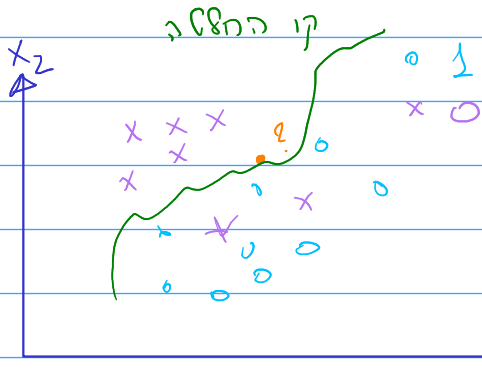


סיווג



מטרה: נתון חזש: האם היא 1 או 0?

\* 2 סוגי פתרונות:

\* מודל שמתזרם או 1

\* מודל שמתזרם הסתברות

← עדיף בין 0-1 שפסגת הסתברות

צוזאטא: אבטון רבול

בציקתם על איתותים

\* סוגי בעיות:

+ יש רק 2 אפשרות

- יש יותר מ-2

מטהם לרוב:

א רק 2 אפשרות

\* צוזאטאם על מודל מלפני

**k-Nearest Neighbors (k-NN)**

שיטה: מפתלים אל א שטתם הכי קרובים ומתחילים לם. כזה שיטה:

\* עוקבות מן חזש  $x_0$

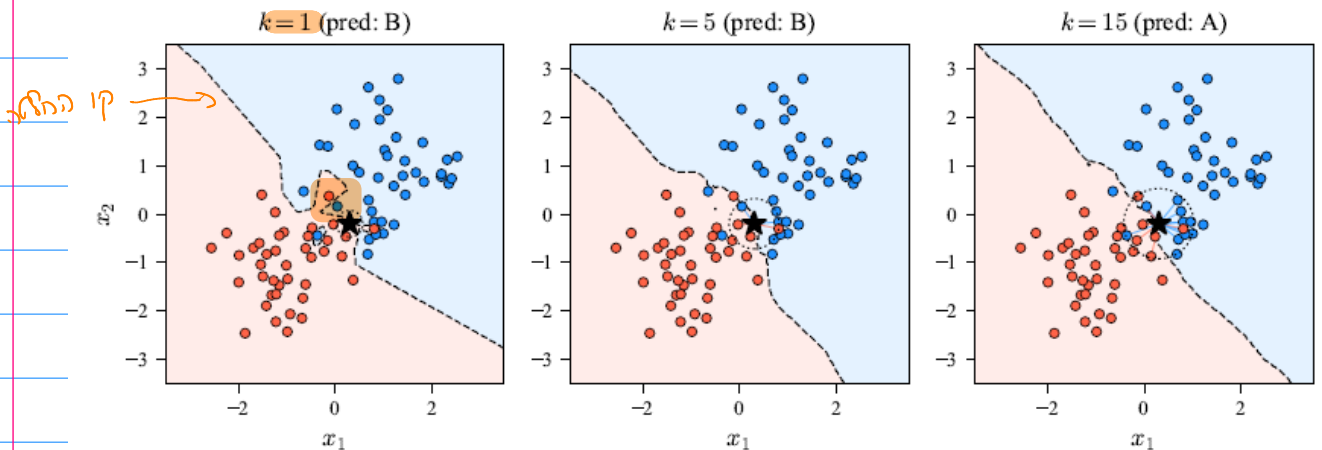
בסיס נתונים  $x_1, \dots, x_n$

\* מנתבים מרחקים:  $M$  מרחקים  $x_1 \rightarrow x_0, x_2 \rightarrow x_0, \dots, x_n \rightarrow x_0$

\* בוחרים א מרחקים הכי קרובים

\* התחלה עם קרוב במק א

כזה אחר מלפני

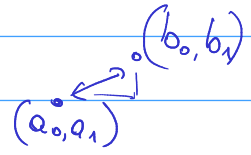


חישוב מרחקים:  $x_0 \leftrightarrow x_1$  איך מרחבים מרחקים?

Euclidean ( $L_2$ )

$$d(a, b) = \|a - b\| = \sqrt{\sum_{j=1}^N (a_j - b_j)^2}$$

Manhattan ( $L_1$ , city block)  $N=2$  \*  
 הישר בהשאלה ישר כווון



$$d(a, b) = \sum_{j=1}^N |a_j - b_j|$$

Minkowski (generalized, hyperparameter  $p$ )  $N=2$  \*

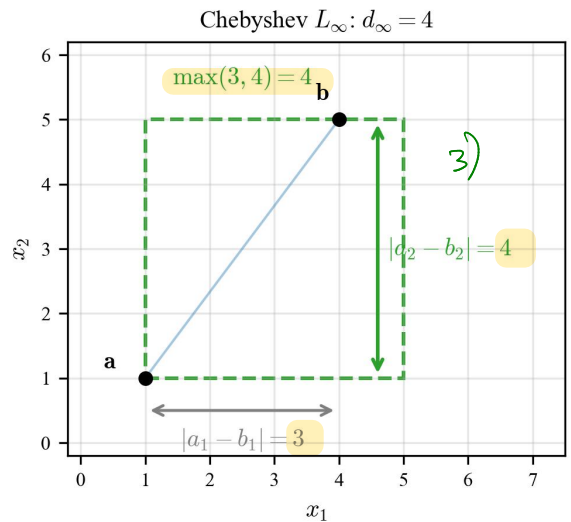
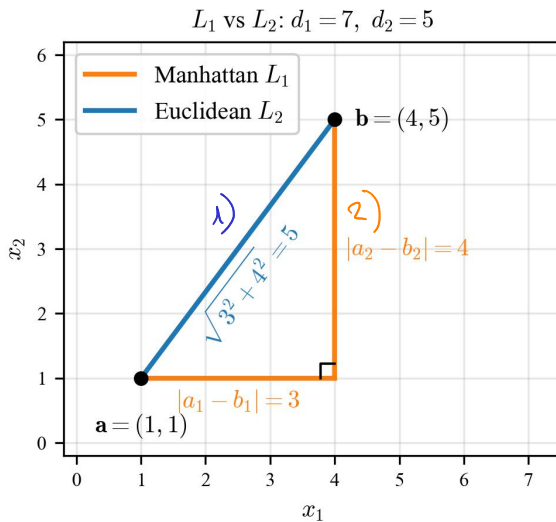
$$d(a, b) = \left( \sum_{j=1}^N |a_j - b_j|^p \right)^{1/p}$$

Special cases:  $p = 1$  gives Manhattan,  $p = 2$  gives Euclidean.

$p = \infty \max |a_j - b_j|$

\* שיקוף הארוך ביותר

על סכמה - מרחק מקסימלי למקף כל המרחבים



תיקו במקד א

\* א כוון עכור בעייה בינארית

\* בעייה עם יותר מ-2 סוגי מטרים



אקראי \* עמדה כן למחשבים. עוקמים  $k=1$  ← עפי שכן הכי קרוב

\* לאחד פשוט

\* עלף צורך באימון לקצרים

\* פרטת אינסטיטוביה

\* למינאם היפר-פרמטרים: 1) א

(2) אופן חישוב המרחק

\* נוח לעבודה מ/base (הערת ביצועים באופן)

עמה לא?

\* רגיש ל Outliers

\* גיועם יורדים עבור  $n$  (מספר נמנים) גבוה.

: curse of dimensionality

\* נמסרת ל מרחק נשתקף - הבדל בין מרחק

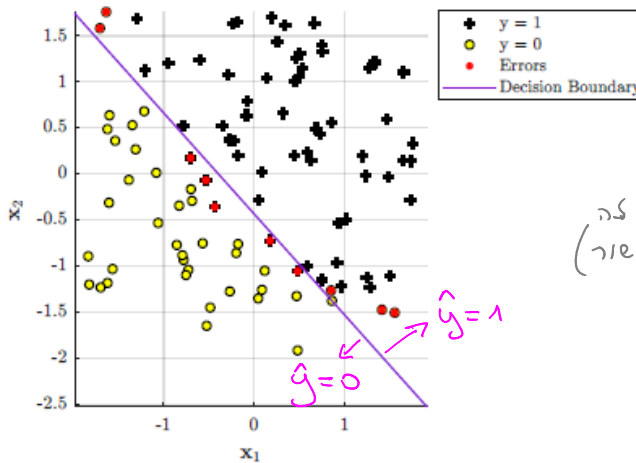
ל גודל וקטן למצבים.

\* פרמטר של  $k$ -M - אי הכי קרוב כנעם עיג

רוב הנעם הוא בקצוות ל האיזור

סיכומי: לא שמשו עבור הרבה נמנים

## Logistic Regression



\* סיטה עיטור = קו הפרדה עיטור.

\* פרמטר הסתבלות

שאלה למקרה: אק נוצאים לשווה ל

הקו הישר?

(ברג-נמנ נה

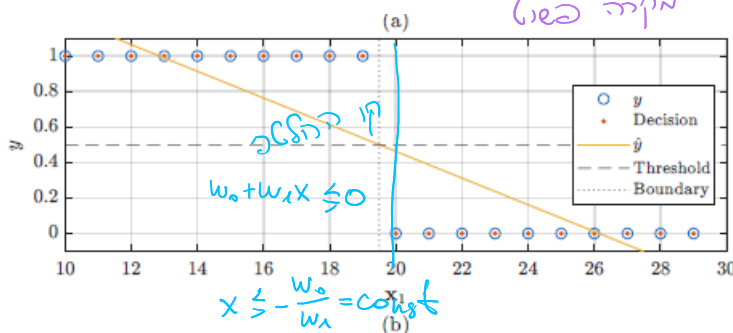
ברגע שיש קו (לקצמים של הקו) נישור)

קק-צמנ

$$w_0 + w_1 x_1 + w_2 x_2 + \dots \geq 0$$

קקצמה - רצוי עיטור. עסיוע (כנע; לא עיטור)

מקרה פשוט



צומת למסר רחב:

$$(x, y)$$

ל פנע ל אים אז

לא עוקב:

\* טצאת  $\hat{y}$  - יכח ערות ל מסר

כישמט?

רוסר

\* לקצמת מרותקת סיטתיה כנענות

לשפסמ יוקר לני

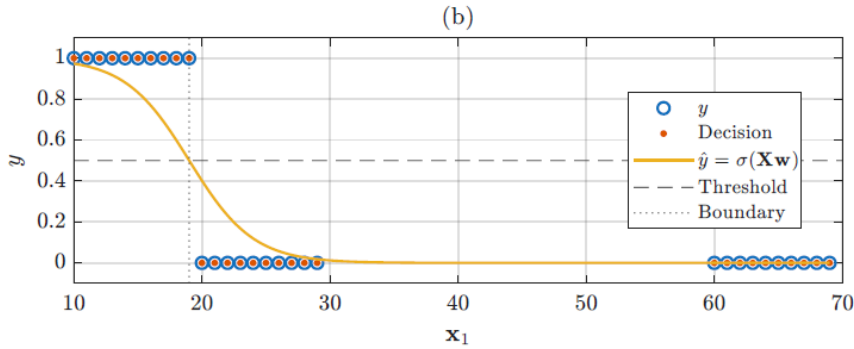
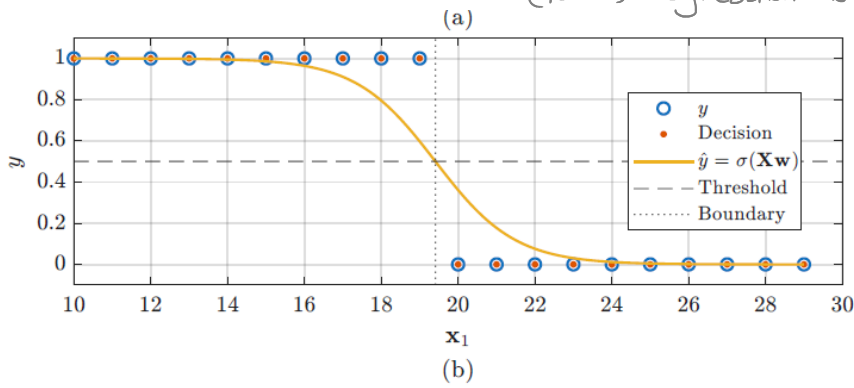
שיטה שיוצא אפילו בקטגורי (ר"ס)

logistic Regression

\*  $\hat{y} \in \{0, 1\}$

\* התחזית עם נקודות

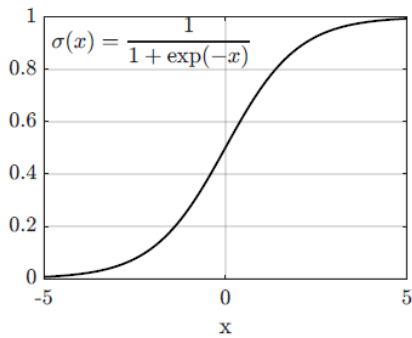
לחיתוך



שיטה

1. ניסוח המודל

$$\hat{y} = \begin{cases} 1 & \text{if } \sigma(w_0 + w_1 x_1 + \dots) > \text{threshold} \\ 0 & \text{otherwise} \end{cases}$$



2. איפון חיפוש הפרמטרים של המודל

חישוב אינדיקטור רגולרי

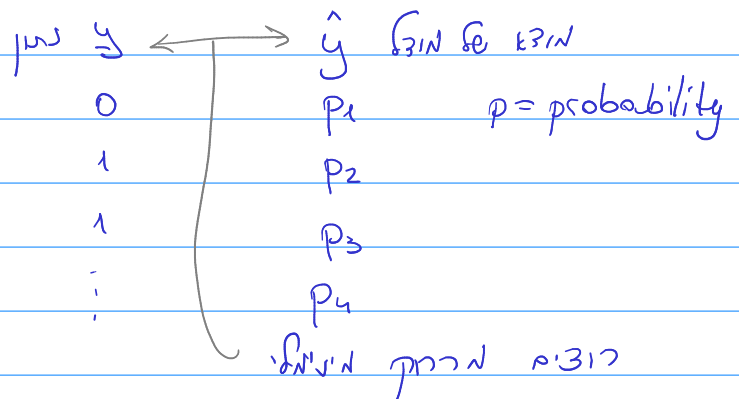
על סכום

$$\mathcal{L}(\cdot) = \frac{1}{2M} \|\hat{y} - y\|^2$$

Cross-entropy loss

מטרה: קצר הן על ידי התפלגות

- \* אין אופטימיזציה
- \* חיבה לקבוע אינדיקטור מקומי
- \* יש שיטה טובה יותר



← מקבלים בואוס: פשוט - הסתברות

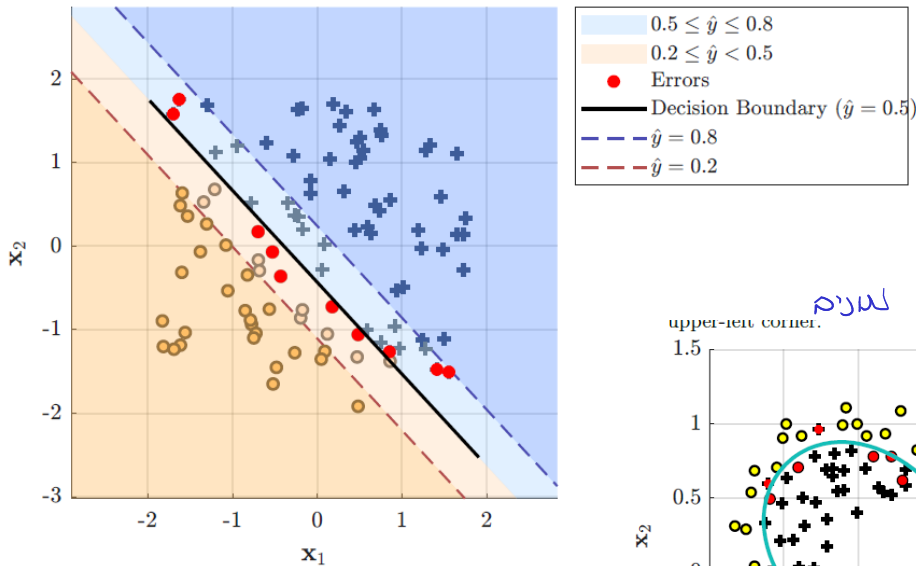
$$\hat{y} = \underset{y=1}{L^{(q)}} \text{ מקבלים } \text{הסתברות}$$

החלטת:  $\hat{y} \approx \frac{1}{2}$  לה שומר קרוב - זה לה פחות

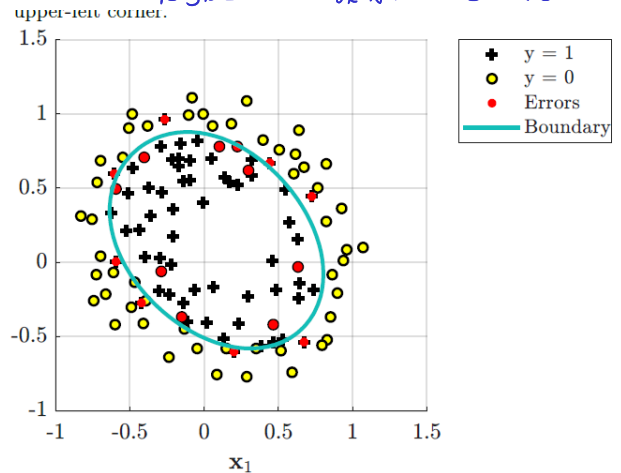
הערת Cross entropy (CE)

\* לשי

\* למתן עולם GD



מיפוי פולינומי של נתונים



← אפשרת עקמת במסוף

ליטר, בעיה שהא עינאני

עבור פ מציג יוצר אלפיסה

יפה גזר למצ

$$x_1, x_2 \rightarrow \varphi(x_1, x_2) = \langle 1, x_1, x_1^2, x_2, x_2^2, x_1x_2, x_1^2x_2, x_1x_2^2, x_1^2x_2^2 \rangle$$

פ מציג, 6 תצוי