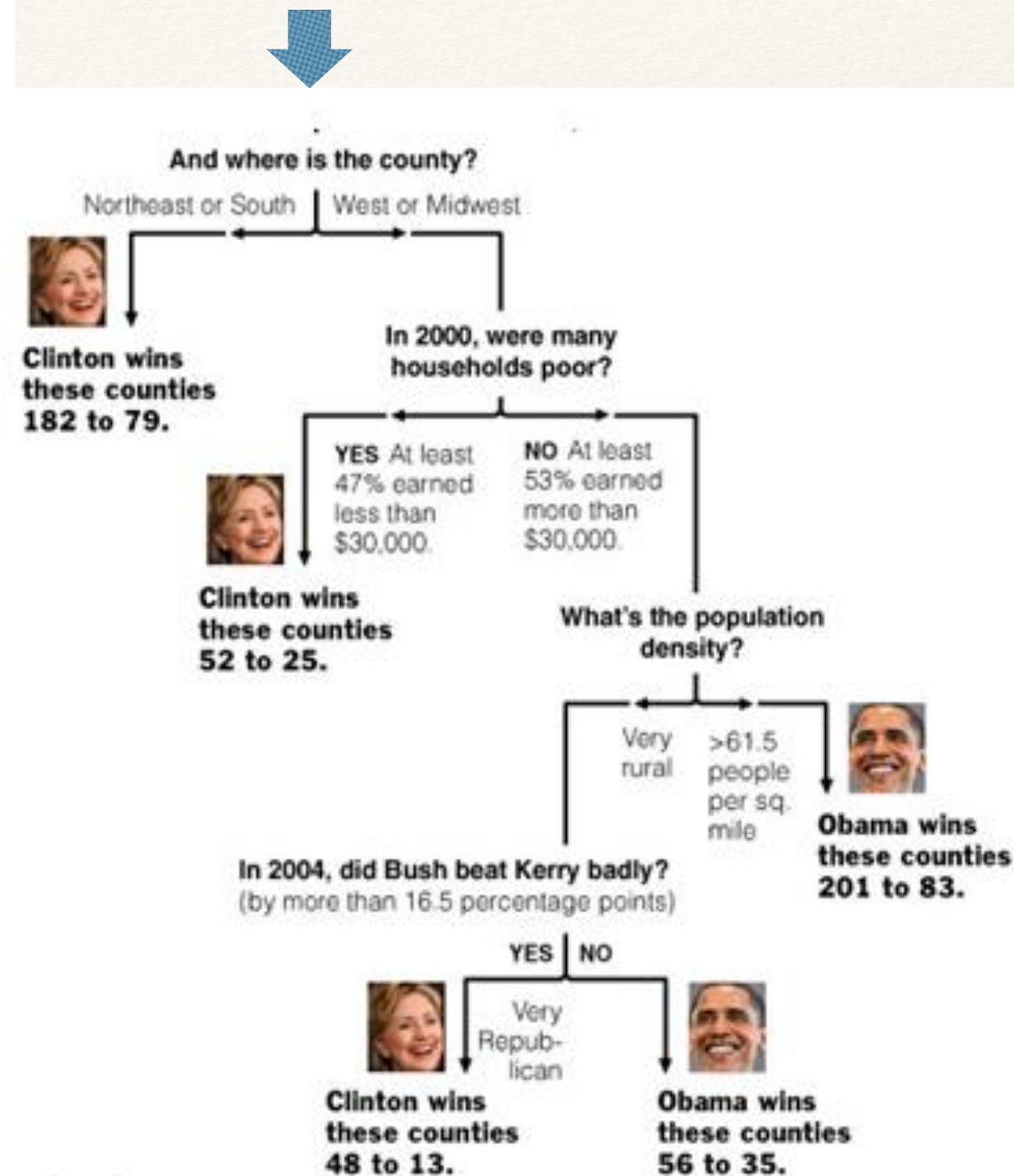
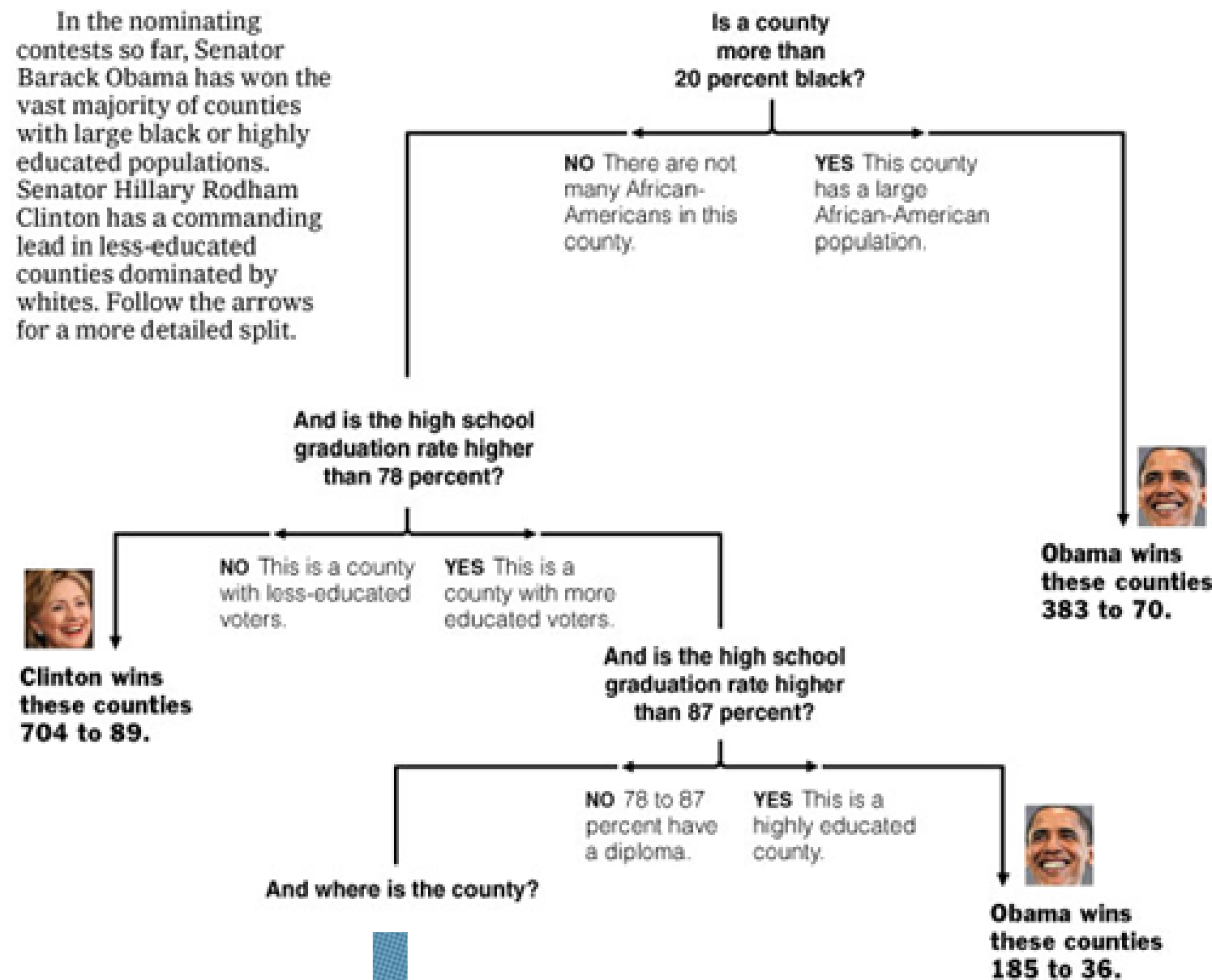
צמתי ביניים: שאלות
המשך

צמתי עלים: ההחלטה
המתקבלת

עצי החלטה - מוטיבציה

Decision Tree: The Obama-Clinton Divide

In the nominating contests so far, Senator Barack Obama has won the vast majority of counties with large black or highly educated populations. Senator Hillary Rodham Clinton has a commanding lead in less-educated counties dominated by whites. Follow the arrows for a more detailed split.



Sources: Election results via The Associated Press; Census Bureau; Dave Leip's Atlas of U.S. Presidential Elections

עצי החלטה - מוטיבציה

❖ מבוסס על סדרה של "מבחנים" לתבנית הקלט, כמו למשל:

❖ "האם התבנית צהובה"

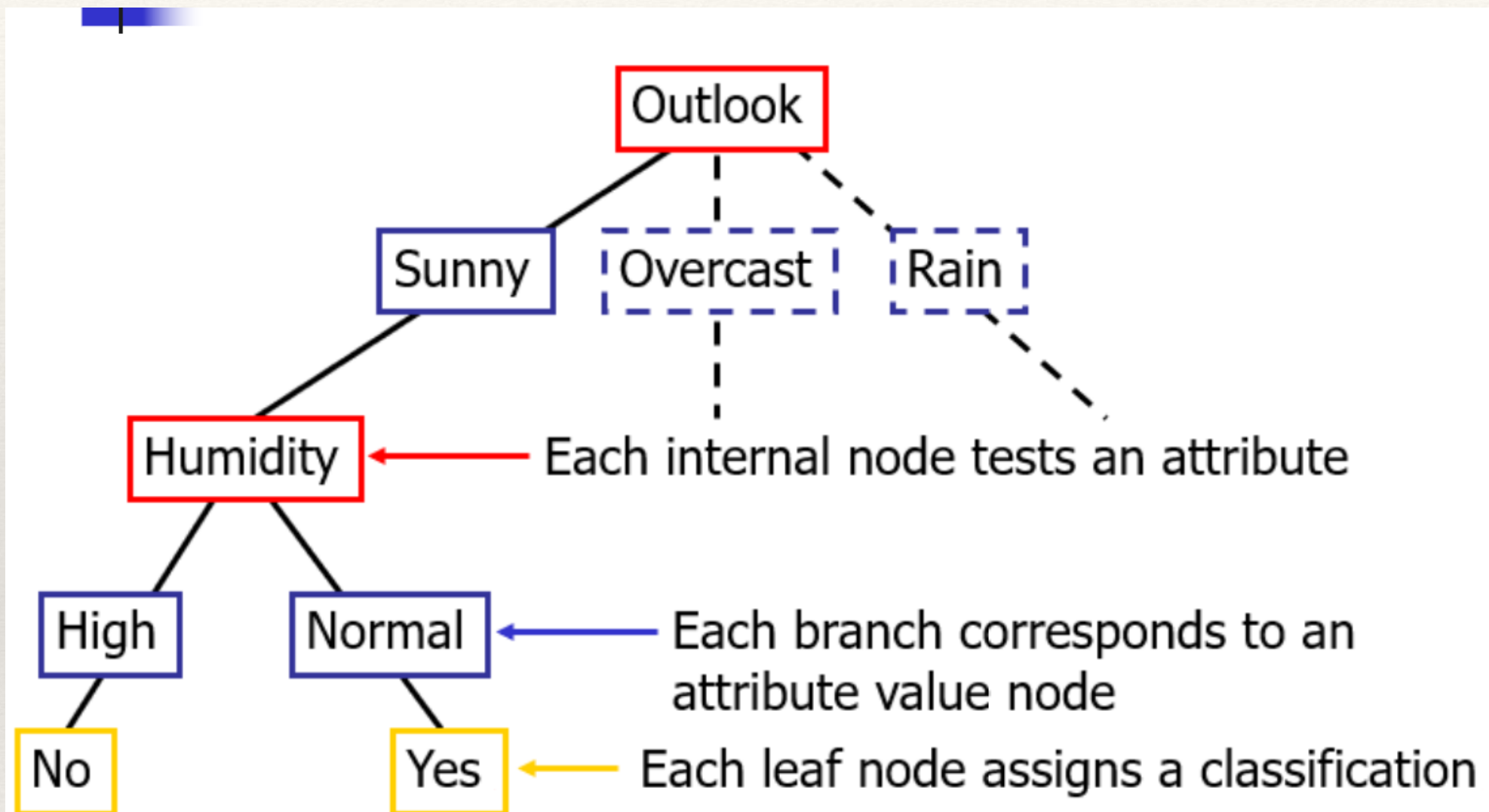
❖ האם המשקל גדול מ-50"

❖

❖ תשובה לסדרת שאלות מסווגת לנו את התבנית.

❖ דוגמא: כתום, עגול, קוטר קטן מ-20 ס"מ ← "תפוז"

עצי החלטה – מוטיבציה והקשר לסיווג



דוגמה: רוצים לחזות
האם מזג האוויר
יאפשר משחק טניס

קודקוד (מסגרת
אדומה) – מאפיין

הקשתות (מסגרת
כחולה) – ערך מאפיין

העלים (מסגרת
צהובה) הסיווג
(לשחק/לא לשחק)

Basic Decision Tree Building Summarized

$\text{BuildTree}(\text{TrainSet}, \text{Output})$

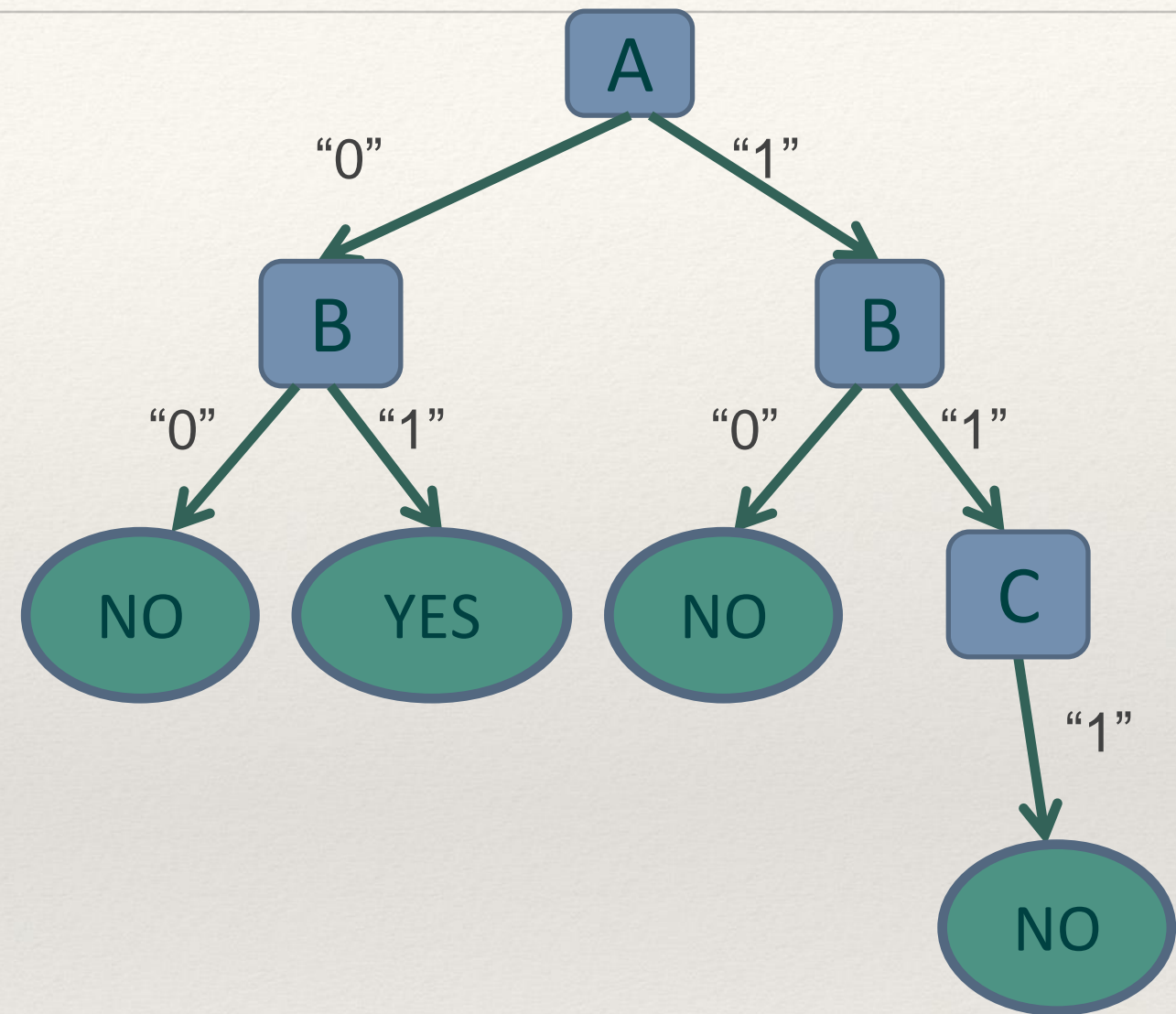
- ❖ Termination Recursion case – choose the category
- ❖ Otherwise (not recursion termination case) – find attribute X with the highest score
- ❖ Suppose X has n_X distinct values (i.e. X has arity n_X).
 - ❖ Create and return a non-leaf node with n_X children.
 - ❖ The i 'th child should be built by calling

$\text{BuildTree}(TS_i, \text{Output})$

Where TS_i built consists of all those records in TrainSet for which $X = i$ th distinct value of X .

דוגמא לסיווג בעזרת עץ החלטה

	A	B	C	Decision
1	0	0	0	"NO"
2	0	1	0	"YES"
3	1	0	1	"NO"
4	1	1	1	"NO"
5	0	0	1	"NO"
6	0	1	1	"YES"
7	1	0	0	"NO"



דוגמא לסיווג בעזרת עץ החלטה

	A	B	C	Decision
1	0	0	0	"NO"
2	0	1	0	"YES"
3	1	0	1	"NO"
4	1	1	1	"NO"
5	0	0	1	"NO"
6	0	1	1	"YES"
7	1	0	0	"NO"

מה עם רשומות חסרות???

איך נסווג את הווקטור $A=1$ $B=1$ $C=0$

בחיים האמיתיים – קבוצת האימון לא תכיל
את כל המצבים האפשריים

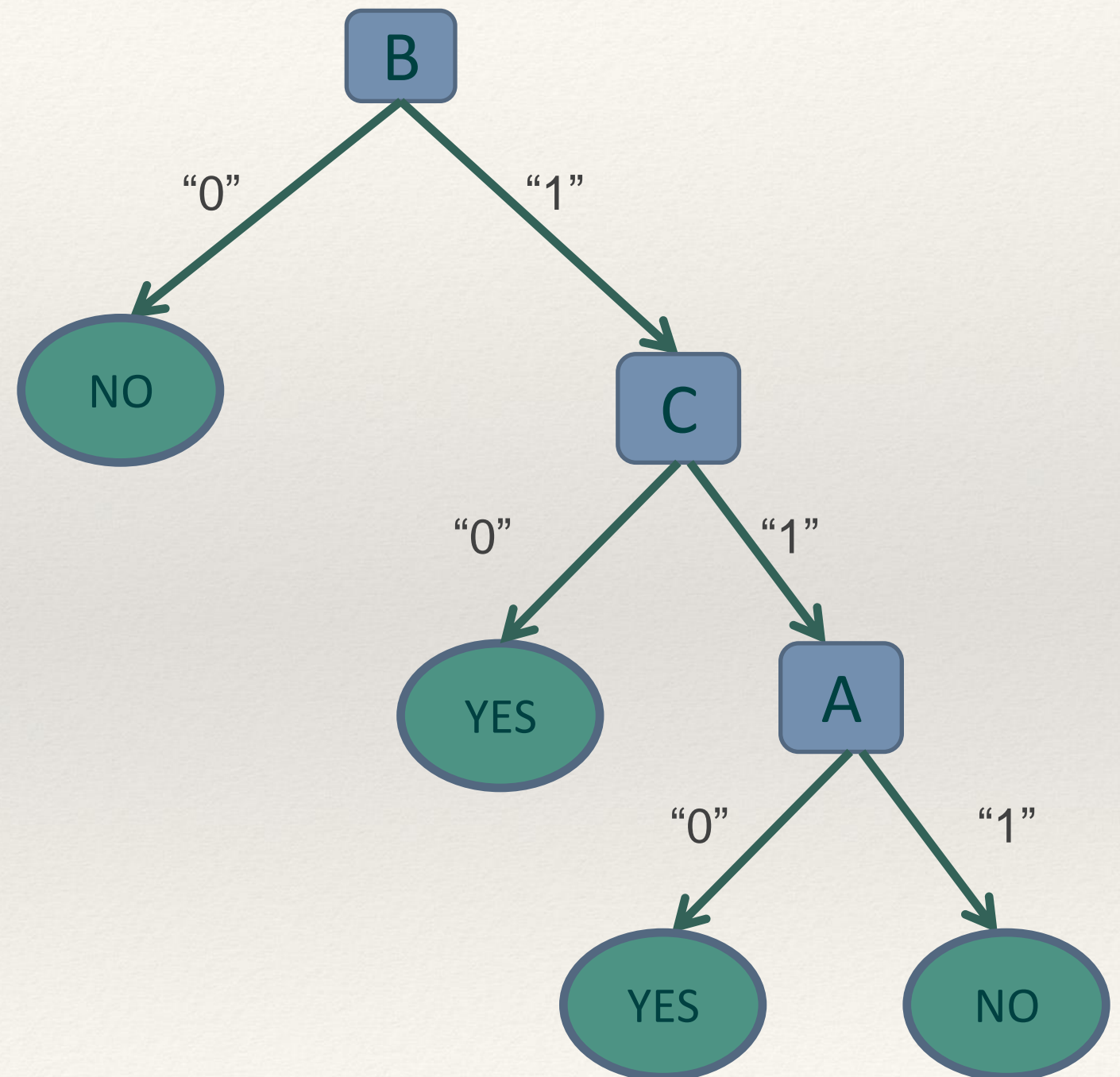
אבל... המטרה שלנו היא כן לדעת לסווג
את כל המצבים האפשריים

בדר"כ נעסוק בריבוי מאפיינים (לא 3 אלא מספר מאות מאפיינים ורוב
המאפיינים אינם בינאריים)

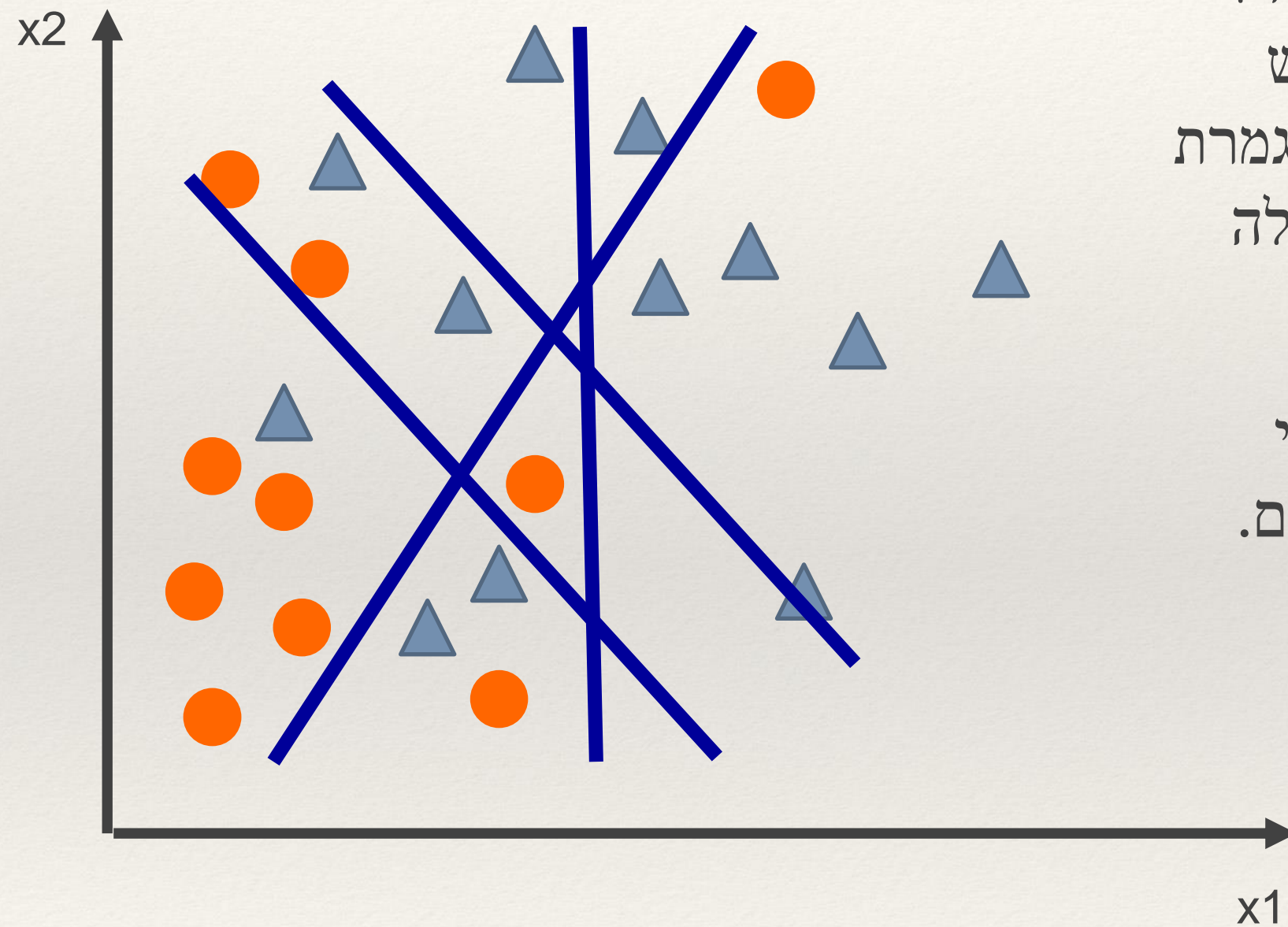
דוגמא לסיווג בעזרת עץ החלטה

אולי ניתן לייצג באמצעות עץ פשוט יותר

	A	B	C	Decision
1	0	0	0	“NO”
2	0	1	0	“YES”
3	1	0	1	“NO”
4	1	1	1	“NO”
5	0	0	1	“NO”
6	0	1	1	“YES”
7	1	0	0	“NO”



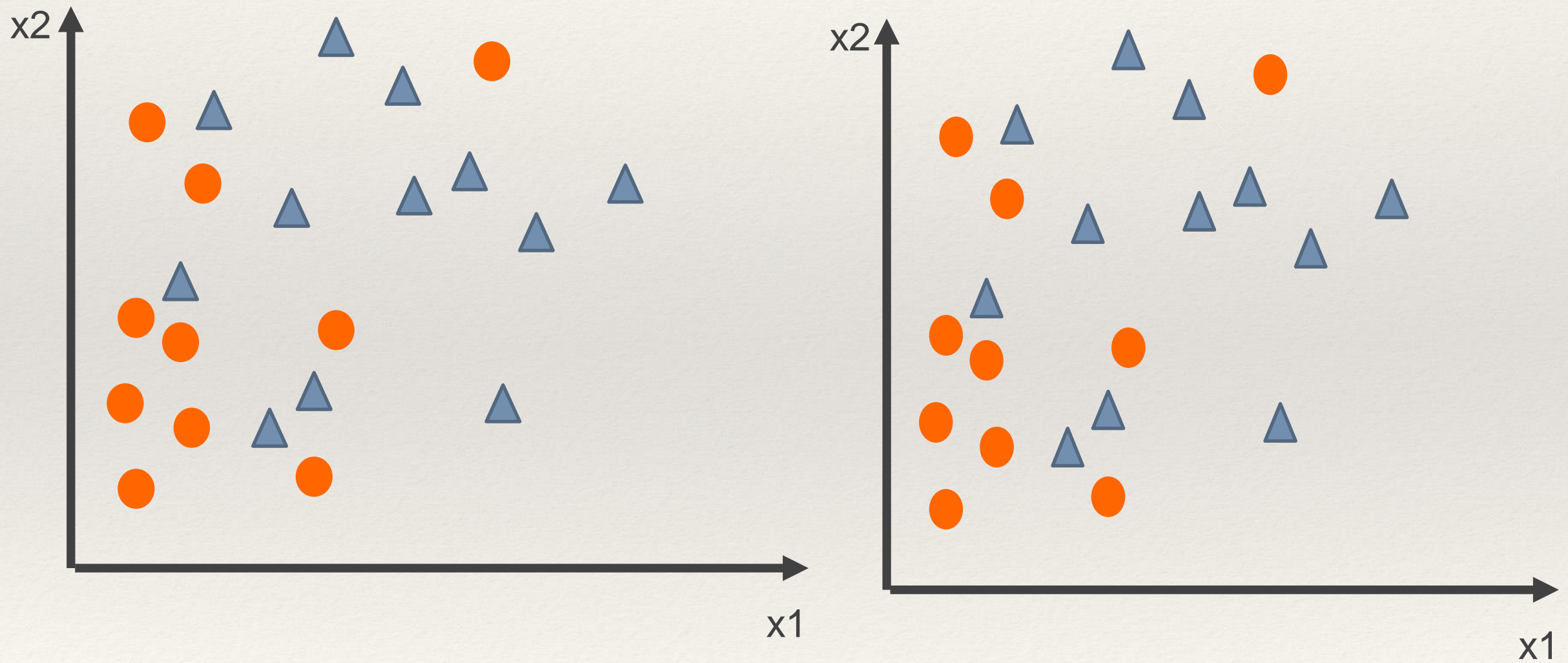
מישור הפרדה



מודל דיסקרמינטיבי - מודל
בו מחפשים הפרדה בה יש
להחליט איפה מתחילה ונגמרת
מחלקה אחת ואיפה מתחילה
ונגמרת מחלקה שניה.

❖ ההחלטה מתבססת לפי
הערכים של המאפיינים.

איזה מישור הפרדה נעדיף



התער של אוקהם Occam's Razor

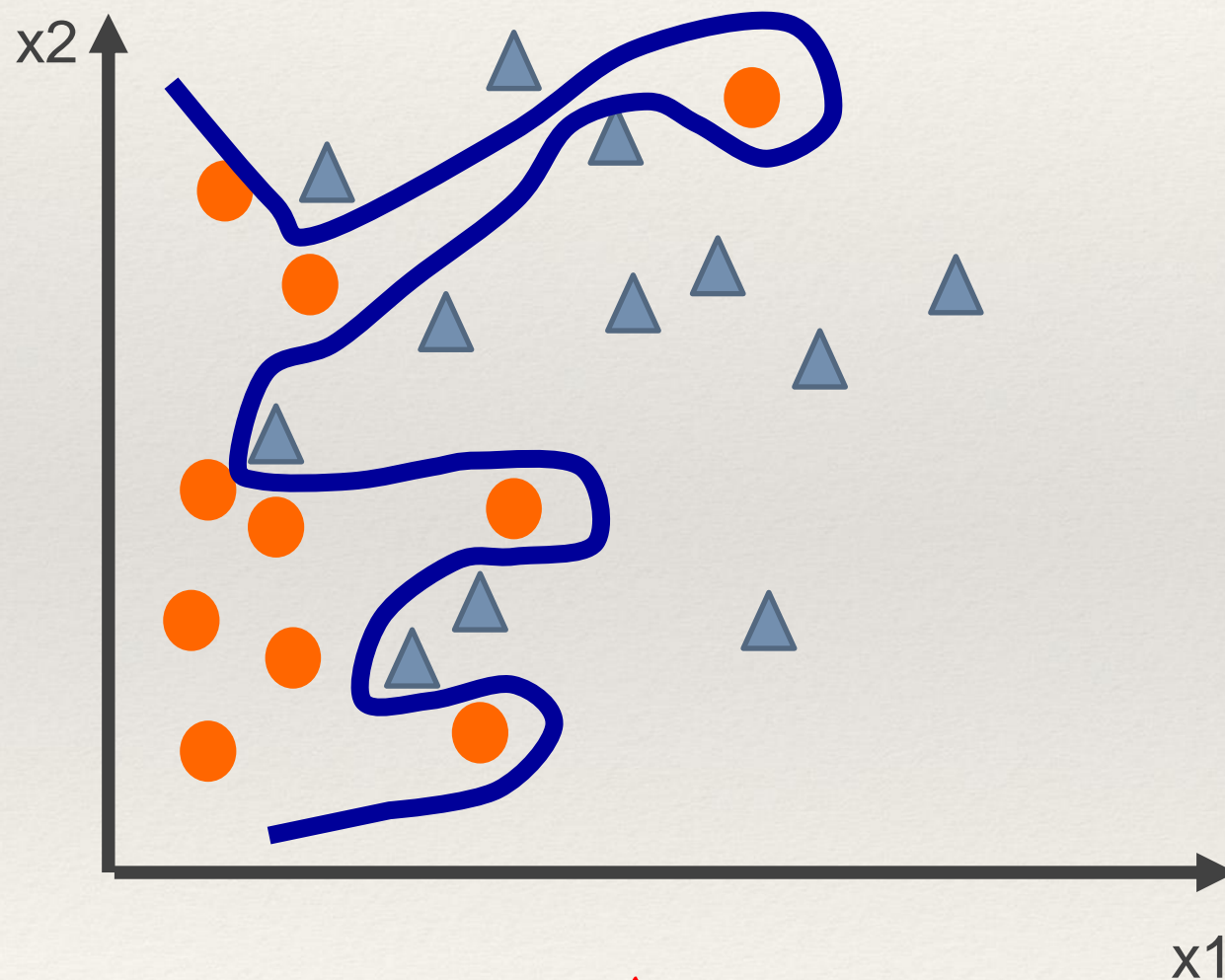
- ❖ עיקרון פילוסופי המיוחס לויליאם איש אוקהם, נזיר פרנציסקני אנגלי בן המאה ה-14.
- ❖ "אין צורך להניח דבר בלא שניתנה לכך סיבה, אלא אם הוא ברור מאליו, או ידוע מהניסיון, או שהוכחתו ניתנת מכוח סמכותם של ספרי הקודש."
- ❖ כאשר קיימים הסברים שונים לאותה תופעה יש לבחור בהסבר הפשוט ביותר, אשר מערב את המספר המועט ביותר של מושגים וחוקים
- ❖ כאשר ישנם שני הסברים שווי ערך לאותה תופעה, יש להעדיף את ההסבר הפשוט או החסכוני יותר

איזה מישור הפרדה נעדיף

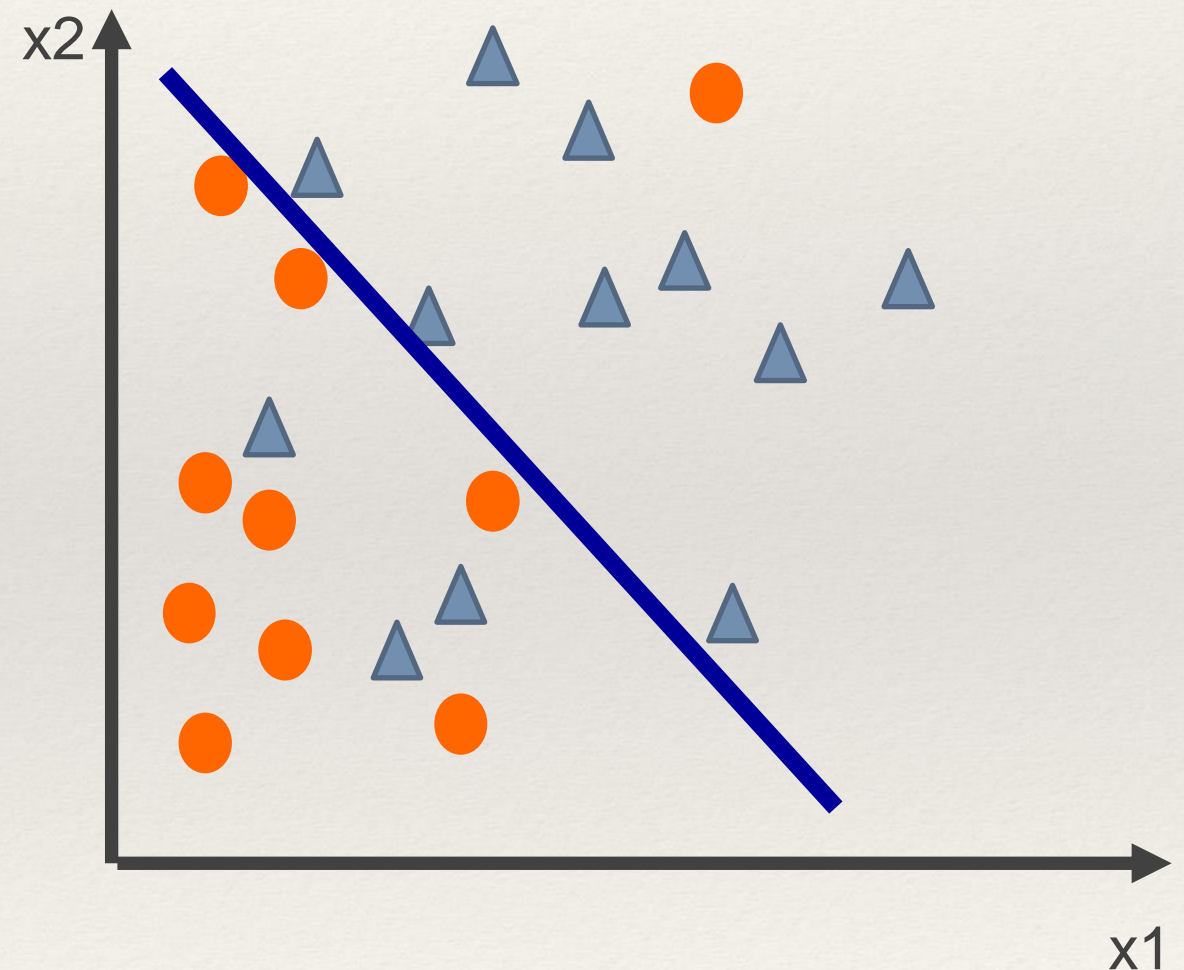
או ההסבר ה"מורכב" ????

את ההסבר ה"פשוט" ????

איך "נתמחר" פשטות מפריד לעומת מס' טעויות

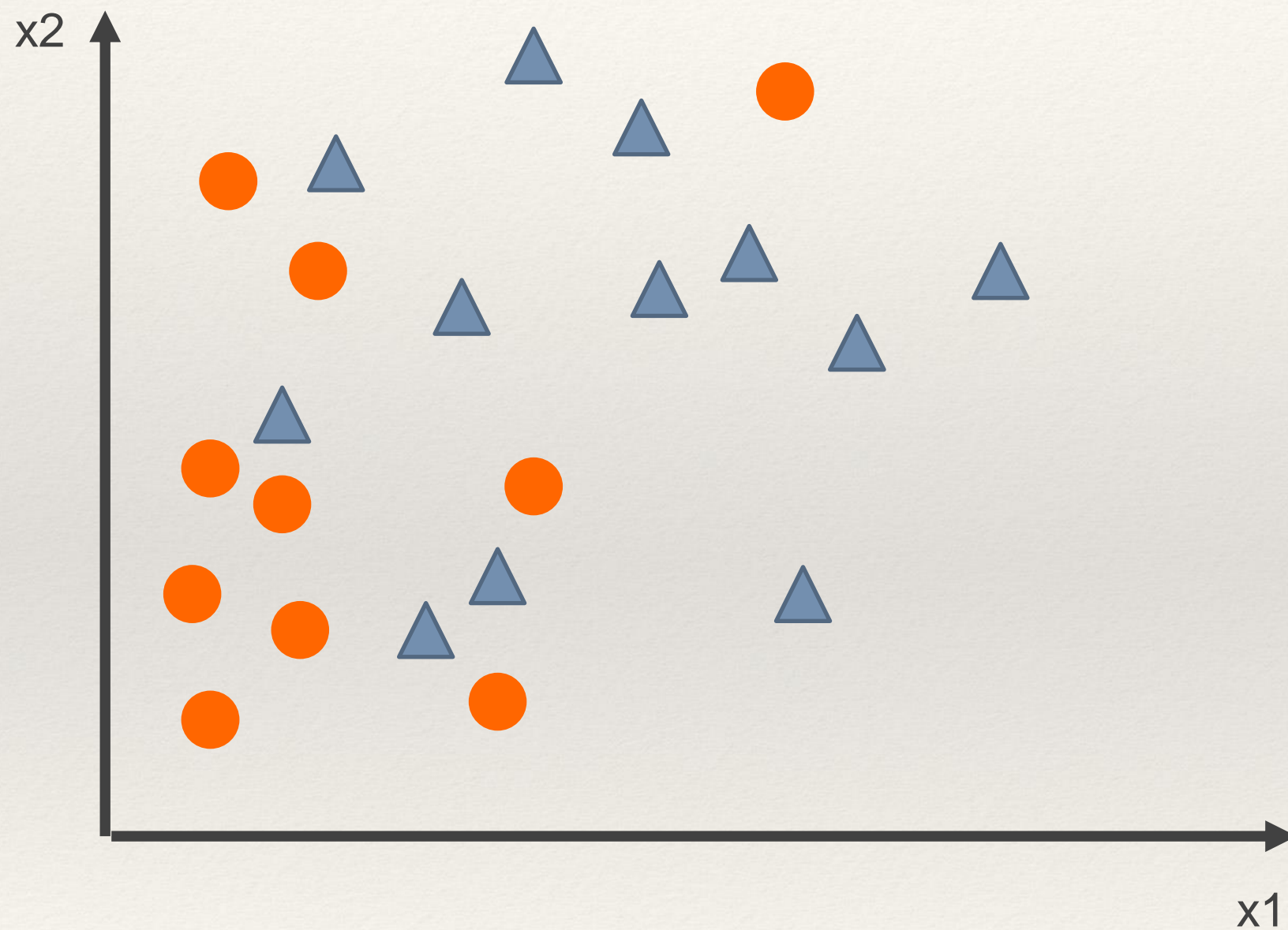


ב



א

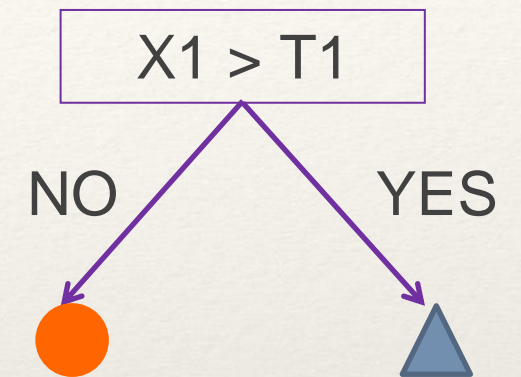
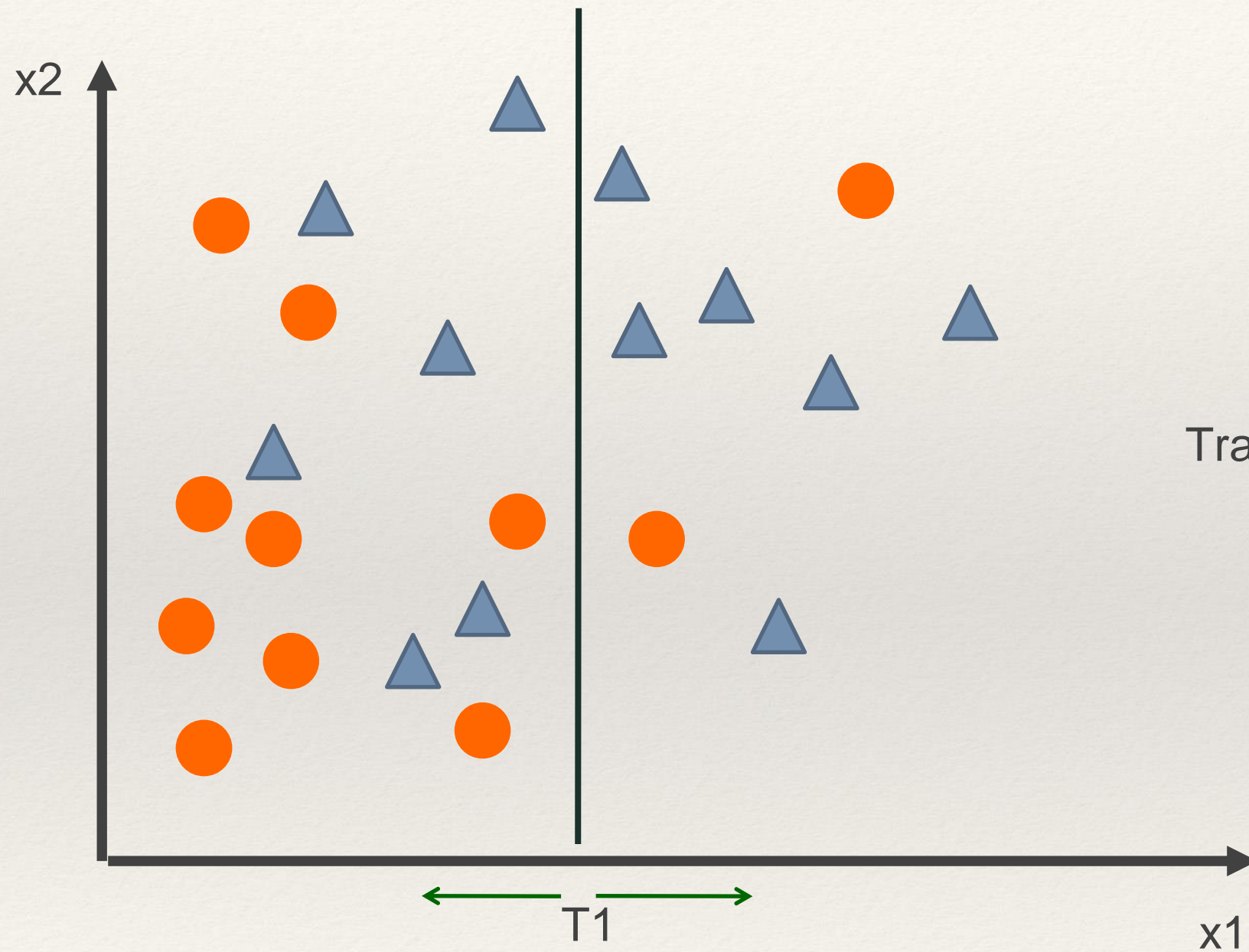
בעיה שרוצים לבנות עבורה עץ החלטה



Training Set Error

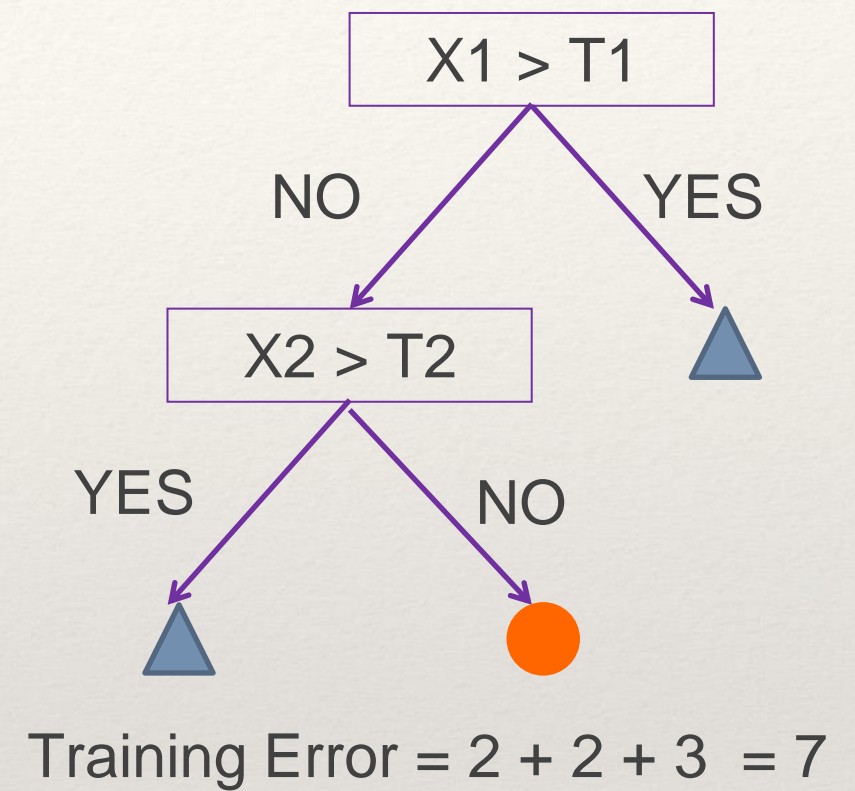
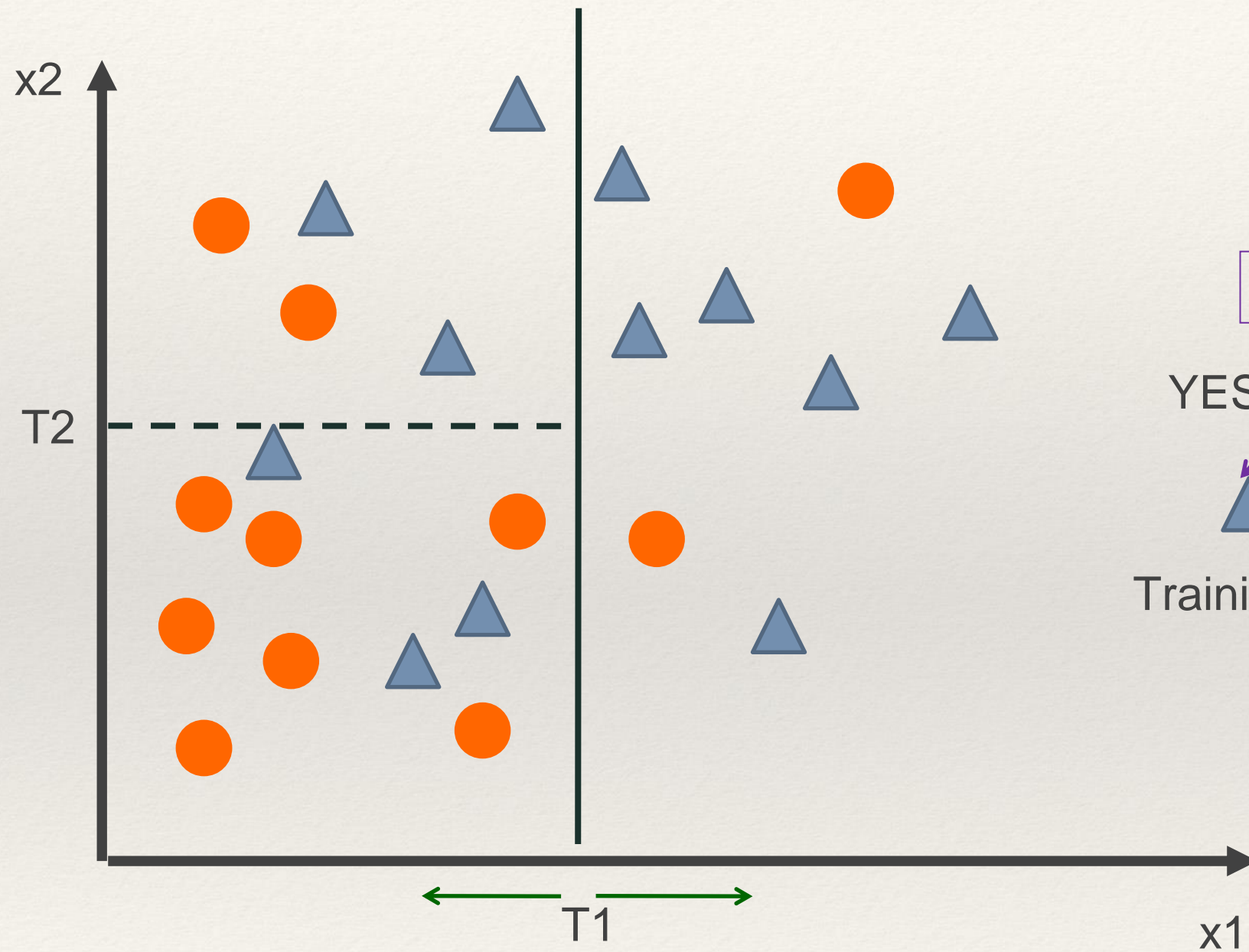
- ❖ עקוב אחר ההחלטה של עץ ההחלטה, עבור כל דוגמה
- ❖ נספור את הדוגמאות עבורם התחזית של עץ ההחלטה אינה תואמת את הערך האמיתי, עבור אותה דוגמה
- ❖ *training set error* – מספר הדוגמאות מה-*train set*, עבור יש חוסר הסכמה.
- ❖ המטרה – ערך נמוך ככל האפשר של ה-*training set error*

דוגמא א - לעץ החלטה

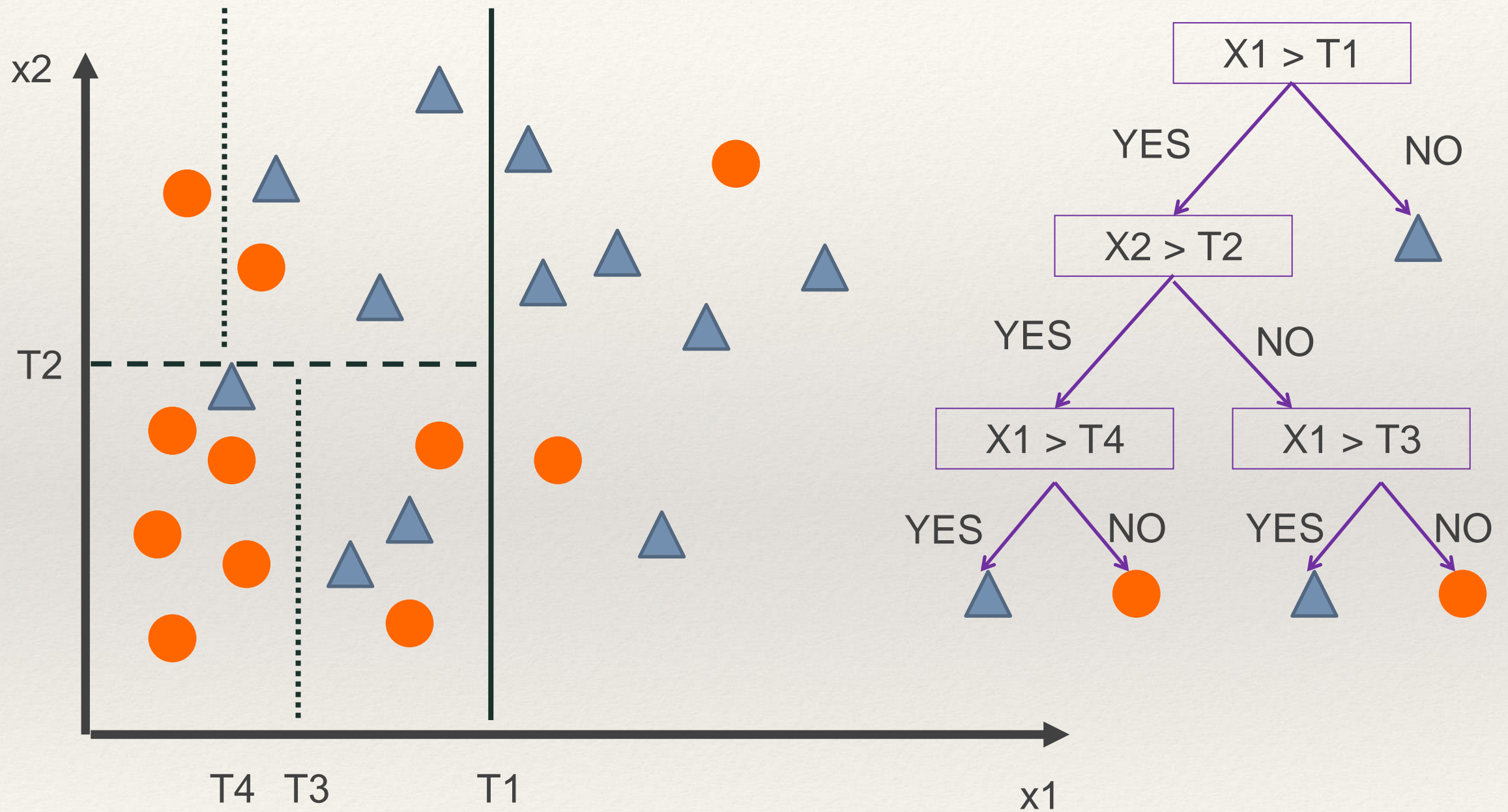


Training Error = 2 + 6 = 8

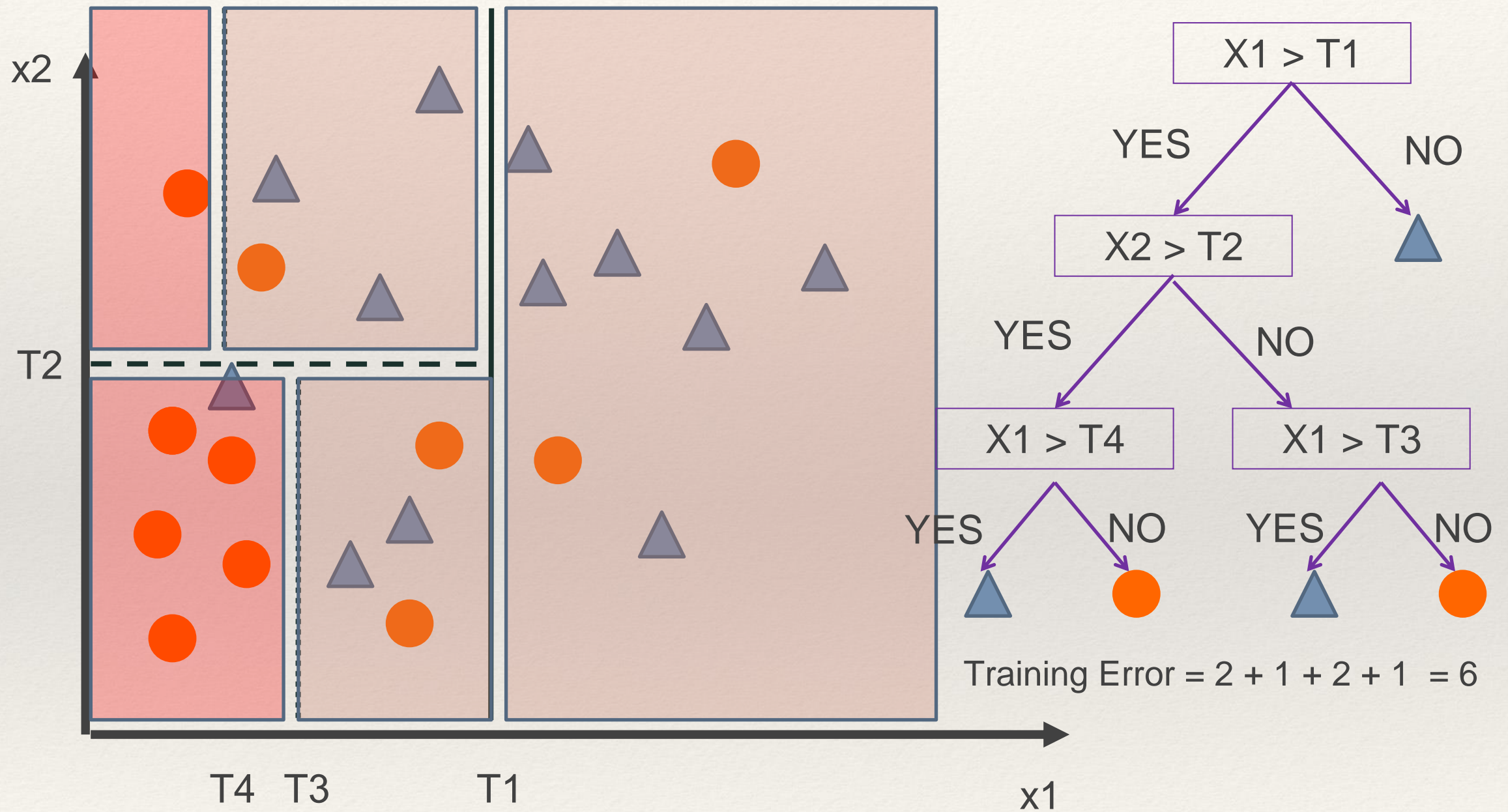
דוגמא לעץ החלטה



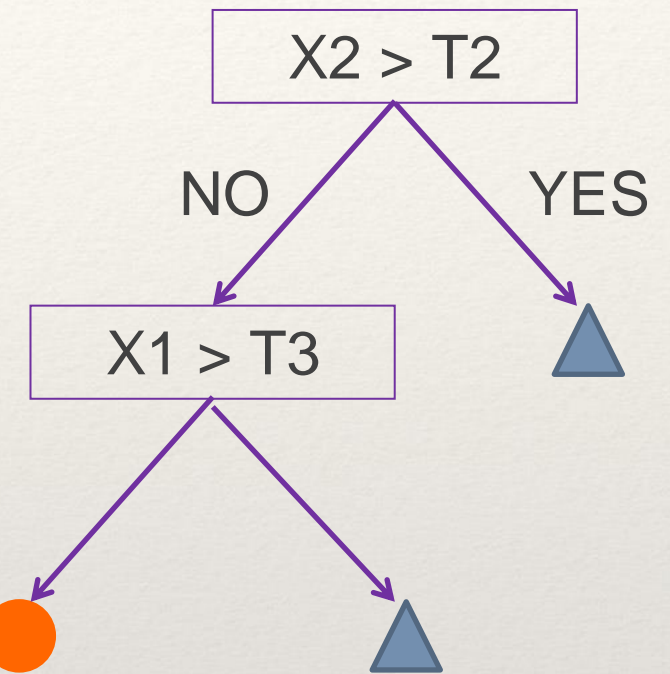
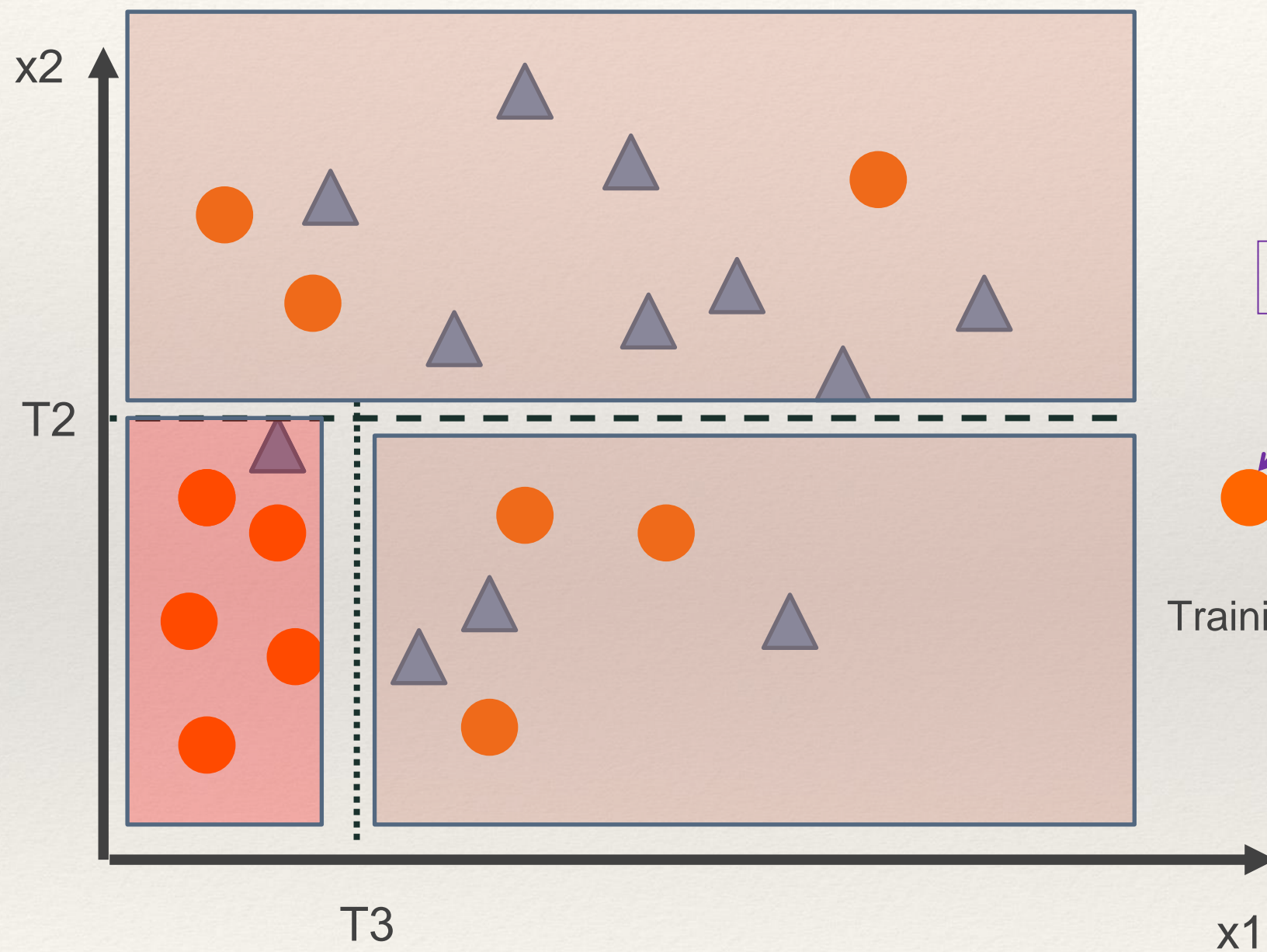
דוגמא לעץ החלטה



דוגמא לעץ החלטה

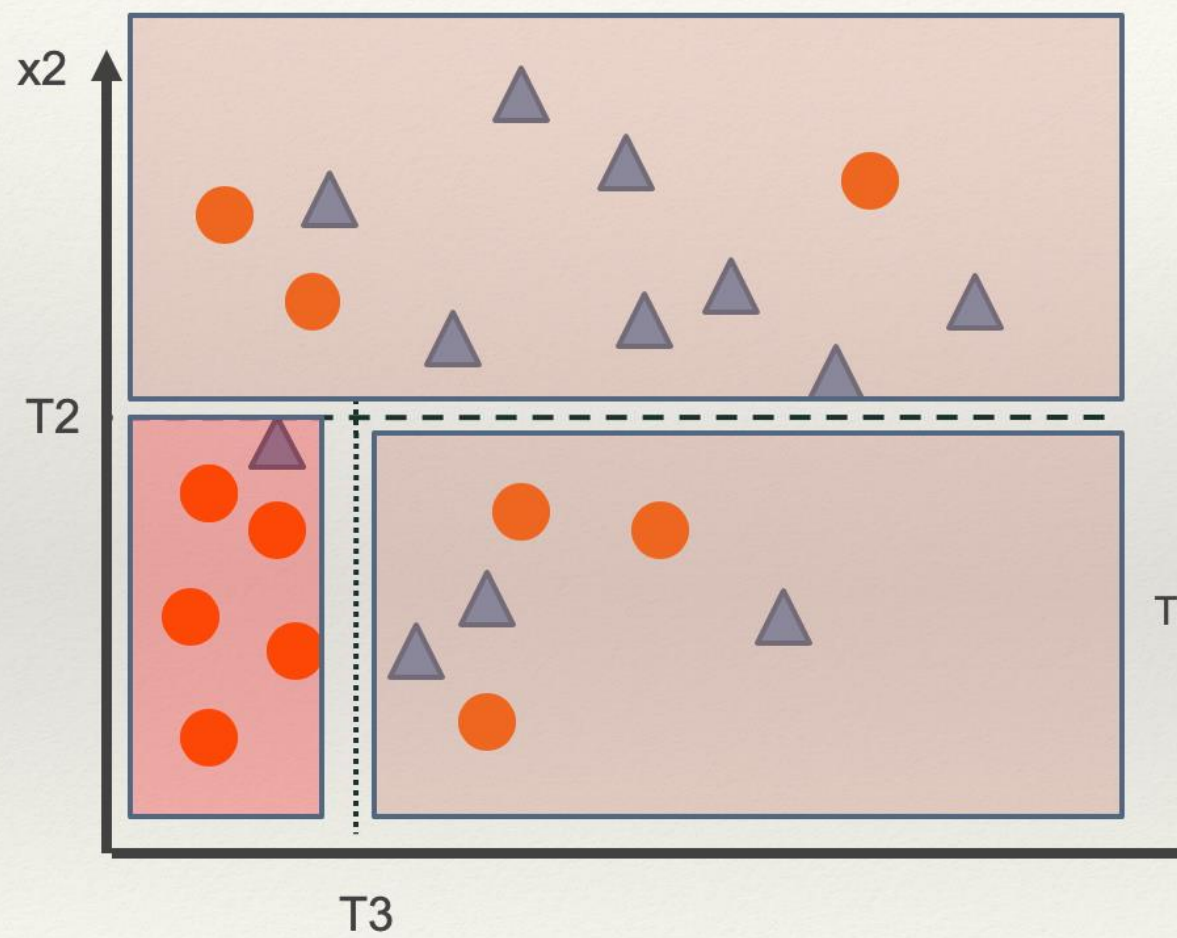


דוגמא ב - לעץ החלטה

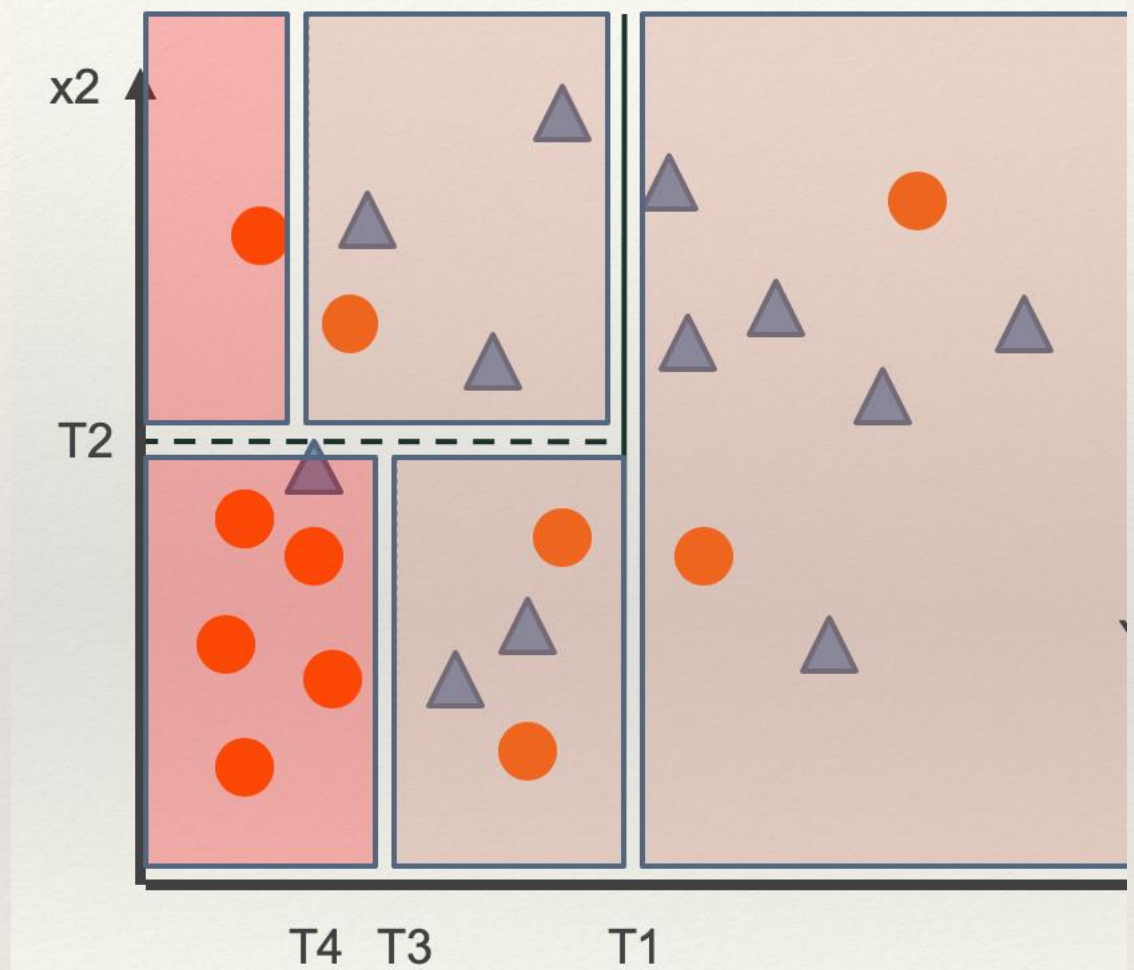


Training Error = 1 + 3 + 3 = 7

מה עדיף?



ב



א

שאלות ביניים

שאלה 1: ממה מורכבים עצי החלטה ומה מייצגים הרכיבים השונים בעץ?

שאלה 2: האם נעדיף הפרדה "מסובכת" או "פשוטה" בין המחלקות?

❖ מה היתרון והחיסרון של על אפשרות?

❖ ומה מנחה אותנו בעצי החלטות?

החלטנו לבנות עץ – מה יהיו השלבים?

- ❖ למצוא מאפיין שמחלק את האוכלוסייה ל-2 קבוצות
 - ❖ למצוא מאפיין חדש לכל קבוצה שמקיים את התנאי הקודם
 - ❖ לסיים – כשכל הקבוצה הומוגנית
-
- ❖ אז מה הופך מאפיין למאפיין טוב בשביל החלוקה?

Good Attribute

❖ אידיאלי כאשר עבור ערך אחד של התכונה נקבל תשובה חיובית ועבור הערך האחר נקבל תשובה שלילית.

Poor Attribute

- ❖ כאשר התכונה לא מייצרת הפרדה בין המחלקות השונות
- ❖ כאשר התכונה לא משנה להחלטה

שאלות פתוחות בשלב זה

איך בונים עץ החלטה? ובפרט

❖ איזו תכונה (attribute) יש לבחור בכל שלב ?

❖ מתי נחליט שצומת מסוים הוא "עלה"

❖ איך מכריעים עבור "עלה"

ID3

- ❖ שיטה אלגוריתמית לבניית עץ החלטה
- ❖ פותחה ע"י J. Ross Quinlan in 1979.
- ❖ שימוש בתורת האינפורמציה (Shanon, 1948)
- ❖ בניית עץ "מלמעלה-למטה" (ללא backtracking)

עץ החלטה: כיצד נבחר את התכונה הבאה

❖ העיקרון הבסיסי

❖ Self information

❖ Entropy

❖ Specific Conditional Entropy

❖ Conditional Entropy

❖ Information Gain

שיטות נוספות: Variance, Gini Index ועוד

עץ החלטה: כיצד נבחר את התכונה הבאה

העיקרון הבסיסי

❖ ננסה לייצר את "המסלול" הקצר ביותר

❖ הרעיון: בכל רמה בעץ ננסה לשאול את השאלה שתשפר לנו בצורה הטובה ביותר את רמת הוודאות בחיזוי

עץ החלטה: כיצד נבחר את התכונה הבאה

שכר	גיל	מגדר	"עובד בקורונה?"
נמוך	30-40	ז	כן
גבוה	40-50	ז	כן
נמוך	30-40	ז	לא
נמוך	50-60	נ	לא
נמוך	50-60	נ	לא
נמוך	30-40	ז	לא
גבוה	30-40	נ	כן

❖ ננסה תמיד לבצע שאילתא על התכונה שמביאה אותנו למקסימום ודאות (מינימום אי-ודאות)

❖ נשים לב שאם נבצע שאילתא על גובה השכר נקבל רמת ודאות גבוהה:

❖ שכר = גבוה \leftarrow עובד בקורונה = כן

❖ שכר = נמוך \leftarrow עובד בקורונה = לא

רמת וודאות

❖ נעדיף תמיד לחזות במקום שבו יש לנו את רמת הוודאות הגבוהה ביותר

❖ מדד לאי הודאות – "אנטרופיה"

Self Information (אינפורמציה עצמית – מושג מתורת האינפורמציה)

❖ דוגמא – הטלת מטבע

❖ עבור מטבע "לא הוגן" שתמיד נופל על "עץ" – לא נהיה מופתעים כאשר גם בפעם הבאה ייפול על "עץ" – קל לחזות = לא נוספה לנו אינפורמציה.

❖ עבור מטבע "הוגן" שנופל בהסתברות 50% על "עץ" – תמיד נהיה מופתעים מתוצאת הזריקה הבאה – קשה לחזות = נוספה לנו "הרבה" אינפורמציה.

Self Information (אינפורמציה עצמית – מושג מתורת האינפורמציה)

❖ נחפש פונקציה שמקיימת את התכונות הבאות:

❖ ככל שהסתברות לאירוע גדלה, האינפורמציה קטנה

❖ נרצה שהאינפורמציה עבור X מאורעות בלתי תלויים תהיה שווה לסכום האינפורמציות עבור כל מאורע בנפרד.

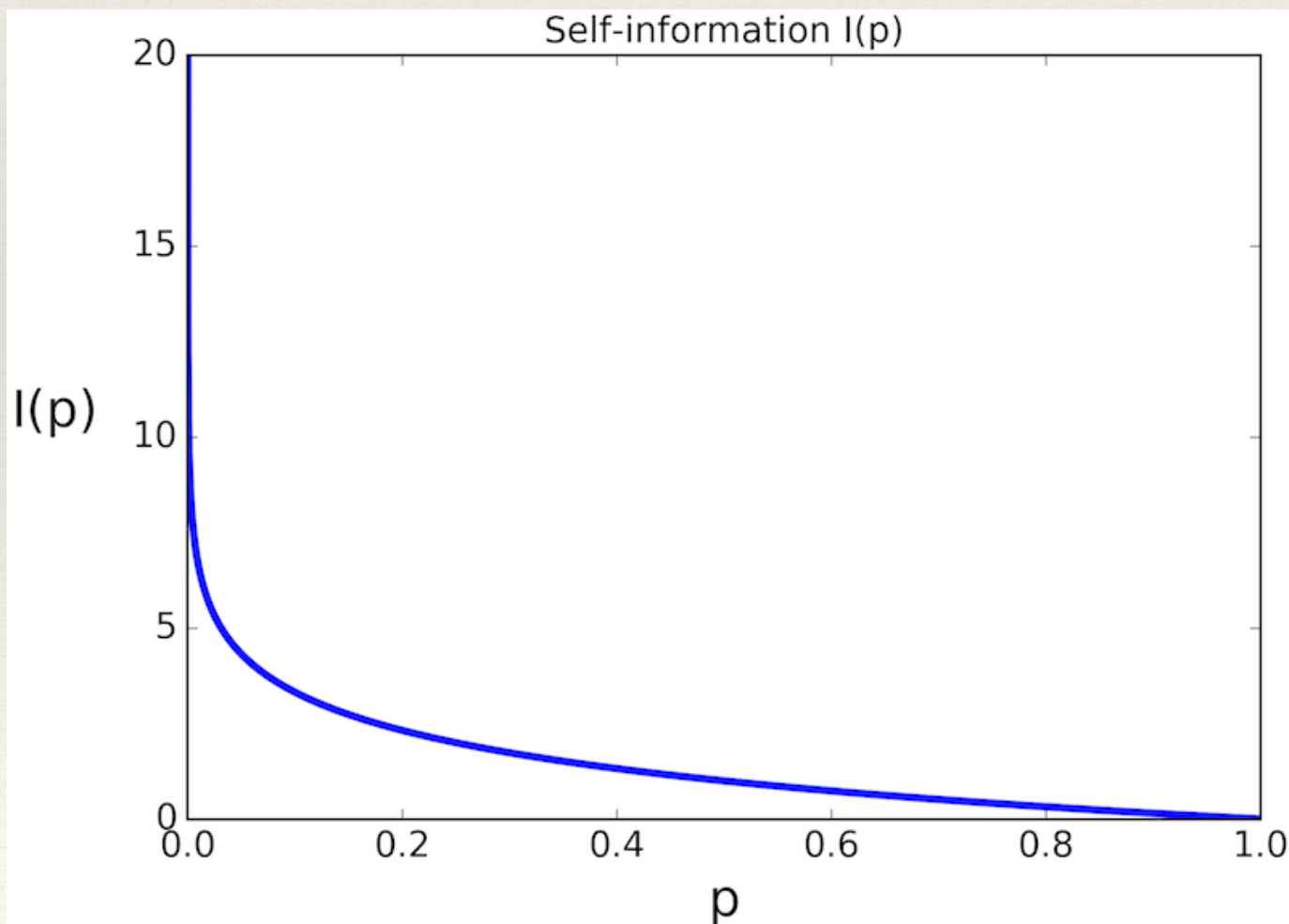
❖ (דוגמא – האינפורמציה של שתי הטלות מטבע תהיה שווה לסכום האינפורמציות של כל הטלה

$$I(p_i \cdot p_j) = I(p_i) + I(p_j) \quad \diamond$$

Self Information (אינפורמציה עצמית – מושג מתורת האינפורמציה)

● פונקציה המתאימה לדרישות אלו

$$I(p_i) = -\log(p_i)$$



Entropy (אנטרופיה)

Shannon entropy is the average self-information (expected value) over all possible values of X

$$H(X) = E_{x \sim P}[I(x)] = \sum_{i=1}^n p(x_i) I(x_i) = - \sum_{i=1}^n p(x_i) \log_b(p(x_i)),$$

Entropy (אנטרופיה)

❖ נתבונן במקרה הכללי בו נתונות לנו הסתברויות:

$$P(X=\alpha_1) = p_1, P(X=\alpha_2) = p_2, \dots P(X=\alpha_n) = p_n$$

נגדיר $H(X)$ כאנטרופיה של X (Entropy)

$$\begin{aligned} H(X) &= -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \dots - p_n \log_2 p_n \\ &= -\sum_{j=1}^n p_j \log_2 p_j \end{aligned}$$

אנטרופיה גבוהה, משמעותה, שההתפלגות יותר דומה לאחידה.

High vs. Low Entropy

❖ "אנטרופיה גבוהה"

❖ התפלגות דומה לאחידה

❖ קשה לחזות

❖ אי וודאות

❖ רמת אי סדר גבוהה

❖ "אנטרופיה נמוכה"

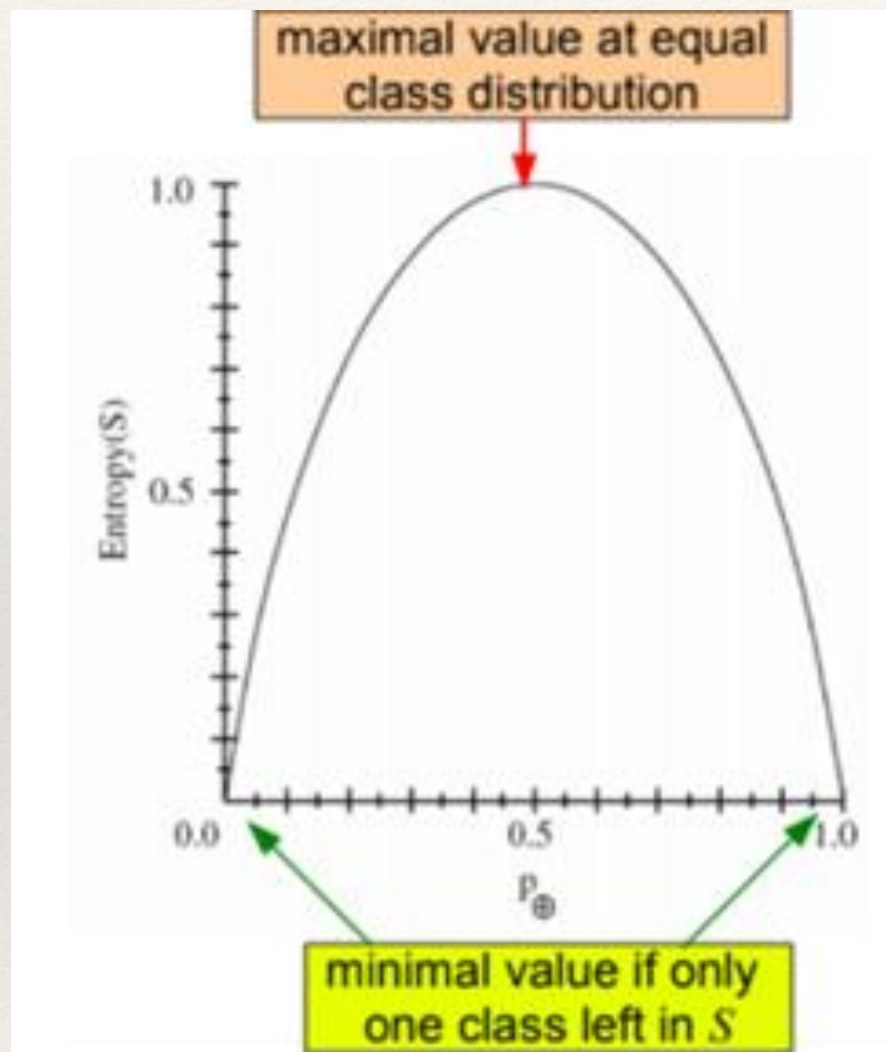
❖ התפלגות מגוונת (בעלת צורה של גבעות ועמקים)

❖ יותר קל לחזות

❖ רמת וודאות גבוהה

❖ רמת אי סדר נמוכה

אנטרופיה עבור 2 מחלקות



❖ S הינה קבוצה של דוגמאות

❖ P^+ הינו החלק היחסי של הדוגמאות החיוביות בקבוצה

❖ P^- הינו החלק היחסי של הדוגמאות השליליות בקבוצה

❖ האנטרופיה של S:

$$H(S) = -p^+ \log_2 p^+ - p^- \log_2 p^-$$

השאיפה: אנטרופיה נמוכה

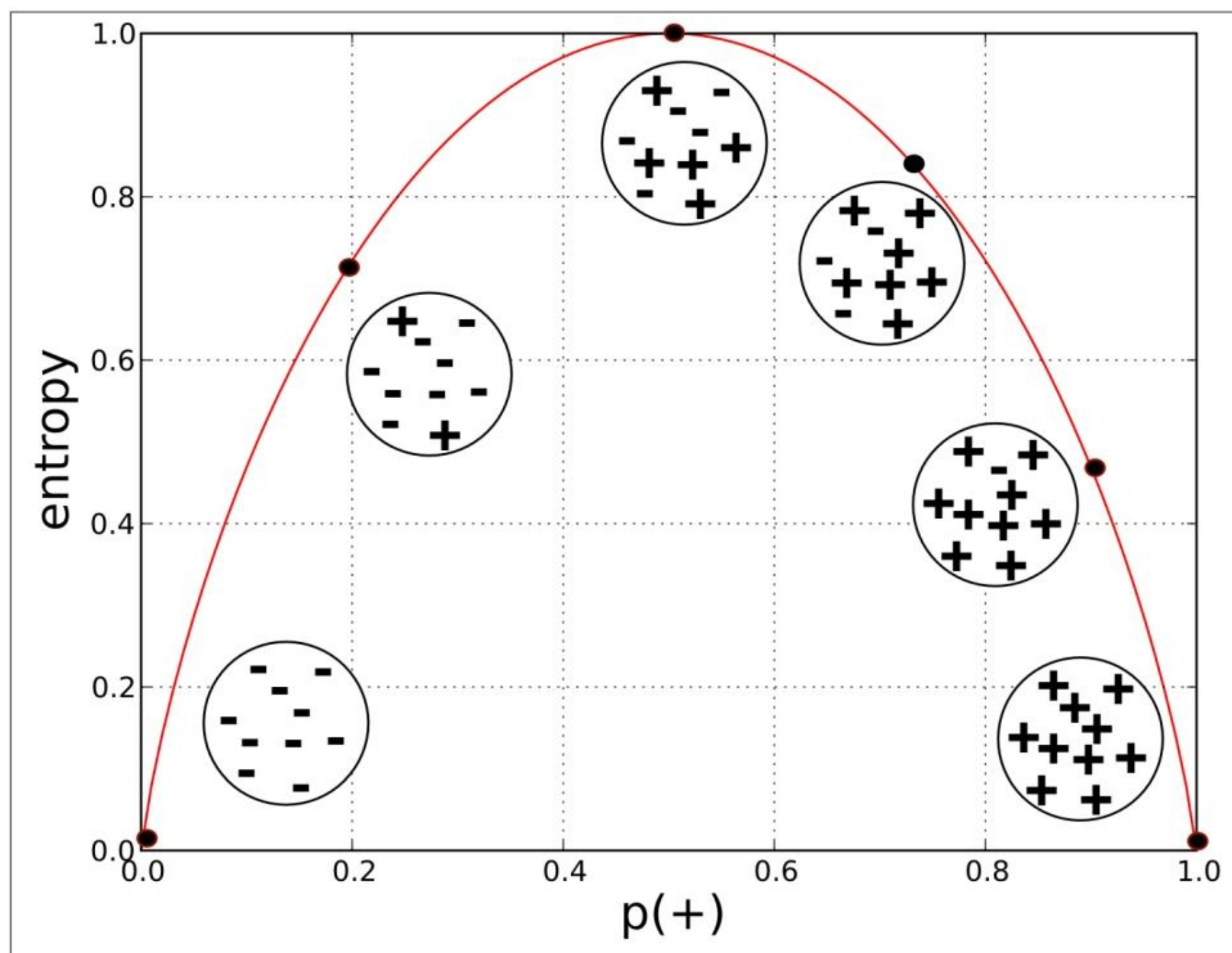


Figure 3-3. Entropy of a two-class set as a function of $p(+)$.

איזו תכונה לבחור בכל שלב

- ❖ בחר את התכונה שממזערת את האנטרופיה בתת-העץ שהיא יוצרת
- ❖ הקשר בין "אי-סדר" ו"אינפורמציה" – ככל שאי הסדר גדל, דרושה יותר אינפורמציה ע"מ לשערך איבר בקבוצה.
- ❖ אנו מחפשים את האסטרטגיה שתוביל אותנו במינימום שאלות אל פתרון הבעיה.



סקר:

כדי למצוא את המאפיין הבא בעץ – מה לדעתכם נחפש:

א. מאפיין שיגדיל את האנטרופיה בכל תת קבוצה של העץ

ב. מאפיין שיקטין את האנטרופיה בכל תת קבוצה של העץ

דוגמא לחישוב אנטרופיה

❖ נתונה הטבלה הבאה:

X	Y
"גיל"	"עובד בקורונה"
צעיר	"כן"
צעיר	"כן"
בינוני	"כן"
צעיר	"לא"
בינוני	"כן"
זקן	"לא"
צעיר	"לא"
זקן	"לא"

$$P(Y = \text{כן}) = 0.5$$

$$P(Y = \text{לא}) = 0.5$$

$$H(Y) = -0.5 \log_2 0.5 - 0.5 \log_2 0.5 = -0.5 \times (-1) - 0.5 \times (-1) = 1$$

$$P(X = \text{צעיר}) = 0.5$$

$$P(X = \text{בינוני}) = 0.25$$

$$P(X = \text{זקן}) = 0.25$$

$$H(X) = -0.5 \log_2 0.5 - 0.25 \log_2 0.25 - 0.25 \log_2 0.25 = -0.5 \times (-1) - 0.25 \times (-2) - 0.25 \times (-2) = 1.5$$

$$H(Y) = 1, \quad H(X) = 1.5$$

אנטרופיה מותנית מסוימת – Specific Conditional Entropy

❖ נגדיר "אנטרופיה מותנית מסוימת"

(Specific Conditional Entropy)

$$H(Y | X = \alpha)$$

כאנטרופיה של Y בין כל אותם הרשומות שבהן
 X מקבל את הערך α

חשבו $H(Y | X = \text{צעיר})$

X "גיל"	Y "עובד בקורונה"
צעיר	"כן"
צעיר	"כן"
בינוני	"כן"
צעיר	"לא"
בינוני	"כן"
זקן	"לא"
צעיר	"לא"
זקן	"לא"

$$P(Y = \text{כן} | X = \text{צעיר}) = 0.5$$

$$P(Y = \text{לא} | X = \text{צעיר}) = 0.5$$

$$H(Y | X = \text{צעיר}) = -0.5 \log 0.5 - 0.5 \log 0.5 = 1$$

אנטרופיה מותנית מסוימת – Specific Conditional Entropy

$$H(Y|X=\alpha)$$

X "גיל"	Y "עובד בקורונה "
צעיר	"כן"
צעיר	"כן"
בינוני	"כן"
צעיר	"לא"
בינוני	"כן"
זקן	"לא"
צעיר	"לא"
זקן	"לא"

$$\begin{aligned} P(Y = \text{כן} | X = \text{בינוני}) &= 1 \\ P(Y = \text{לא} | X = \text{בינוני}) &= 0 \\ H(Y | X = \text{בינוני}) &= -1\log 1 = 0 \end{aligned}$$

$$\begin{aligned} P(Y = \text{כן} | X = \text{זקן}) &= 0 \\ P(Y = \text{לא} | X = \text{זקן}) &= 1 \\ H(Y | X = \text{זקן}) &= -1\log 1 = 0 \end{aligned}$$

$$\begin{aligned} P(Y = \text{כן} | X = \text{צעיר}) &= 0.5 \\ P(Y = \text{לא} | X = \text{צעיר}) &= 0.5 \\ H(Y | X = \text{צעיר}) &= -0.5\log 0.5 - 0.5\log 0.5 = 1 \end{aligned}$$

$$H(Y|X=\text{צעיר}) = 1$$

$$H(Y|X=\text{בינוני}) = 0$$

$$H(Y|X=\text{זקן}) = 0$$

אנטרופיה מותנית – Conditional Entropy

X "גיל"	Y "עובד בקורונה"
צעיר	"כן"
צעיר	"כן"
בינוני	"כן"
צעיר	"לא"
בינוני	"כן"
זקן	"לא"
צעיר	"לא"
זקן	"לא"

האנטרופיה המותנית $H(Y|X)$ הינה ממוצע משוקלל של האנטרופיות ה"ספציפיות" של Y

$$H(Y | X) = \sum_j P(X = \alpha_j) H(Y | X = \alpha_j)$$

$$H(Y|X=\text{צעיר}) = 1 \quad P(X=\text{צעיר}) = 0.5$$

$$H(Y|X=\text{בינוני}) = 0 \quad P(X=\text{בינוני}) = 0.25$$

$$H(Y|X=\text{זקן}) = 0 \quad P(X=\text{זקן}) = 0.25$$

$$H(Y | X) = 0.5 \times 1 + 0.25 \times 0 + 0.25 \times 0 = 0.5$$

סקר:

עבור המאפיין X , המסמן שכר, קיבלנו ש- $H(Y | X) = 0.5$.

נניח שעבור המאפיין X_2 , המסמן מגדר, האנטרופיה המותנית $H(Y | X_2) = 0.6$.

בחרו את הטענה נכונה:

א. נעדיף את מאפיין X בצומת הבא של העץ

ב. נעדיף את מאפיין X_2 בצומת הבא של העץ

ג. לא ניתן להכריע איזה מאפיין עדיף

ד. שתי המאפיינים X ו- X_2 בעלי כדאיות שווה לצומת הבא בעץ, מכיוון שערכם קטן מ-1

סקר:

עבור המאפיין X , המסמן שכר, קיבלנו ש- $H(Y | X) = 0.5$.

נניח שעבור המאפיין X_2 , המסמן מגדר, האנטרופיה המותנית $H(Y | X_2) = 0.6$.

בחרו את הטענה נכונה:

א. נעדיף את מאפיין X בצומת הבא של העץ

ב. נעדיף את מאפיין X_2 בצומת הבא של העץ

ג. לא ניתן להכריע איזה מאפיין עדיף

ד. שתי המאפיינים X ו- X_2 בעלי כדאיות שווה לצומת הבא בעץ, מכיוון שערכם קטן מ-1

Information Gain

Information Gain($Y|X$)

הינה ההפחתה הצפויה באנטרופיה של Y בגלל מיון עפ"י תכונה X

$$Gain(Y|X) = H(Y) - H(Y|X)$$

דוגמה - בחירת התכונה לצומת

- ❖ נתונה קבוצת האימון בה 12 דוגמאות מסווגות כחיובי ו-9 דוגמאות כשלילי.
- ❖ נתונה תכונה A (מייצגת "גיל") לה שני פיצולים אפשריים:
- ❖ "אם גדול מ-30" – הסיווג הוא YES, "אם קטן/שווה ל-30" הסיווג הוא NO.
- ❖ נסמן זאת $(0+, 9-)$ ו- $(12+, 0-)$
- ❖ נתונה תכונה B ("צבע") לה שלושה פיצולים אפשריים:
- ❖ "אם שחור – אנו נשארים עם קבוצה של 2 חיוביים וחמישה שליליים",
- ❖ "אם לבן – נשארים עם קבוצה של 5 חיוביים ואחד שלילי",
- ❖ "אם צהוב – נשארים עם קבוצה של חמישה חיוביים ושלושה שליליים"
- ❖ $(2+, 5-)$, $(5+, 1-)$ ו- $(5+, 3-)$

איזו תכונה עדיפה כצומת הבא בעץ?

דוגמה - בחירת התכונה לצומת – פתרון בעזרת IG

נחשב את האנטרופיה של המחלקה ...

חישבנו:

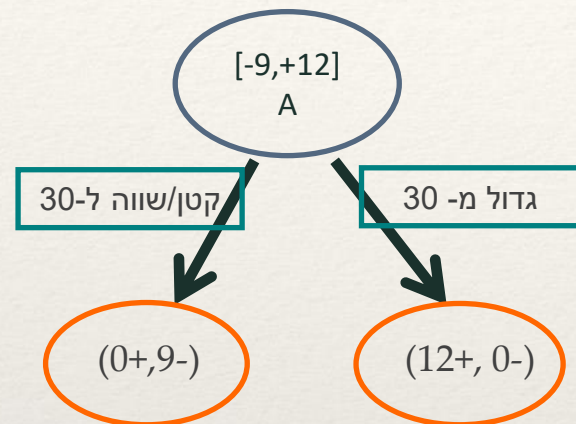
$$p(\text{סיווג} = Yes) \approx 0.57$$

$$p(\text{סיווג} = No) \approx 0.43$$

נחשב את האנטרופיה של המחלקה ...

$$\begin{aligned} H(Y) &\approx -0.57 \cdot \log_2(0.57) - 0.43 \cdot \log_2(0.43) = \\ &= -0.57 \cdot (-0.807) - 0.43 \cdot (-1.22) = 0.4611 + 0.5228 = \mathbf{0.98} \end{aligned}$$

דוגמה - בחירת התכונה לצומת – פתרון בעזרת IG



נחשב את האנטרופיה המותנת $H(Y|B)$...

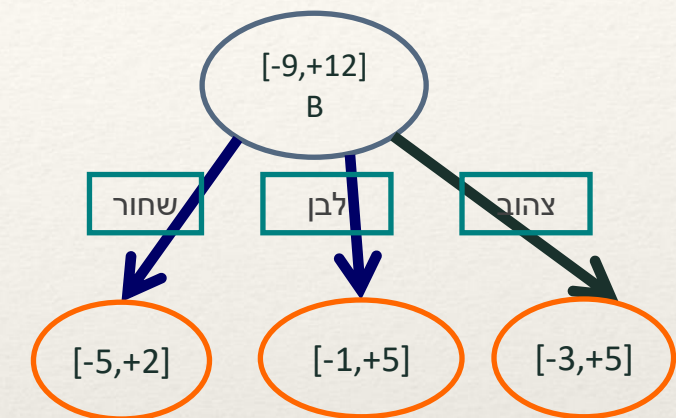
$$H(Y|B = \text{black}) \approx -0.286 \cdot \log_2(0.286) - 0.714 \cdot \log_2(0.714) \approx 1.002$$

$$H(Y|B = \text{white}) \approx -0.833 \cdot \log_2(0.833) - 0.166 \cdot \log_2(0.166) \approx 0.645$$

$$H(Y|B = \text{yellow}) \approx -0.625 \cdot \log_2(0.625) - 0.375 \cdot \log_2(0.375) \approx 0.954$$

$$P(B = \text{black}) = \frac{7}{21} \approx 0.333 \quad P(B = \text{white}) = \frac{6}{21} \approx 0.286 \quad P(B = \text{yellow}) = \frac{8}{21} \approx 0.381$$

$$H(Y|B) \approx 0.333 \cdot 1.002 + 0.286 \cdot 0.645 + 0.381 \cdot 0.954 \approx 0.882$$



נחשב את האנטרופיה המותנת $H(Y|A)$...

$$H(Y|A > 30) = -1 \cdot \log_2(1) - 0 \cdot \log_2(0) = 0$$

$$H(Y|A \leq 30) = -0 \cdot \log_2(0) - 1 \cdot \log_2(1) = 0$$

$$H(Y|A) = p(A \leq 30) \cdot 0 + p(A > 30) \cdot 0 = 0$$

דוגמה - בחירת התכונה לצומת – פתרון בעזרת IG

חישובנו:

$$H(Y|A) = 0$$

$$H(Y|B) \approx 0.882$$

$$H(Y) = 0.98$$

נחשב את האנטרופיה המותנת $IG(Y|B)$...

$$IG(Y|B) = H(Y) - H(Y|B) \approx 0.98 - 0.882 \\ = \mathbf{0.098}$$

נחשב את האנטרופיה המותנת $IG(Y|A)$...

$$IG(Y|A) = H(Y) - H(Y|A) = \mathbf{0.98}$$

ולכן נבחר ב-A כתכונה בצומת הבא

שיטות נוספות - GINI INDEX

- Instead of entropy, impurity can be measured by the Gini index

$$Gini(S) = 1 - \sum_i p_i^2$$

- error probability

$$E[1 - p] = \sum_i p_i(1 - p_i) = \sum_i p_i - \sum_i p_i^2 = 1 - \sum_i p_i^2$$

שיטות נוספות - GINI INDEX

- Instead of entropy, impurity can be measured by the Gini index

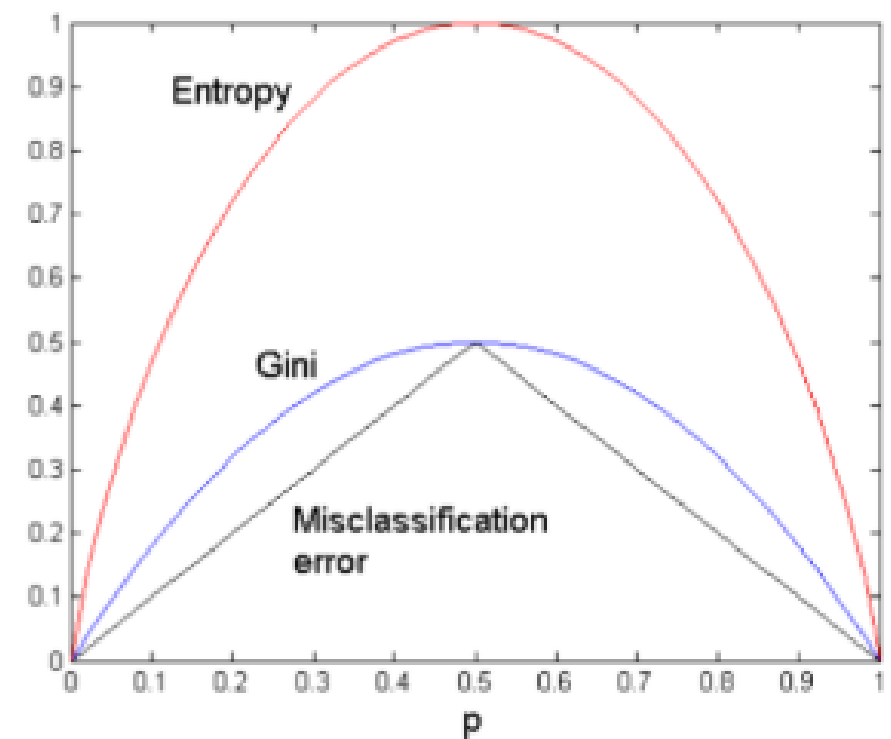
$$Gini(S) = 1 - \sum_i p_i^2$$

- average Gini index

$$Gini(S, A) = \sum_i \frac{|S_i|}{|S|} \cdot Gini(S_i)$$

- Popular for regression trees

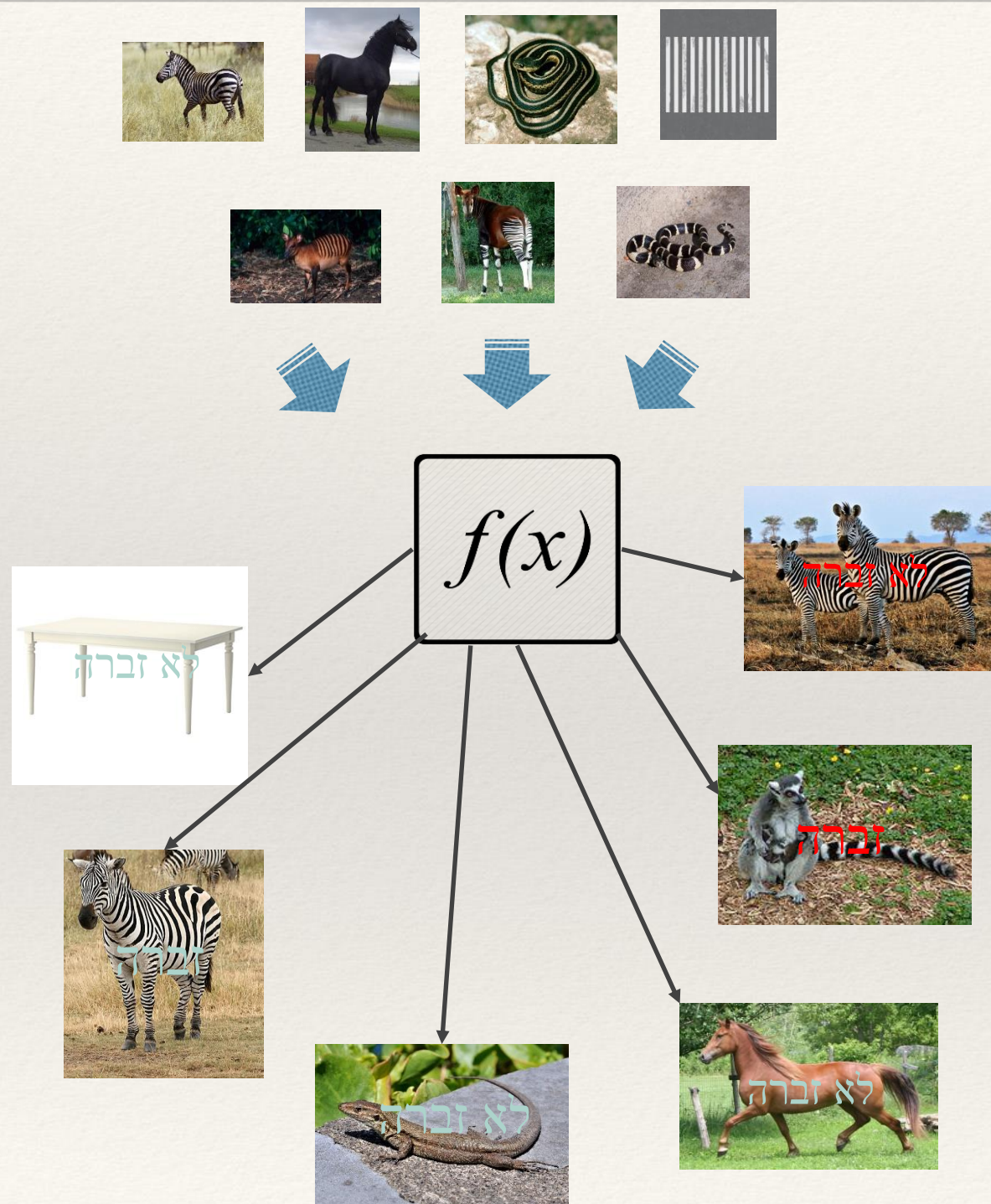
For a 2-class problem:



שערוך המודל (evaluation)



שערוך המודל (evaluation)



שערוך המודל (evaluation):

- שערוך המודל מבוסס על בחינתם של תוצאות הניבוי של המודל
- ההצלחה או אי הצלחה, נאמדת (estimated) על דוגמאות שהמודל לא ראה

שערוך המודל (evaluation)

Confusion matrix:

	Predicted Yes	Predicted No
Actual Yes	True Positive (TP)	False Negative (FN)
Actual No	False Positive (FP)	True Negative (TN)

$$\text{accuracy} = \frac{\#correct\ predictions = \#TP + \#TN}{\#test\ instances = \#TP + \#TN + \#FP + \#FN}$$

$$\text{Error (rate)} = 1 - \text{accuracy} = \frac{\#incorrect\ predictions = \#FP + \#FN}{\#test\ instances = \#TP + \#TN + \#FP + \#FN}$$

שערוך המודל – דוגמה

בדיקה חדשה לגילוי קורונה נוסתה על 500 איש. מתוכם 400 בריאים ו 100 חולים – והתקבלו הנתונים הבאים:

$$\text{accuracy} = \frac{\#correct\ predictions = \#TP + \#TN}{\#test\ instances = \#TP + \#TN + \#FP + \#FN}$$

Calculate the accuracy ...

	סוג בחולה	סוג כלא חולה
חולה בפועל	95	5
לא חולה בפועל	15	385

שערוך המודל – דוגמה

בדיקה חדשה לגילוי קורונה נוסתה על 500 איש. מתוכם 400 בריאים ו 100 חולים – והתקבלו הנתונים הבאים:

$$\text{accuracy} = \frac{\#correct\ predictions = \#TP + \#TN}{\#test\ instances = \#TP + \#TN + \#FP + \#FN}$$

Calculate the accuracy ...

	סוג בחולה	סוג כלא חולה
חולה בפועל	95	5
לא חולה בפועל	15	385

Accuracy =

$$\frac{(95+385)}{(95+5+15+385)} = \frac{480}{(500)} \sim 0.96$$