

*Machine learning*

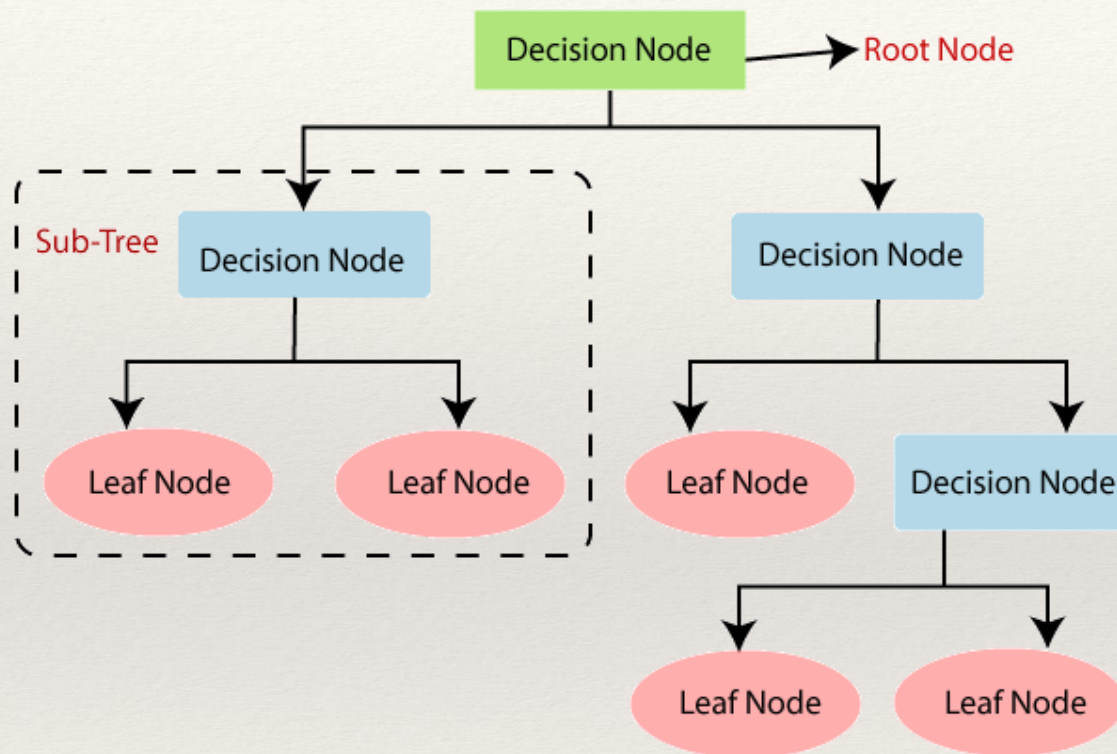
---

# Decision Trees – Part 2

Exercise IV

פיתוח:  
ד"ר יהונתן שלר  
משה פרידמן

# עצי החלטה – משמעות וקריאת עץ החלטה - תזכורת



עץ החלטה: סדרת  
השאלות שמביאה אותנו  
להחלטה

צומת שורש: השאלה  
הראשונה בעץ ההחלטה

צמתי ביניים: שאלות  
המשך

צמתי עלים: ההחלטה  
המתקבלת



---

# עץ החלטה: כיצד נבחר את התכונה הבאה - תזכורת

---

## 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

נדרגל 2 פונקציות לבחירת מאפיין לצומת:

- אנטרופיה
- Information gain

# אנטרופיה - entropy - תזכורת

❖ 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)

$$H(X) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \dots - p_n \log_2 p_n$$

$$= -\sum_{j=1}^n p_j \log_2 p_j$$

אנטרופיה גבוהה, משמעותה, שההתפלגות יותר דומה לאחידה.



---

# אנטרופיה מותנית - תזכורת

---

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

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

# תזכורת - Information Gain

Gain( $Y|X$ ) הינה ההפחתה הצפויה באנטרופיה של  $Y$  בגלל מיון עפ"י תכונה  $X$  ❖

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



# בניית עצי החלטה (אלגוריתם ID3) - תזכורת

לולאה:

- ❖ מצא את המאפיין הטוב ביותר  $X_i$  ושים אותו בצומת.
- ❖ עבור כל ערך של  $X_i$  צור קשת לכל אפשרות וצומת היוצאת ממנה
- ❖ מיין את הדוגמאות ב-train-set לצמתים החדשים (מהשלב הקודם)
- ❖ אם שילוב המאפיין והערך שלו, מוביל להחלטה טובה מספיק, צומת זה הוא עלה ומייצג החלטה.
- ❖ אם לא נוכל לקבל "החלטה טובה", נשאיר צומת זה כעלה
- ❖ אחרת, בצע את אותו תהליך עבור הצומת הזה

# תרגיל 5 - בחירת התכונה לצומת

(התרגיל הופיע בשבוע שעבר)

סיווג	גיל גדול מ-30	צבע
YES	כן	שחור
NO	לא	לבן
NO	לא	צהוב
YES	כן	שחור
YES	כן	צהוב
YES	כן	לבן
YES	כן	צהוב
NO	לא	שחור
NO	לא	שחור
YES	כן	לבן
YES	כן	לבן
NO	לא	שחור
YES	כן	צהוב
NO	לא	שחור
YES	כן	צהוב
NO	לא	שחור
YES	כן	צהוב
YES	כן	לבן
YES	כן	לבן
NO	לא	צהוב
NO	לא	צהוב

נתון ה-train-set הבא,  
עם 2 מאפיינים:  
צבע (שחור/לבן/צהוב)  
גיל (גדול מ-30/קטן או שווה ל-30)  
וקטגוריה: סיווג (Yes/No)



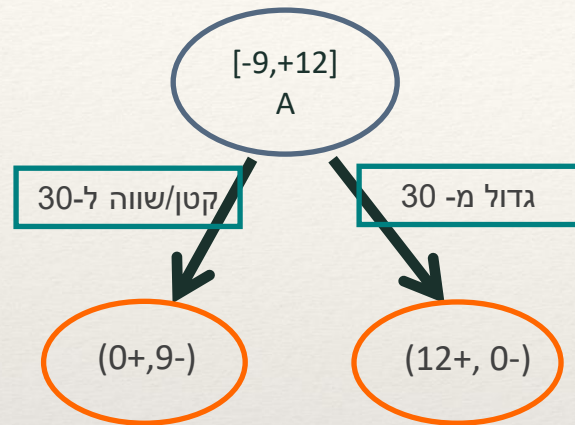
# תרגיל 5 - בחירת התכונה לצומת

- ❖ 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?**

# תרגיל 5 - בחירת התכונה לצומת – פתרון בעזרת IG

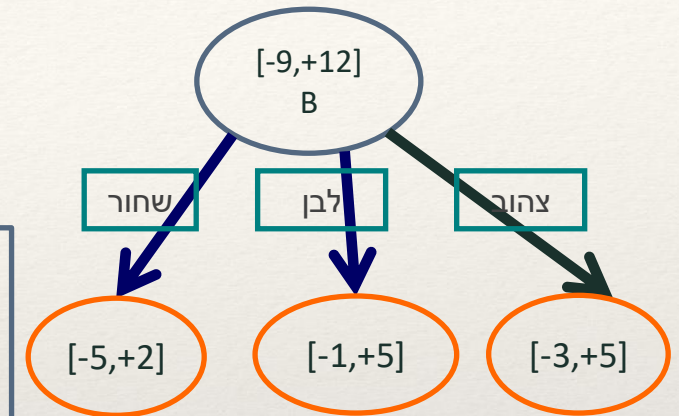
## נחשב את ההסתברויות ...



$$\begin{aligned} p(\text{סיווג} = \text{Yes} | \text{age} \leq 30) &= \frac{0}{9} = 0 \\ p(\text{סיווג} = \text{No} | \text{age} \leq 30) &= \frac{9}{9} = 1 \\ p(\text{סיווג} = \text{Yes} | \text{age} > 30) &= \frac{12}{12} = 1 \\ p(\text{סיווג} = \text{No} | \text{age} > 30) &= \frac{0}{12} = 0 \end{aligned}$$

ראשית נחשב הסתברויות ...

$$\begin{aligned} p(\text{סיווג} = \text{Yes} | \text{color} = \text{black}) &= \frac{2}{7} \approx 0.286 \\ p(\text{סיווג} = \text{No} | \text{color} = \text{black}) &= \frac{5}{7} \approx 0.714 \\ p(\text{סיווג} = \text{Yes} | \text{color} = \text{white}) &= \frac{5}{6} \approx 0.833 \\ p(\text{סיווג} = \text{No} | \text{color} = \text{white}) &= \frac{1}{6} \approx 0.166 \\ p(\text{סיווג} = \text{Yes} | \text{color} = \text{yellow}) &= \frac{5}{8} = 0.625 \\ p(\text{סיווג} = \text{No} | \text{color} = \text{yellow}) &= \frac{3}{8} = 0.375 \end{aligned}$$



$$\begin{aligned} p(\text{סיווג} = \text{Yes}) &= \frac{12}{21} \approx 0.57 \\ p(\text{סיווג} = \text{No}) &= \frac{9}{21} \approx 0.43 \end{aligned}$$



# תרגיל 5 - בחירת התכונה לצומת – פתרון בעזרת IG

## נחשב את האנטרופיה של המחלקה ...

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

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

$$p(\text{סיווג} = \text{No}) \approx 0.43$$

חישובנו:

נחשב את האנטרופיה של המחלקה ...

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

# תרגיל 5 - בחירת התכונה לצומת – פתרון בעזרת IG

חישובנו:

$$p(\text{סיווג} = \text{Yes} | \text{age} \leq 30) = 0$$

$$p(\text{סיווג} = \text{No} | \text{age} \leq 30) = 1$$

$$p(\text{סיווג} = \text{Yes} | \text{age} > 30) = 1$$

$$p(\text{סיווג} = \text{No} | \text{age} > 30) = 0$$

$$p(\text{סיווג} = \text{No} | \text{color} = \text{white}) \approx 0.166$$

$$p(\text{סיווג} = \text{Yes} | \text{color} = \text{yellow}) = 0.625$$

$$p(\text{סיווג} = \text{No} | \text{color} = \text{yellow}) = 0.375$$

$$p(\text{סיווג} = \text{Yes} | \text{color} = \text{black}) \approx 0.286$$

$$p(\text{סיווג} = \text{No} | \text{color} = \text{black}) \approx 0.714$$

$$p(\text{סיווג} = \text{Yes} | \text{color} = \text{white}) \approx 0.833$$

נחשב את האנטרופיה המותנת  $H(Y|B)$  ...

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

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

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

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

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

נחשב את האנטרופיה המותנת  $H(Y|A)$  ...

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

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

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

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



# תרגיל 5 - בחירת התכונה לצומת – פתרון בעזרת IG

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

חישבנו:

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

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

$$H(Y) = 0.98$$

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

# תרגיל 6 – האם (מבחינת מזג האוויר) מתאים לשחק מחר טניס?

Day	Outlook	Temp.	Humidity	Wind	Play?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

❖ ה- training-set שמציג נתוני עבר

המאפיינים:

❖ Outlook – תחזית מבחינת גשם

❖ Temp – טמפרטורה

❖ Humidity – לחות

❖ Wind - רוחות

המחלקה: Play – האם התחזית אפשרה משחק או לא

שאלה: האם (מבחינת מזג האוויר) מתאים לשחק מחר  
טניס?



# תרגיל 6 – האם (מבחינת מזג האוויר) מתאים לשחק מחר טניס?

Day	Outlook	Temp.	Humidity	Wind	Play?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

❖ שלב 1: עבור כל מאפיין ועבור כל ערך  
נבדוק את התפלגות המחלקה (כמה דוגמאות  
חיוביות וכמה שליליות) – בצורה זו:

$Values(Outlook) = Sunny, Overcast, Rain$

$$S = [9+, 5-]$$

$$S_{Sunny} = [2+, 3-]$$

$$S_{Overcast} = [4+, 0-]$$

$$S_{Rain} = [3+, 2-]$$

# תרגיל 6 – האם (מבחינת מזג האוויר) מתאים לשחק מחר טניס?

Day	Outlook	Temp.	Humidity	Wind	Play?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

❖ שלב 2: עבור כל מאפיין, נחשב את ה- Info gain המותנה של המחלקה במאפיין

$$\begin{aligned}Gain(S, Outlook) &= Entropy(S) \\&\quad - (5/14)Entropy(S_{Sunny}) \\&\quad - (4/14)Entropy(S_{Overcast}) \\&\quad - (5/14)Entropy(S_{Rain}) \\&= .940 - (5/14).971 - (4/14)0 - (5/14).971 \\&= .247\end{aligned}$$

$$\begin{aligned}Gain(S, Temperature) &= Entropy(S) \\&\quad - (4/14)Entropy(S_{Hot}) \\&\quad - (6/14)Entropy(S_{Mild}) \\&\quad - (4/14)Entropy(S_{Cool}) \\&= .940 - (4/14)1 - (6/14).811 - (4/14).971 \\&= .029\end{aligned}$$



# תרגיל 6 – האם (מבחינת מזג האוויר) מתאים לשחק מחר טניס?

❖ שלב 2 (המשך): עבור כל מאפיין, נחשב את ה-Info gain המותנה של המחלקה במאפיין

Day	Outlook	Temp.	Humidity	Wind	Play?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

$$\begin{aligned} \text{Gain}(S, \text{Humidity}) &= \text{Entropy}(S) \\ &\quad - (7/14)\text{Entropy}(S_{\text{High}}) \\ &\quad - (7/14)\text{Entropy}(S_{\text{Normal}}) \\ &= .940 - (7/14).985 - (7/14).592 \\ &= .152 \end{aligned}$$

$$\begin{aligned} \text{Gain}(S, \text{Wind}) &= \text{Entropy}(S) \\ &\quad - (8/14)\text{Entropy}(S_{\text{Weak}}) \\ &\quad - (6/14)\text{Entropy}(S_{\text{Strong}}) \\ &= .940 - (8/14).811 - (6/14)1 \\ &= .048 \end{aligned}$$

נבחר את המאפיין outlook,  
עבור הצומת הבא



# תרגיל 7 – בניית עץ החלטה

❖ נניח שברשותי שלושה מטבעות: מטבע של 10 שקלים, מטבע של חמישה שקלים ומטבע של שקל. אני זורק את שלושת המטבעות בו-זמנית ורושם את תוצאות הזריקה ("עץ" או "פלי"). נתונה טבלת המאורעות:

	שקל"	10 ש"ח	סיווג - 5 ש"ח
1	"עץ"	"עץ"	"פלי"
2	"עץ"	"פלי"	"עץ"
3	"פלי"	"עץ"	"עץ"
4	"עץ"	"פלי"	"פלי"
5	"פלי"	"פלי"	"פלי"
6	"עץ"	"פלי"	"עץ"

חשבו עץ החלטה המסווג תוצאת זריקת מטבע של חמישה ₪ על סמך ידיעת הטלת מטבע שקל ומטבע של 10 ₪

האם בכלל ניתן לחזות תוצאת הטלת מטבע 5 ₪ על בסיס ידע של הטלות של מטבעות אחרים???



# תרגיל 7 – בניית עץ החלטה

- ❖ נניח שברשותי שלושה מטבעות: מטבע של 10 שקלים, מטבע של חמישה שקלים ומטבע של שקל. אני זורק את שלושת המטבעות בו-זמנית ורושם את תוצאות הזריקה ("עץ" או "פלי"). נתונה טבלת המאורעות:

	A: "שקל"	B: 10 ש"ח	Y - סיווג - 5 ש"ח
1	"עץ"	"עץ"	"פלי"
2	"עץ"	"פלי"	"עץ"
3	"פלי"	"עץ"	"עץ"
4	"עץ"	"פלי"	"פלי"
5	"פלי"	"פלי"	"פלי"
6	"עץ"	"פלי"	"עץ"

פתרון: נסמן הטלת מטבע 5 כ-Y, הטלת מטבע "שקל" כ-A והטלת מטבע 10 כ-B. נחשב  $\text{InformationGain}(Y | A)$  ונחשב  $\text{InformationGain}(Y | B)$  ונכריע לפי איזו תכונה כדאי לשאול בראש העץ

# תרגיל 7 – בניית עץ החלטה

- ❖ נניח שברשותי שלושה מטבעות: מטבע של 10 שקלים, מטבע של חמישה שקלים ומטבע של שקל. אני זורק את שלושת המטבעות בו-זמנית ורושם את תוצאות הזריקה ("עץ" או "פלי"). נתונה טבלת המאורעות:

	A: "שקל"	B: 10 ש"ח	Y - סיכוי - 5 ש"ח
1	"עץ"	"עץ"	"פלי"
2	"עץ"	"פלי"	"עץ"
3	"פלי"	"עץ"	"עץ"
4	"עץ"	"פלי"	"פלי"
5	"פלי"	"פלי"	"פלי"
6	"עץ"	"פלי"	"עץ"

נחשב את האנטרופיה של מטבע 5 ש"ח  $H(Y)$  עבור ששת הזריקות:

$$P(\text{ets}) = 3/6 = 0.5$$

$$P(\text{Pali}) = 3/6 = 0.5$$

$$H(Y) = -0.5 \log_{0.5} - 0.5 \log_{0.5} = 0.5 + 0.5 = 1$$



# תרגיל 7 – בניית עץ החלטה

	A: "שקל"	B: 10 ש"ח	Y - סיווג - 5 ש"ח
1	"עץ"	"עץ"	"פלי"
2	"עץ"	"פלי"	"עץ"
3	"פלי"	"עץ"	"עץ"
4	"עץ"	"פלי"	"פלי"
5	"פלי"	"פלי"	"פלי"
6	"עץ"	"פלי"	"עץ"

נחשב את האנטרופיה הספציפית של מטבע 5 ש"ח בהינתן שתוצאת הטלת מטבע "שקל" הינה "עץ".

$$P(Y = ets | A = ets) = 0.5$$

$$P(Y = pali | A = ets) = 0.5$$

$$H(Y | A = ets) = 1$$

# תרגיל 7 – בניית עץ החלטה

	A: "שקל"	B: 10 ש"ח	Y - סיווג - 5 ש"ח
1	"עץ"	"עץ"	"פלי"
2	"עץ"	"פלי"	"עץ"
3	"פלי"	"עץ"	"עץ"
4	"עץ"	"פלי"	"פלי"
5	"פלי"	"פלי"	"פלי"
6	"עץ"	"פלי"	"עץ"

נחשב את האנטרופיה הספציפית של מטבע 5 ש"ח בהינתן שתוצאת הטלת מטבע "שקל" הינה "פלי".

$$P(Y = ets | A = pali) = 0.5$$

$$P(Y = pali | A = pali) = 0.5$$

$$H(Y | A = pali) = 1$$



# תרגיל 7 – בניית עץ החלטה

	A: "שקל"	B: 10 ש"ח	Y - סיווג - 5 ש"ח
1	"עץ"	"עץ"	"פלי"
2	"עץ"	"פלי"	"עץ"
3	"פלי"	"עץ"	"עץ"
4	"עץ"	"פלי"	"פלי"
5	"פלי"	"פלי"	"פלי"
6	"עץ"	"פלי"	"עץ"

נחשב את האנטרופיה הספציפית של מטבע 5 בהינתן הטלת מטבע "שקל"

$$\begin{aligned} H(Y | A) &= \sum_j P(A = \alpha_j) H(Y | A = \alpha_j) \\ &= p(A = ets) H(Y | A = ets) + p(A = pali) H(Y | A = pali) \\ &= \frac{2}{3} \times 1 + \frac{1}{3} \times 1 = 1 \end{aligned}$$

# תרגיל 7 – בניית עץ החלטה

	A: "שקל"	B: 10 ש"ח	Y - סיווג - 5 ש"ח
1	"עץ"	"עץ"	"פלי"
2	"עץ"	"פלי"	"עץ"
3	"פלי"	"עץ"	"עץ"
4	"עץ"	"פלי"	"פלי"
5	"פלי"	"פלי"	"פלי"
6	"עץ"	"פלי"	"עץ"

נחשב את האנטרופיה הספציפית של מטבע 5 ש"ח בהינתן הטלת מטבע 10 ש"ח

$P(Y = ets   B = pali) = 0.5$	$P(Y = ets   B = ets) = 0.5$
$P(Y = pali   B = pali) = 0.5$	$P(Y = pali   B = ets) = 0.5$
$H(Y   B = pali) = 1$	$H(Y   B = ets) = 1$

$$\begin{aligned} H(Y | B) &= \sum_j P(B = \alpha_j) H(Y | B = \alpha_j) \\ &= p(B = ets) H(Y | B = ets) + p(B = pali) H(Y | B = pali) \\ &= \frac{1}{3} \times 1 + \frac{2}{3} \times 1 = 1 \end{aligned}$$



# תרגיל 7 – בניית עץ החלטה

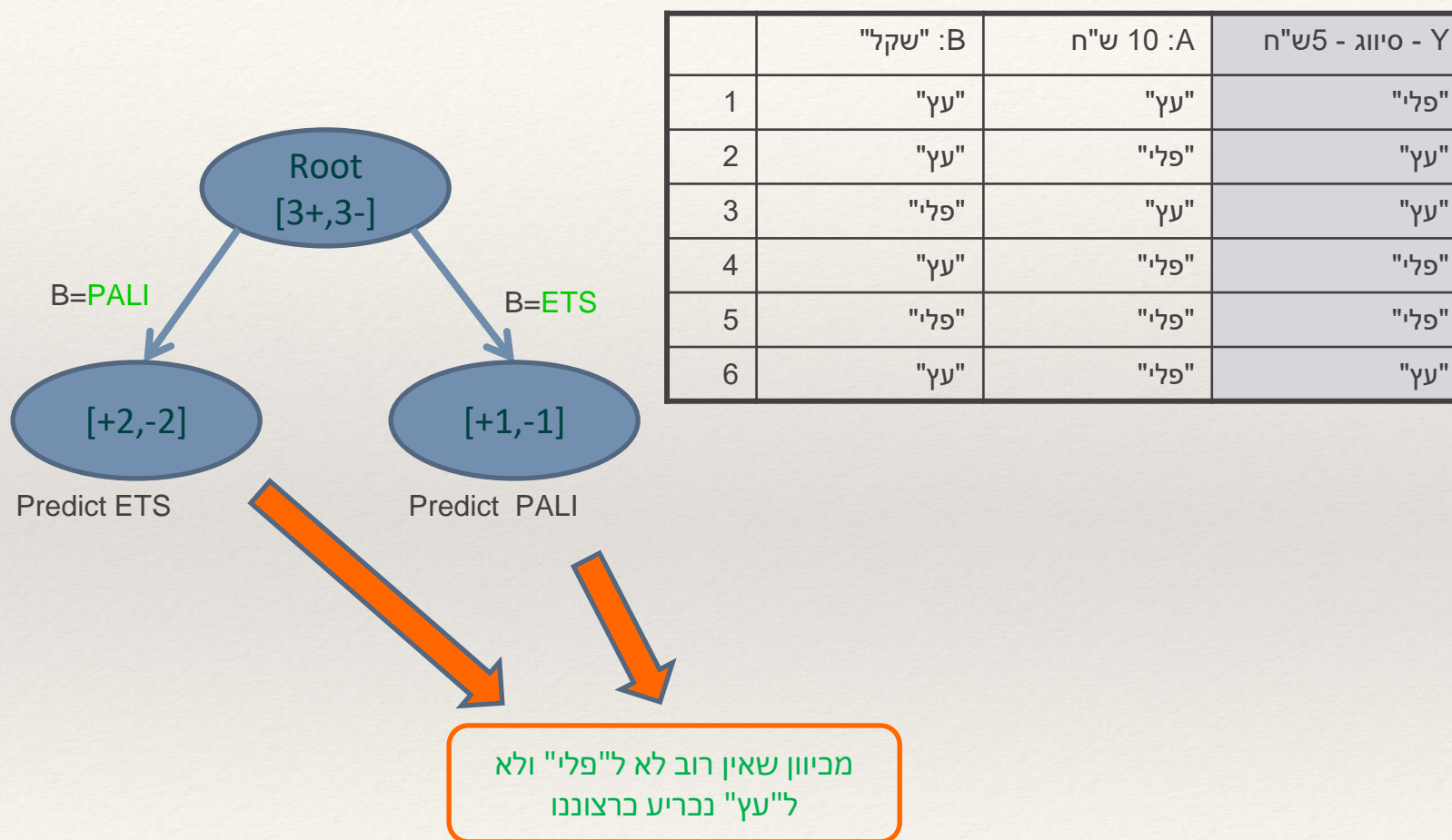
	A: "שקל"	B: 10 ש"ח	Y - סיווג - 5 ש"ח
1	"עץ"	"עץ"	"פלי"
2	"עץ"	"פלי"	"עץ"
3	"פלי"	"עץ"	"עץ"
4	"עץ"	"פלי"	"פלי"
5	"פלי"	"פלי"	"פלי"
6	"עץ"	"פלי"	"עץ"

$$IG(Y | A) = H(Y) - H(Y | A) = 1 - 1 = 0$$

$$IG(Y | B) = H(Y) - H(Y | B) = 1 - 1 = 0$$

שתי התכונות שוות ב"איכות" שלהן. כפי שראינו בהרצאה  
infoGain=0 זו לא סיבה להפסיק בבניית העץ. נבחר את התכונה  
10 נה לשורש של העץ

# תרגיל 7 – בניית עץ החלטה





# מתי הופך צומת לעלה (כלומר מתי לא לפצל)?

מקרה בסיסי 1 (Base Case 1): לא לפצל את הצומת (כלומר הצומת תהפוך לעלה), אם כל הדוגמאות שמויננו מה-training-set בעלי אותה קטגוריה

מקרה בסיסי 2 (Base Case 2): לא לפצל את הצומת (כלומר הצומת תהפוך לעלה), אם ערכי המאפיינים בדוגמאות שנותרו זהה.

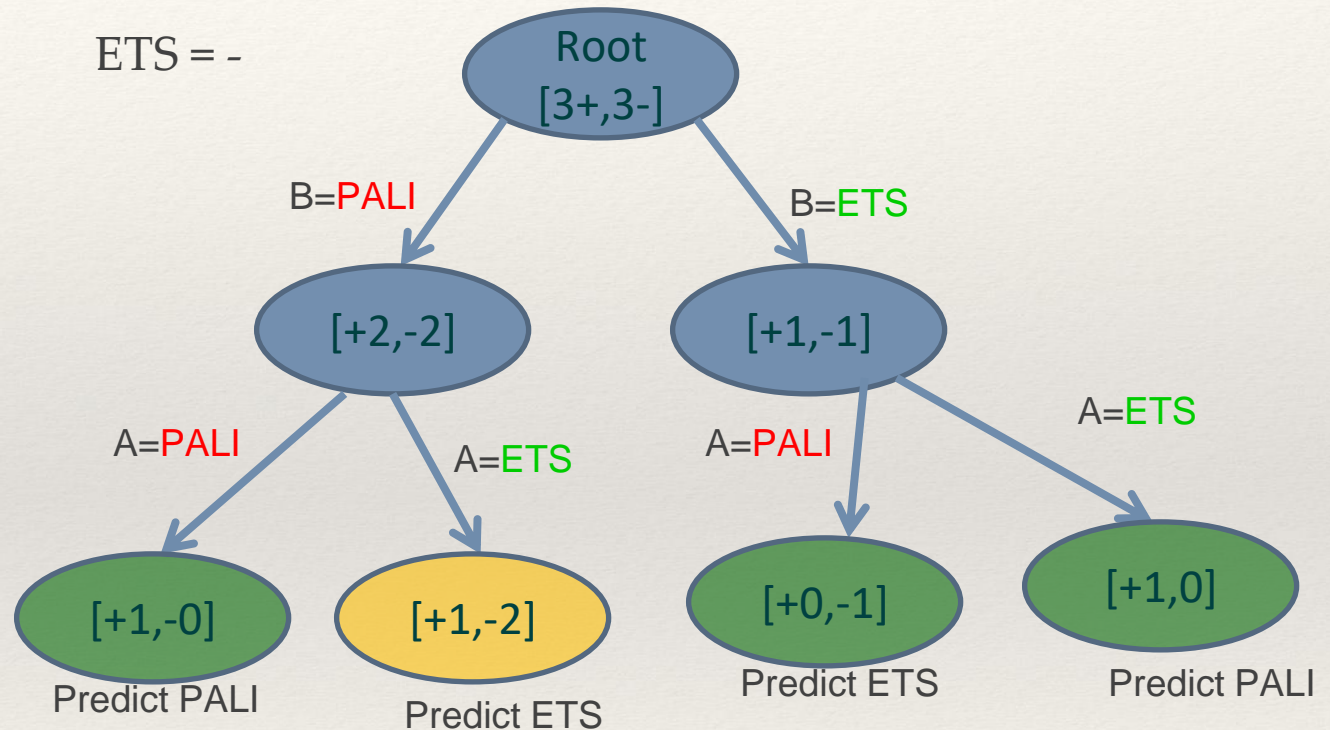
❖ במקרה זה נסווג את קטגוריית הרוב

# חזרה לתרגיל 7 – בניית עץ החלטה

PALI = +  
ETS = -

ה"עלים" הירוקים  
הם  
Base case1

ה"עלים" הכתומים  
הם  
Base case2





Confusion matrix:

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

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

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

# חזרה לתרגיל 7 – בניית עץ החלטה

PALI = +

ETS = -

מהו הדיוק אליו הגענו על קבוצת האימון? חשב את ה-TP ואת ה-FP

Positive = PALI, Negative = ETS

תשובה: סה"כ  $\#P = 3$ ,  $\#N = 3$

$TP=2$  ;  $FP=0$  ;  $FN=1$  ;  $TN=3$

$Accuracy = 5/6 \approx 0.83$

האם זה אומר שאני נביא שיכול לחזות ע"י הטלת מטבעות של 1 ₪ ושל 10 ₪ את תוצאת הטלת מטבע 5 ₪ בהסתברות יותר טובה ממקרית?

תשובה: וודאי שלא. זה אומר שקבוצת האימון קטנה מדי. לאחר מספר חזרות רב של הניסוי הטעות שלנו גם על קבוצת האימון תגדל (יהיו שורות עם ערך זהה עבור הטלת 1 ₪ והטלת 10 ₪ אבל ערכים שונים עבור הטלת מטבע 5 ₪)



---

# שאלות?

---