

2 שנה 8

1.1.

לפי הנתונים, המידע של הנתונים הוא 1.0
לפי הנתונים, המידע של הנתונים הוא 1.0

$$H(S) = -0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1$$

$$P(S | Top) = 0.3$$

$$P(S | Middle) = 0.4$$

$$P(S | Bottom) = 0.3$$

$$H(S | Top) = -\frac{2}{3} \log_2 \left(\frac{2}{3}\right) - \frac{1}{3} \log_2 \left(\frac{1}{3}\right) = 0$$

$$H(S | Middle) = -\frac{2}{4} \log_2 \left(\frac{2}{4}\right) - \frac{2}{4} \log_2 \left(\frac{2}{4}\right) = 1$$

$$H(S | Bottom) = -\frac{3}{3} \log_2 \left(\frac{1}{3}\right) - \frac{2}{3} \log_2 \left(\frac{2}{3}\right) = 0$$

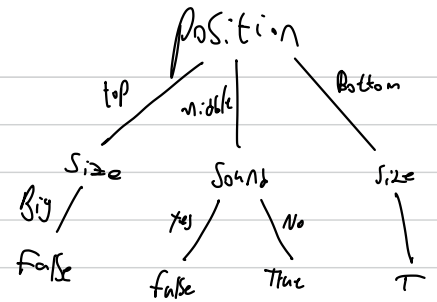
$$IG(S | Position) = 1 - 0.4 = 0.6$$

המידע של הנתונים
המידע של הנתונים

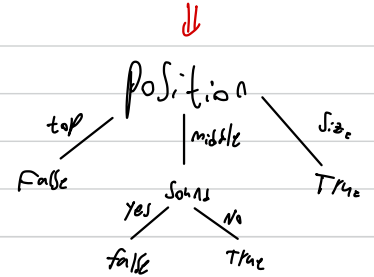
Clicked	Size	Position	Sound
F	Big	Top	No
F	Small	Middle	Yes
F	Small	Middle	Yes
T	Small	Bottom	No
T	Big	Bottom	No
F	Big	Top	Yes
T	Big	Bottom	Yes
T	Small	Middle	No
T	Small	Middle	No
F	Big	Top	No

1.2

Clicked	Size	Position	Sound
F	Big	Top	No
F	Small	Middle	Yes
F	Small	Middle	Yes
T	Small	Bottom	No
T	Big	Bottom	No
F	Big	Top	Yes
T	Big	Bottom	Yes
T	Small	Middle	No
T	Small	Middle	No
F	Big	Top	No



Position	leaf	N_j	$H(p_j)$
top	1	3	$-\frac{1}{3} \log \frac{1}{3} - \frac{2}{3} \log \frac{2}{3} = 0.9182$
middle	2	4	$-\frac{1}{4} \log \frac{1}{4} - \frac{1}{4} \log \frac{1}{4} = 1$
bottom	3	3	$-\frac{1}{3} \log \frac{1}{3} - \frac{2}{3} \log \frac{2}{3} = 0.9182$



Position = top

Size	Sound	Clicked
Big	No	F
Big	Yes	F
Big	No	F

Size	leaf	N_j	$H(p_j)$
Big	1	3	$-\frac{1}{3} \log \frac{1}{3} - \frac{2}{3} \log \frac{2}{3} = 0.9182$

Position = middle

Size	Sound	Clicked
Small	Yes	F
Small	Yes	F
Small	No	T
Small	No	T

Sound	leaf	N_j	$H(p_j)$
Yes	1	2	$-\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} = 1$
No	2	2	$-\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} = 1$

Position = bottom

Size	Sound	Clicked
Small	No	T
Big	Yes	T

Size	leaf	N_j	$H(p_j)$
Small	1	1	$-\frac{1}{1} \log \frac{1}{1} = 0$
Big	2	1	$-\frac{1}{1} \log \frac{1}{1} = 0$

2. Naive Base

Clicked	Size	Position	Sound
F	Big	Top	No
F	Small	Middle	Yes
F	Small	Middle	Yes
T	Small	Bottom	No
T	Big	Bottom	No
F	Big	Top	Yes
T	Big	Bottom	Yes
T	Small	Middle	No
T	Small	Middle	No
F	Big	Top	No

if: FG, with: pp'n, Ge Ge : new

	Big	Small
Yes	$\frac{2}{5}$	$\frac{3}{5}$
No	$\frac{2}{5}$	$\frac{3}{5}$

	Top	Middle	Bottom
Yes	$\frac{0}{5}$	$\frac{1}{5}$	$\frac{2}{5}$
No	$\frac{3}{5}$	$\frac{2}{5}$	$\frac{0}{5}$

	Sound Yes	Sound No
Yes	$\frac{1}{5}$	$\frac{4}{5}$
No	$\frac{3}{5}$	$\frac{2}{5}$

New $\Rightarrow p(\text{Size} = \text{big}, \text{Position} = \text{Middle}, \text{Sound} = \text{No})$

$$p(\text{Yes} | \text{New}) = p(\text{Yes}) \cdot p(\text{big} | \text{Yes}) \cdot p(\text{Middle} | \text{Yes}) \cdot p(\text{No} | \text{Yes}) = 0.5 \cdot \frac{2}{5} \cdot \frac{1}{5} \cdot \frac{4}{5} = 0.064$$

$$p(\text{No} | \text{New}) = p(\text{No}) \cdot p(\text{big} | \text{No}) \cdot p(\text{Middle} | \text{No}) \cdot p(\text{No} | \text{No}) = 0.5 \cdot \frac{2}{5} \cdot \frac{2}{5} \cdot \frac{2}{5} = 0.048$$

$$p(\text{Yes} | \text{New}) > p(\text{No} | \text{New})$$

↓

Yes is new

1. Describe the analytical solution for linear regression with MSE as a distance function. (4 points)
2. What is the problem with information gain? Describe any solution for it. (4 points)
3. Why do we use Gradient Descent or Neotun Roffson for Linear Regression? (4 points)
4. Explain how a Decision tree is used for regression problems. (4 points)

1. רגרסיה ליניארית היא מודל סטטיסטי המשמש לחיזוי משתנה תלוי (Y) כפונקציה ליניארית של משתנים בלתי תלויים (X). פונקציית האובדן הנפוצה ביותר ברגרסיה ליניארית היא סכום הריבועים הממוצע (MSE), המודד את המרחק הממוצע בין הערכים החזויים לערכים האמיתיים.

2.

- **הטיה לטובת תכונות עם מספר רב של ערכים:** תכונות עם מספר רב של ערכים (כגון קטגוריות טקסטואליות) נוטות לקבל רווח מידע גבוה יותר, גם אם הן אינן רלוונטיות לסיווג.
- **רגישות לרעש בנתונים:** ערכים חריגים בנתונים יכולים להשפיע באופן משמעותי על חישוב רווח המידע.
- **אי התאמה למטרות סיווג ספציפיות:** רווח מידע אינו מתייחס באופן ישיר למטרות סיווג ספציפיות, כגון דיוק או זיכרון.

3. גרדיאנט דיסנט וניוטון-רפסון ברגרסיה ליניארית

גרדיאנט דיסנט:

- קל ליישום ולהבנה.
- עובד טוב עם מערכי נתונים גדולים כיוון שהוא מעדכן משקולות צעד אחר צעד.
- ניתן לשנות אותו בקלות עבור פונקציות עלות ומודלים שונים.
- מציע גמישות בהגדרת קצב למידה וקריטריונים לעצירה..

ניוטון-רפסון:

- התכנסות מהירה יותר ברוב המקרים, במיוחד עבור פונקציות עלות עם התנהגות טובה.
- דורש פחות איטרציות כדי להגיע למינימום.

- **ניוטון-רפסון:** מתאים עבור מערכי נתונים קטנים יותר, פונקציות עלות עם התנהגות טובה וכאשר רצויה התכנסות מהירה יותר. עם זאת, יש להבטיח יציבות מספרית ואתחול נכון.

4. עצי החלטה לרגרסיה שימושיים לסיווג קשרים מורכבים בין תכונות קלט ומשתני יעד שהם רציפים. הם ניתנים לפירוש, קלים להמחשה ויכולים להתמודד עם קשרים שהם לא ליניאריים בנתונים.