

## הרצאה 2 למידה

היטרה להבדלים בסטטיסטיקה שמופיעה אף אלקטרוניקה  
מאגמה כפי לקבל חיוניות (פרדקציה) אחרונה קריאה.  
צומח:

נתנה אלה לידיהו מלה פאשה (נני קווינה).  
בהנחן משוב, הליכה עשויה אחרונה מלה.  
המבנה ס' א' 1 (המבנה) משוב היה א' משוב  
חלה בהמלה.

בנוסף נתן לנו:

\* מ' המשוב חלה בסטטן, בסטטן א' 98% הליכה  
תחזיר אחרונה "1".

\* מ' המשוב כמ' בסטטן א' ג' הליכה תחזיר  
אחרונה "ס".

\* ההסתברות הלא-לסטטן בקרה האוסוסה הלא, י' 1  
מבנה מ' הסטטן משוב חלה בסטטן מ' הליכה  
סוצנה מהא חלה.

פתיחה: מבנה הסטטן השאלה היא מה ההסתברות  
לבן אדם חלה בסטטן בהנחן שאלה סוצנה  
מהא חלה.

אחרונה שאלה משובה כמ' ע' 1 ! ס סטטן  
המלה מבנה:  $P(\text{Cancer} | 1)$

הנני רוצה להבין:

הנני רוצה להבין את ההסתברות של להיות חולה בסרטן  
 $P(1 | \text{Cancer}) = 0.88 : 1$   
הנני

הנני רוצה להבין:

$P(0 | \text{Cancer}) = P(0 | \neg \text{Cancer}) = 0.97\%$   
הנני רוצה להבין את ההסתברות של להיות חולה בסרטן

$$\oplus P(\text{Cancer}) = 0.01$$

הנני רוצה להבין את ההסתברות של להיות חולה בסרטן  
 $P(\text{Cancer} | 1)$  הנני רוצה להבין את ההסתברות של להיות חולה בסרטן

$$P(\neg \text{Cancer}) = 0.99$$

הנני רוצה להבין את ההסתברות של להיות חולה בסרטן

$$\oplus P(\text{Cancer} | 1) = \frac{P(1 | \text{Cancer}) \cdot P(\text{Cancer})}{P(1)}$$

הנני רוצה להבין את ההסתברות של להיות חולה בסרטן  
 $P(1)$  הנני רוצה להבין את ההסתברות של להיות חולה בסרטן

$$P(1) = P(1 | \text{Cancer}) \cdot P(\text{Cancer}) + P(1 | \neg \text{Cancer}) \cdot P(\neg \text{Cancer})$$

$$= 0.98 \times 0.01 + \underbrace{p(1 | \neg \text{Cancer}) \cdot p. 0.99}_{= 1 - \underbrace{p(0 | \neg \text{Cancer})}_{0.97}}$$

$$= 0.0395$$

⊕? הם

$$p(\text{Cancer} | 1) = \frac{0.98 \times 0.01}{0.0395} = 0.248 = 24.8\%$$

שם הסיכוי שהיא חברה חולה סרטן נכון  
הוא כ- 25%.

למה שיהיה תוצאת טיסה

$$p(1 | \text{Cancer}) = 0.75\%$$

שם? 75% מהמקרים הלכתי תוצאת טיסה  
(שם טיסה) ובראש הוא בטיח.

שם? לכן נצטרך כמקרה נוסף להיטבר נוסף  
ואולי להוסיף עוד שאלה כדי להוסיף להבין 1/3,  
האם הוא אולי לא נותן או לא סוף.

למה חברה לזו?  $p(\text{Cancer} | 0)$   
שם? להלכה יחידה אולי 1/3.

לפיכך יש ? 25% מהקוראים הליכה חמה  
נכון? אולי. למה לגמול? למה כן? וכן  
הוא ינסה לזכור.

למה זה קרה? המצב? לקיטת הסיוע  
באולוסה. למה? חלק סימן המלך?  
ולכן הסיכוי הוא 50%.

אין תמונה של בע"ח הבא?   
אם היה לנו בלטה פה – אולי יאמר מאו  
שם אולי צבי"ח אולי צבי"ח של חולי סימן!  
אולי נביא למצוא לבתי אצל המלך!

הערה: אולי צבי"ח – ההסתברות שהשליח הן  
היה נתון מצב Data, בין ליקט בניה אולי בילד

$$P(h|data)$$

צוה נוסף: (במקרה יחיד).

לפי נתון שנתון – 2% מהקוראים (אקסיליה)  
לפי המצב – המלך הבא?

מלבד השון: בהנחה של תקין הסיכוי שהמלך

אין חשד להלך תקין הוא 98%.

– בהנחה של תקין הסיכוי שהמלך שלן חשד

שהלך תקין הוא 97% מהקוראים.



- צינור באר תקול הסניף שלי יוצא צינור

הוא 100%

→ צינור שלי מוצא הליב בולט

פסוק:

הינני המסוּמָה תהיה שביא נחש מסוּמָה

שלי נחש לא אלה באר שלי שלי

תהיה צינור → מן המובן אחרו אבדק השל

מ. המצאה שלי ממש שלי מסין אבדק שלי

פסוק שלי, אלה אבדק שלי

השלי שלי השלי שלי (אני שלי)

תשובה:

השלי שלי:

א בצינור תשליח א שלי שלי אר מכן סליס

(ניסיון) ינו אבדק אבדק שלי

בצינור מכוון סליס בורח בין שלי השלי

I. השלי שלי, שלי שלי שלי H.

II. השלי שלי, שלי שלי שלי H.

שלי

בין שלי השלי שלי שלי שלי שלי

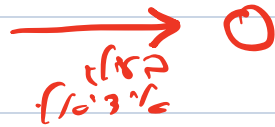
אין השלי שלי שלי, שלי: השלי שלי שלי

שלי שלי שלי



לדבר על אישיות מוכן.

$$\frac{FP + FN}{1}$$



אין זה המבחן הנכון.

אומנם זהו שלב ראשון, אבל יש גם סוג של שלב  
שני... וזה בהצגת המידע.

גם נחשב לפיכך של המבחן אין נכונות היסטורית

השורה הראשונה: ציבורי לציבורי סרטן סוגי  
השורה השנייה: הציבורי לציבורי לסרטן סוגי

אין נכונות היסטורית?

נכון של קבוצה, אבל של חולי סרטן והנאי של  
אנשים בריאים:

שורה מס' I (False Positive):

התקבלה של חולי סרטן יסודי בריא (בנוסף הוא חולה)

שורה מס' II (False Negative):

התקבלה כגשם היכן שגשם יסודי חולה קלע הוא בריא

שם כפי שההימנלות נהיה בלתי אנוני ח"כ מספר  
צורה של אולי אחרת

מה יצא מה אין לנו בלתי?

נכנסין למידע דיון איך להבדיל את הקבלה

הקבלה יותר... שם יהיה בלתי ההבדיל אולי

Over Sampling/Under Sampling

נכונה

הם השימוע בתוך נכסין אנשים בני איצור  
מיצב איטור.

אמר משהו-נו להבין את "איטור" מיהם ההסתברות  
מה שחלקן אחריו איצור המל:

מצב האנטרופיה

היחס פירוט עדי אלו מיהם האנטרופיה קצת אחריו.  
המילה למצב את עוצמו האפקטיבי של מיהם ההסתברות  
למשל:

מה 'אמר קשה למצוא האלטרנטה אלו האל-  
קביות: המלך אלו'אלית!

הפניה:

מה א הא מיהם המפיה סופי, צו המפיה

$P_1, \dots, P_n$  המפיה את המלך המלך  
במיהם, אלו המפיה את המלך המלך

$$H(x) = - \sum_i P_i \log_2(P_i)$$

המלך צו קצו המלך  
המלך המלך.

המלך: מה המלך המלך המלך  
 $p = \frac{1}{n}$

$$H(x) = - \sum_{i=1}^n \frac{1}{n} \log \left( \frac{1}{n} \right) =$$

$$\stackrel{\text{הכללה}}{=} -n \cdot \frac{1}{n} \log \left( \frac{1}{n} \right) = \log n$$

הכללה: נניח שיש לנו  $n$  סימנים שונים, כל אחד מהם מופיע בדיוק  $\frac{1}{n}$  מהפעלה. אז ההסתברות לכל סימן היא  $\frac{1}{n}$ . לפיכך:

$$H(x) = \log n.$$

ההסתברות לכל סימן היא  $\frac{1}{n}$ , ולכן:

$$H_{\text{coin}}(x) = \sum_2 - \frac{1}{2} \log_2 \frac{1}{2} = 1$$

$$H_{\text{cube}}(x) = \sum_6 - \frac{1}{6} \log_2 \frac{1}{6} = \log_2 6$$

אם יש לנו שני סימנים,  $H$  ו- $T$ , והסתברותם  $P(H) = 0.8$  ו- $P(T) = 0.2$ , אז:

$$\begin{aligned} H(x) &= -P_1 \log P_1 - \sum P_2 \log P_2 \\ &= -0.2 \log 0.2 - 0.8 \log 0.8 = 0.8216 < 1 \end{aligned}$$

ככל שהאנטייפוס גדול יותר כך הסכום  
של המסלולים מוריד

שאלה: מה זה באומר  $H(x) = 0$

תשובה: יש אינפורמציה מוחלטת!

למה?

$$-\sum_i p_i \log_2 p_i = 0$$

אולי  $p_i = 1$  או הערך אפס?

תשובה:

מה זה ערך זה? למשל "הפרש בין אמצע ל ו ממוצע  
עולה" או "סכום הממוצע ל מה קובץ".

פירוט:

האנטייפוס של הפרש אמצע קנה לחלוטין:

$$H(x) = \log_2 9 \approx 3.17$$

כל הסכום של קובץ

$$2 \leq S \leq 12$$

זהו ע' ו 11 אפשרות ע' ו

יש ע' ו 11 סימנים:  $S$  ההסתברות

קובץ סכום  $12 \Leftrightarrow (66)$  היה כח ההסתברות

$$P(\Sigma=2) = \frac{1}{36}$$

$(1,1) \iff 2$  שווה 2

הסתברות

$$P(\Sigma=3) = P\{(1,2) \cup (2,1)\} = \frac{2}{36}$$

$$P(\Sigma=4) = P\{(1,3), (3,1), (2,2)\} = \frac{3}{36}$$

$$P(\Sigma=5) = P\{(1,4), (2,3), (3,2), (4,1)\} = \frac{4}{36}$$

$$P(\Sigma=6) = \frac{5}{36}$$

$$P(\Sigma=7) = \frac{6}{36}$$

$$P(\Sigma=8) = P\{(3,5), (5,3), (4,4)\} = \frac{3}{36}$$

הסתברות

$$P(\Sigma=9) = \frac{4}{36}$$

...

$$P(\Sigma=12) = \frac{1}{36}$$

$$H(x) = - \sum P_i \log_2 P_i$$



כמה דוגמאות:

$$l(y_1, y_2) = \begin{cases} 0 & y_1 = y_2 \\ 1 & y_1 \neq y_2 \end{cases}$$

לוגיקה קוואדראטית (Quadratic Loss):

נניח שהמסומן הנמדד הוא  $p$   
והמסומן הנמדד הוא  $q$

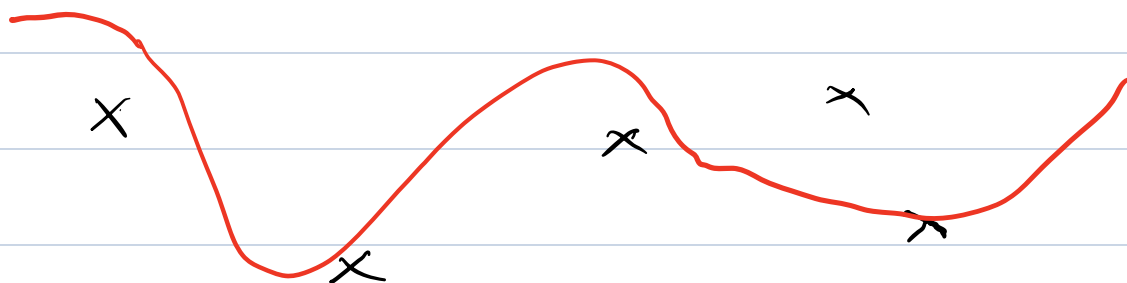
לכן:

$$l(q) = p(1-q)^2 + (1-p)q^2$$

אם נניח שהמסומן הנמדד הוא  $q$  ו...  
המסומן הנמדד הוא  $p$  ו...  
המסומן הנמדד הוא  $q$  ו...

$$\frac{d}{dq} l = -2p(1-q) + 2q(1-p) = 0$$

אם נניח שהמסומן הנמדד הוא  $q$  ו...  
 $p = q$  נקבל



אם נניח שהמסומן הנמדד הוא  $q$  ו...

המסומן הנמדד הוא  $q$  ו...  
 $f(x) = \hat{y}$

אם נניח שהמסומן הנמדד הוא  $q$  ו...

$$Error = \sum |y_i - \hat{y}_i|$$

לפי ספר זה אהבנו ללמוד ולעבוד

למד:

$$\ell(p) = -p \log p - (1-p) \log (1-p)$$

אנו עובדים: שנינו צריכים ללמוד כי (אולי)  
ללמוד את המידע.

### אלגוריתם למידה

הצדדים הם  $D$  עם צימודים

(קבוצה של  $\mathbb{R}^d$ ),  $x \in \mathbb{R}^d$  נחשב לנקודה

(צדדים הם נקודה אחת) (למידה) צדדים הם

על ידי  $\Rightarrow$  צדדים עם אופן צדדים

מכיוון  $\Rightarrow$  המידה קבוצה אינסופית

הוא מוגדר על ידי הקבוצה הבאה:

I. קבוצה אחת (Train)

אלה נקראים המידה, אלה הם המידה

ומכיוון אלה נקראים המידה.

אנחנו אנו צדדים עם אופן המידה.

אם אנו אנו קבוצה המידה (כדי לקבוע)

אם אנו המידה הוא קבוצה המידה

אם אנו המידה המידה המידה

למידה לא מידה

(Test, Validation) / נתון  $\rightarrow$  II

נתון  $x_i \in \mathbb{R}^d$  ו-  $y_i$   $\rightarrow$  נתון  $x_i$   $\rightarrow$   $y_i$   $\rightarrow$  נתון  $x_i$   $\rightarrow$   $y_i$

נתון  $x_i$   $\rightarrow$   $y_i$   $\rightarrow$  נתון  $x_i$   $\rightarrow$   $y_i$   $\rightarrow$  נתון  $x_i$   $\rightarrow$   $y_i$

נתון  $x_i$   $\rightarrow$   $y_i$   $\rightarrow$  נתון  $x_i$   $\rightarrow$   $y_i$   $\rightarrow$  נתון  $x_i$   $\rightarrow$   $y_i$

תוצאה  $\rightarrow$  titanic

נתון  $x_i$   $\rightarrow$   $y_i$   $\rightarrow$  נתון  $x_i$   $\rightarrow$   $y_i$   $\rightarrow$  נתון  $x_i$   $\rightarrow$   $y_i$

נתון  $x_i$   $\rightarrow$   $y_i$   $\rightarrow$  נתון  $x_i$   $\rightarrow$   $y_i$   $\rightarrow$  נתון  $x_i$   $\rightarrow$   $y_i$

$\{x_i\}_{i=1}^m \in \mathbb{R}^d$

נתון  $x_i$   $\rightarrow$   $y_i$   $\rightarrow$  נתון  $x_i$   $\rightarrow$   $y_i$   $\rightarrow$  נתון  $x_i$   $\rightarrow$   $y_i$

$\{(x_i, \hat{y}_i)\}_{i=1}^m$

נתון  $x_i$   $\rightarrow$   $y_i$   $\rightarrow$  נתון  $x_i$   $\rightarrow$   $y_i$   $\rightarrow$  נתון  $x_i$   $\rightarrow$   $y_i$

עין נתת סכום? נצטרך להוסיף  
CON המיון על לפי נתון כי להחיל  
עד יחסית את כל הנתונים על  
הנלכדים על איכות המידע

מיון:

Train data ← נתון למידה (פה יש יע)

Validation-data ← בדיקה של מידת קצב CON הנתון  
לפי בדיקה מיון (פה יש יע) יע יע יע יע יע

Test-data ← למידה את הנתון  
הנתון נתון לפי  $x_i$  את הנתון  $y_i$

מה 'קרה' עם את נתון Validation Set?

המחשב יתן תוצאות מוזנות באמצעות  
למידה שהם לא מבינים תוצאות עלולות

המחשב הישן נקרא Overfitting  
והיא קורא - כאן המידע מופץ טוב מ'3'!!

מיון:

יש לי מידע 20 פה

להחיל את המידע על המידע

data Scientist את יתן לי פה

מיון יתן יתן יתן יתן יתן

אם  $x_i \in \mathbb{R}^2$  אז  $y_i \in \mathbb{R}$

$$x_i \in \mathbb{R}^2 \leftarrow x_i \in \mathbb{R}^2$$

אם  $f$  היא פונקציה אז  $f(x_i) = y_i$



אם  $y_i \in \mathbb{R}$  אז  $x_i \in \mathbb{R}^2$

אם  $y_i \in \mathbb{R}$  אז  $x_i \in \mathbb{R}^2$

אם  $y_i \in \mathbb{R}$  אז  $x_i \in \mathbb{R}^2$