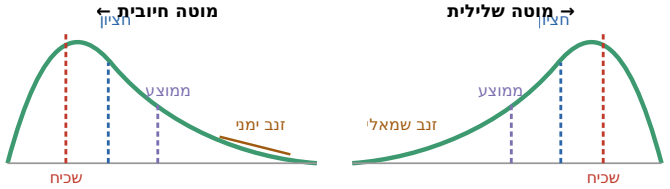
דף נוסחאות ומדריך פתרון · מבוא ללמידת מכונה ומדעי הנתונים (12112123)

1 · מדדי מרכז — ממוצע, חציון, שכיח

- **ממוצע** $= \Sigma x/n$ — **רגיש מאוד לחריגים**: ערך קיצוני בודד "מושך" אותו לכיוונו.
- **חציון** — הערך האמצעי אחרי מיון — **עמיד לחריגים** (כמעט לא זז). **שכיח** — הערך הנפוץ ביותר = פסגת ההיסטוגרמה.
- **הדגמה:** הנתונים [1, 2, 3, 4, 100] ← ממוצע = 22, חציון = 3. חריג אחד הזיז את הממוצע פי 7 מהחציון! לכן בנתונים מוטים (שכר!) מדווחים **חציון**, וגם משלימים ערכים חסרים בחציון.

2 · הטיה (Skew) — הנושא שפותח כמעט כל מבחן



הזנב מושך את הממוצע לכיוונו; השכיח נשאר בפסגה; החציון תמיד באמצע בין שניהם

- **מוטה חיובית (זנב ימני):** שכיח < חציון < ממוצע . דוגמאות: שכר, מחירי דירות, זמני המתנה, משקלי ארגזים עם מיעוט כבדים.
- **מוטה שלילית (זנב שמאלי):** ממוצע < חציון < שכיח . דוגמאות: ציוני מבחן קל, גיל פרישה.
- **סימטרית (וכל נורמלית):** שלושתם מתלכדים במרכז.
- **הכיוון ההפוך:** נתון "ממוצע > חציון" ← מוטה חיובית. "ממוצע = חציון" ← עקבי עם סימטריה — **אבל לא מוכיח נורמליות** (גם אחידה ודו-פסגתית סימטרית מקיימות!).
- **זיהוי מציור:** מחפשים לאן משתרך הזנב — לשם ההטיה. פסגה משמאל + זנב ימינה = חיובית (לא שלילית!).

נוסחאות בסיס

- **שונות:** $\sigma^2 = \Sigma(x_i-\mu)^2/n$ · סטיית תקן = √שונות. הדגמה: סטיות −2,0,2,4−,4 ← σ = √(40/5) = √8 = 2.83
- **מתאם פירסון:** $r = cov(x,y)/(\sigma_x \cdot \sigma_y)$, טווח [−1,1]. r=1 קו עולה מושלם · r≈0 אין קשר **לינארי** (ייתכן קשר אחר!) · **לא תלוי סקאלה**.
- **מתאם ≠ סיבתיות:** לחפש **משתנה שלישי**: גלידה↔כרישים (קיץ), תחנות כיבוי↔שריפות (גודל עיר).

3 · היסטוגרמה מנורמלת (סכום גבהים = 1)

מתכון קריאה:

1. מספר פריטים בעמודה שלמה = גובה × N . עמודה בגובה 0.15 עם N=200 ← 30 פריטים.
2. שאלה על **חצי עמודה** (למשל "כמה מעל 172.5" כשהעמודה 170–175) ← **לא ניתן לחשב!** נותנים רק חסמים: [סכום העמודות שמעבר, אותו סכום + העמודה החצויה].
3. **חציון:** סוכמים גבהים משמאל עד שמגיעים בדיוק ל-0.5 — שם החציון. אם 0.5 נופל בתוך עמודה — לא ניתן לקבוע מדויק.
4. ממוצע וסטיית תקן — **לעולם לא ניתנים לחישוב מדויק** מהיסטוגרמה (רק הערכה).

- **צורת הפיזור מספרת על השונות:** צורת U (מסה בקצוות) = שונות גבוהה ← רווח סמך **רחב**; פעמון צר = שונות נמוכה ← רווח צר. מדגם U אינו מנורמלי — אבל CI עדיין תקף (CLT).

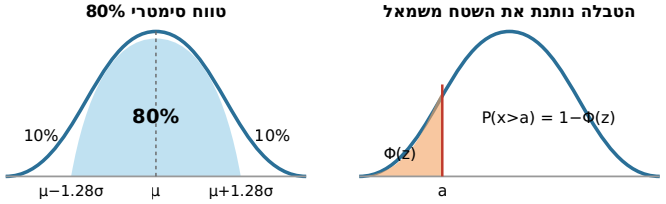
4 · סוגי משתנים

סוג	מבחן	דוגמאות
קטגורי נומינלי	אין שום סדר	קבוצת דם, צבע, עיר, סוג חיה, **ת"ז / מיקוד / מס' טלפון (!)**
קטגורי אורדינלי	יש סדר, אין "מרחק"	השכלה, מידות S/M/L, רמת חריפות
כמותי	חשבון משמעותי	משקל, גובה, שעות שינה, שכר

הכלל: מספר שמשמש כ"תווית" (אי אפשר לחבר/למצע אותו בהיגיון) = קטגורי. ת"ז היא המלכודת האהובה עליהם.

5 · התפלגות נורמלית N(μ, σ)

$$Z = (x - \mu) / \sigma$$



שמאל: טווח סימטרי — z לפי Z(0.5 + אחוז/2) · ימין: הטבלה תמיד מצטברת משמאל — ל"גדול מ" הופכים!

מתכון P(x < a):

1. תקנון: $z = (a-\mu)/\sigma$
2. טבלה: Φ(z). עבור z שלילי: $\Phi(-z) = 1 - \Phi(z)$
3. "גדול מ": $P(x>a) = 1 - \Phi(z)$ — **לא לשכוח להפוך!**
4. טווח [a,b]: $\Phi(z_b) - \Phi(z_a)$ · שני זנבות: מחברים אותם.

- **דוגמה מלאה (חזרה ב-3 מבחנים):** N(70,10), מתחת ל-50 או מעל 85: Φ(−2) + [Φ(1.5)−1] = 0.0228 + 0.0668 = **0.0896 ≈ 9%**
- **טווח סימטרי של X%:** $z = Z(0.5 + X/2)$ ← $\mu \pm z \cdot \sigma$. דוגמה 80%: N(70,10) ← Z(0.90)=1.28 נותן **[57.2, 82.8]**
- **טווח לא סימטרי:** מפרקים לצדדים: Y% משמאל ל-μ ← גבול שמאלי ב-X; Z(0.5−Y)% מימין ← Z(0.5+X). חלוקת X=Y+5 מתוך 85% ← Y=40, X=45. **בדיוק 50% מצד אחד ← הגבול ∞ (כל הזנב)!**
- **כלל אמפירי:** ±2σ ≈ 95% · σ ≈ 68% (בסטנדרטית: [−1.96, 1.96]) · ±3σ ≈ 99.7%
- **סטנדרטית:** μ=0, σ=1 ← שונות = סטיית תקן = 1 (מתלכדות). התלכדות ממוצע=חציון=שכיח — בכל נורמלית.

דוגמאות תקנון מלאות (N(70,10))

שאלה	תקנון	טבלה	תשובה
P(x<50)	z=−2	Φ(−2)	0.0228
P(x>85)	z=1.5	1−Φ(1.5)	0.0668
P(60<x<80)	z=∓1	Φ(1)−Φ(−1)	0.6826
מרכזי 80%	z=1.28	70±1.28·10	[57.2, 82.8]

מלכודות נורמלית: ① "גדול מ" ← z−1 ② Φ. שלילי ← Φ(−z)=1−Φ(z). ③ בדיוק חצי מצד אחד ← הגבול ∞. ④ התלכדות מדדים ≠ נורמליות.

Φ	0.85	0.875	0.90	0.925	0.94	0.95	0.975	0.985	0.995
Z	1.04	1.15	1.28	1.44	1.56	1.65	1.96	2.17	2.58
ביטחון	70%	75%	80%	85%	88%	90%	95%	97%	99%

CLT — התפלגות דגימה של ממוצעים

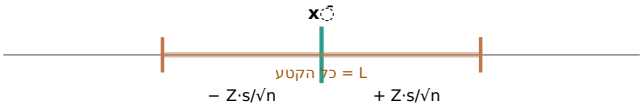
$$\bar{x} \sim N(\mu, \sigma/\sqrt{n}) \cdot SE = \sigma/\sqrt{n}$$

- התפלגות הממוצעים נורמלית **לכל** התפלגות מקור (גם אחידה/מוטה) — לכן מצורת הפעמון של הממוצעים **אי אפשר להסיק כלום על המקור**.
- עקומה **צרה** יותר ⇔ n **גדול** יותר. התוחלת נשמרת; סטיית התקן קטנה פי √n (לא נשארת σ, ולא הופכת סטנדרטית).

6 · רווחי סמך (CI) לתוחלת

$CI = \bar{x} \pm Z \cdot s/\sqrt{n}$ · $L = 2 \cdot Z \cdot s/\sqrt{n}$

(L = האורך הכולל של הרווח — פעמיים המרווח)



מתכון "[left, right] של X% מהשטח" (חזר 8+ פעמים!):

1. סימטרי: $z = Z(0.5 + X/2)$ ← $left = \mu - z \cdot \sigma$, $right = \mu + z \cdot \sigma$. דוגמה N(70,10), 80%: z=1.28 ← [57.2, 82.8].
2. לא-סימטרי (Y% משמאל, X% מימין): גבול שמאלי ב- $\mu - Z(0.5-Y) \cdot \sigma$, ימני ב- $\mu + Z(0.5+X) \cdot \sigma$ בהתאמה . **צד עם 50%** ← **∞.**

מתכון חישוב CI (4 שלבים — להראות הכל!):

1. אם נתונה **שונות** ← $s = \sqrt{\text{שונות}}$ (המלכודת מס' 1!)
2. Z לפי רמת ביטחון C: $Z((1+C)/2)$ — או ישר משורת "ביטחון" בטבלת עמוד 1. 99%← 2.58 · 90%← 1.65 · 85%← 1.44
3. מרווח: $Z \cdot s/\sqrt{n}$
4. תשובה: [$\bar{x}$ + מרווח, $\bar{x}$ − מרווח]

מחושב: שונות 2.5, $\bar{x}$=500.8, n=100, ביטחון 99%: s=1.58 ← מרווח 2.58·1.58/10 = 0.41 ← **[500.39, 501.21]**

מתכון גודל מדגם (אורך נדרש L):

1. $\sqrt{n} = 2 \cdot Z \cdot s/L$ ← מעלים בריבוע
2. **מעגלים תמיד למעלה!** (519 אריזות · 1153 ל-679 · IQ צמיגים)

מתכון הפוך (נתון L, מחפשים שונות):

1. $s = L \cdot \sqrt{n}/(2Z)$
2. מבקשים שונות? **להעלות בריבוע.** (L=0.25, n=729, ביטחון 95% ← s^2 ← 0.25·27/3.92=1.72≈**2.96**)

מלכודות CI: ① שונות נתונה ← שורש לפני הצבה. ② "אורך/רוחב" = פעמיים המרווח. ③ התפלגות המקור אחידה/מוטה? **לא מפריע** — CLT מציל במדגם גדול. ④ תקן/דרישה מחוץ לרווח ← לא ניתן להצהיר עמידה בתקן.

פרשנות נכונה — נשאל שוב ושוב

- ✓ "בביטחון C, **תוחלת** האוכלוסייה נמצאת בטווח."
- ✗ "C אחוז מהפרטים בטווח" · ✗ "פרט אקראי ייפול בטווח בהסתברות C" — הרווח על הממוצע **בלבד**, והוא מצטמצם עם √n; פיזור הפרטים רחב בהרבה.

מה מצמצם את הרווח

n ↑ רמת ביטחון ↓ (קטן Z) שונות ↓ — וההפכים מרחיבים.

אל תבלבלו CI עם טווח נורמלי! CI לתוחלת = $\bar{x} \pm Z \cdot \sigma/\sqrt{n}$ (עם √n). טווח סימטרי של X% מהאוכלוסייה = $\mu \pm Z \cdot \sigma$ (בלי √n!). שאלת "[left,right] של 80% מהשטח" = הטווח, לא CI.

7 · פייתון — דפוסים ומלכודות

- **מפתח מילון חייב hashable:** tuple/str/int · TypeError ← list ← תקין.
- **tuple = immutable:** השמה t[i]=x ← TypeError. list — מותר.
- **פרמטרים לפונקציה:** lst.append(...) / lst[i]= משנים את **המקור**; param = [...] **מנתק** — המקור לא משתנה. מחרוזות immutable — לעולם לא משתנות דרך פרמטר.
- "X".join(s.split("a")) שקול ל- s.replace("a","X")
- count(x)>1 = חוזרים · count(x)==1 = יחידאיים — לקרוא את התנאי בעיון!
- **לזהות מה הקוד עושה — סימנים מנחים:** בניית מספר הפוך (r = r*10 + num%10) ← פלינדרום/היפוך · r%a==0 and r%b==0 עד r++ :LCM · d[c1]!=c2 עם reversed ← צופן/מיפוי · set(list) בתוך comprehension ← הסרת כפילויות · key % len ← hash table.
- [x if cond for x in L] ← **SyntaxError**. תקין: [x for x in L if cond]
- **map** עם כמה iterables — איבר מקביל מכל אחד; חיבור מחרוזות = שרשור ללא רווח.
- **hash table:** אינדקס = key % N ; התנגשויות נערמות ברשימת התא.
- **set על tuple מסיר כפילויות!** set((1,3,4,5,5,2,1,1)) = {1,2,3,4,5} — קריטי לפני jaccard.

Pandas — groupby

- **שתי עמודות:** df.groupby(['a','b'])['c'].sum() — **רשימה!** ['a']['b'] ← שגיאה. איטרציה: for key, g in df.groupby(...) ; g.shape = (שורות, עמודות). ברירת מחדל: ממוין לפי מפתח, והמפתח הופך ל-**index**.
- **as_index=False** (נבחן!): שומר את עמודת הקיבוץ כעמודה רגילה במקום index — נוח לפני merge: df.groupby('class', as_index=False)['grade'].mean()
- **apply עם lambda** (חוזר): יצירת עמודה מחושבת — df['y'] = df['date'].apply(lambda d: int(d.split('/')[1])) (חילוץ חודש מ-DD/MM/YYYY).

Pandas — merge (תאר פלט בצורת dataframe!)

df1 (name, team)

Moshe · development
Adam · research
Emily · HR

df2 (team, est)

development · 2008
research · 2013
finance · 2010

how = ...

inner	רק המשותפים: Adam ,Moshe
left	כל df1: Emily+ עם est=NaN
outer	הכל: Emily+)NaN(finance+)NaN

שאלת ההמשך הקבועה: מדדים על העמודה אחרי merge

'left'how=, עמודת est: [2008, 2013, NaN]

.sum() = 4021.0 .count() = 2

count ו-sum מדלגים על NaN — לא 3!

פעולות קבוצה (set) — 6 הופעות

- A.intersection(B) = חיתוך · A.union(B) = איחוד · גודל החיתוך חלקי גודל האיחוד = jaccard על קבוצות.
- **set() מסיר כפילויות תחילה!** set((1,3,4,5,5,2,1,1)) = {1,2,3,4,5} (5 איברים, לא 8).

מתכון merge:

1. **inner** = חיתוך המפתחות (רק ערכים שקיימים בשני הצדדים).
2. **left** = כל שורות השמאלי; חסרים בימני ← **NaN**.
3. **outer** = איחוד; חסרים משני הצדדים ← NaN.
4. מדד על העמודה: **count מדלג על NaN** (סופר רק ערכים), **sum מדלג על NaN**. mean = sum/count (על הקיימים).

דפוס הזהב — "מעל הממוצע בקבוצה"

```
g = df.groupby('dept', as_index=False)['sal'].mean()
g.rename(columns={'sal':'avg'}, inplace=True)
m = df.merge(g, on='dept')
print(m[m['sal'] > m['avg']])
```

rename ← groupby (שלא יתנגש) ← merge (מצמיד ממוצע לכל שורה) ← סינון בוליאני. חוזר עם עובדים ועם תלמידים/כיתות.

קודים מוכרים (הפלט מאומת בהרצה)

קוד	מה עושה	פלט
set(...)+count>1	ערכים חוזרים בלבד	[1,2,5,9]
ראשוניים 2..19 (מסכה)	סכום	**77**
פלינדרום (בניית הפוך)	756 / 757	True / False
reversed + zip	צופן מראה a↔z	a→z, b→y
LCM (while)	lcm(7,6)	**42**
len(set(s))/len(s)	גיוון לקסיקלי	16/19
sorted(key=d.get)	תו שכיח	'e'

מתכון "מה מבצעת התוכנית (במשפט אחד)":

1. להריץ בראש על הקלט הנתון ולרשום את הפלט.
2. לנסח את **הכוונה**, לא את השורות: "סוכמת את הראשוניים בין 2 ל-20", "בודקת האם המספר פלינדרום" — משפט אחד ברמת מהות.

8 · רגרסיה לינארית

$MSE = (1/n)\cdot\Sigma(y - \hat{y})^2$ · $R^2 = 1 - SSE/SST$

- **R^2 = שונות מוסברת:** $1 - \Sigma(y-\hat{y})^2/\Sigma(y-\bar{y})^2$. ההגדרה נשענת על ניבוי רציף ← לא ישימה ללוגיסטית (שם pseudo-R^2).
- **מקדמי רגרסיה פשוטה:** $b1 = \Sigma(x-\bar{x})(y-\bar{y})/\Sigma(x-\bar{x})^2$ · $b0 = \bar{y} - b1\cdot\bar{x}$ ← כשאין קשר (רעש): b1=0 ו-**b0 = ממוצע y בדיוק**.
- **loss חלופי — הכלל:** מעניש כל שגיאה (בלי קיזוז) + קטן=טוב.

חלופה ל-MSE	תקין?	למה
|Δ| (MAE)	✓	חיובי, מעניש כל שגיאה
Δ^4	✓	זוגי, רגיש לחריגים
Δ^3	✗	סימנים מתקזזים
Δ^(−2)	✗	מתגמל שגיאות ענק

- **adjusted R^2:** R^2 עולה תמיד עם עוד משתנה (גם חסר תועלת). adjusted "מעניש" משתנים מיותרים — פער גדל בין השניים = מודל עמוס. הסרת משתנה לא-מובהק ← adjusted R^2 עשוי לעלות.

רעש: b1=0, ȳb0= · wise-piece: חיפוש חיתוך · U ← לא לינארי ✗ · לינארי ✓ · 0;gt&b1

לזהות את הצורה לפני שבוחרים מודל: U ← פולינומי/piece-wise (חיפוש נק' חיתוך שממזערת MSE כולל של שני הקטעים)

מתכון טבלת OLS (חזרה ב-6 מבחנים!):

1. לכל משתנה להסתכל **רק** על P>|t|: p<0.05 = מובהק ← הסרתו **תוריד** p. R^2 גדול = לא מובהק ← הסרה לא משפיעה; **adjusted R^2 אף עשוי לעלות**.
2. R^2 **לעולם לא עולה** מהסרת משתנה.
3. "למי ההשפעה הגדולה ביותר?" בלי scaling ← **לא ניתן לדעת!** (מקדם תלוי סקאלה). אחרי scaling — כן, לפי |מקדם|.
4. במטריצת **מתאמים** (פירסון) — כן משווים תמיד: |r| הגבוה מול y מנצח.
5. CI של β1 שכולל 0 (למשל [−3.11, 2.51]) ← אין עדות לקשר לינארי.

Multi-collinearity

- = מתאם חזק בין משתנים **בלתי תלויים** (x2–x4: 0.84 · שטח↔חדרים).
- **פוגע:** פרשנות מקדמים, סימן הפוך, מובהקות מזויפת (p גדול למרות קשר אמיתי).
- **לא פוגע:** R^2, חיזוי, test.
- זיהוי **קל** — מטריצת מתאמים. פתרון: להסיר את בעל המתאם הנמוך יותר עם y. scaling / dummy — לא פותרים.
- **דוגמה (2025א):** Y = 3.36 + 0.2·size − 0.52·bedrooms. הצבה (3, 1500): 3.36+300−1.56 = **301.8** ✓. "כל חדר מוריד 0.52" ✗ — מקדם שלילי בגלל collinearity; המודל תקף לניבוי, R^2 גבוה.

מתכון "האם מתאים לרגרסיה לינארית?" (מטבלת נתונים):

1. לצייר/לדמיין את הנקודות. מגמה מונוטונית (רק עולה או רק יורד) ← מתאים, b1≠0.
2. צורת U / הפוך-U (יורד ואז עולה) ← **לא מתאים** — קו ישר יפספס. (חשבון חשמל מול טמפ' = U קלאסי.)
3. אומדן "על מפית": b1 ≈ Δy/Δx מהקצוות · b0 ≈ y ההתחלתי פחות b1·x.

scaling · קידוד · חסרים · לוגיסטית

- scaling ברגרסיה: **לא** משנה את טיב המודל; **כן** משנה מקדמים (גם scaling של y). לא הכרחי — בניגוד ל-KNN/KMeans/רשתות.
- נומינלי ← **dummy (one-hot)**. קידוד אזורים 1-5 ← כופה סדר ומרחק מלאכותיים ✗ (כשל: מחיר גבוה באמצע הציר).
- ערכים חסרים בהתפלגות מוטה ← **חציון**.
- לוגיסטית: **sigmoid** σ(z)=1/(1+e^−z) ← הסתברות; אימון **log loss** (לזהות במבחן!): −[y·log(ŷ)+(1−y)·log(1−ŷ)] — לא MSE.
- **למה לא R^2 רגיל בלוגיסטית?** R^2 מבוסס ריבועי סטיות של ניבוי **רציף** סביב ממוצע; כאן התוויות 0/1 והפלט הסתברות ← ההגדרה לא משקפת טיב סיווג. לכן pseudo-R^2 (McFadden). הערכה נוספת: accuracy, confusion, precision/recall, ROC/AUC.
- **loss 0-1** "עובד" (סופר טעויות) אבל גס — נותן עונש זהה לטעות ב-0.49 ולטעות ב-0.01; log loss מבחין ולכן עדיף לאימון.

9 · סיווג והערכה

איזה מודל לפי צורת הפיזור? נתונים שזורים / טבעת / לא-לינארי ← **KNN, NNR, MLP**. ✓. רגרסיה לוגיסטית ← גבול **לינארי** בלבד ✗. רגרסיה לינארית ← לא מסווג כלל ✗. KMeans ← לא מונחה, לא מנבא מחלקות ✗.

	pred A	pred B	pred C
act A	**50**	10	10
act B	20	**50**	20
act C	5	5	**50**

עמודה אדומה = precision(A) = 50/75

שורה כחולה = recall(A) = 50/70

accuracy = אלכסון/הכל = 150/220

אותו תא אלכסון (50), מכנים שונים: precision = עמודה (מה שחזיתי) · recall = שורה (מה שאמיתי)

מתכון confusion:

1. סה"כ דוגמאות = סכום כל התאים (220).
2. precision(X) = תא האלכסון של X חלקי **סכום עמודת** predicted X.
3. recall(X) = אותו תא חלקי **סכום שורת** actual X.
4. "כמה דוגמאות B יש" = סכום שורת B (90).

- P, R ∈ [0,1] · סכומם על מחלקות לא חייב 1 · ΣR = 3 ייתכן (מסווג מושלם) · P+R ≤ 2 תמיד · **P = 1 לא גורר R = 1** (מסווג "שמרן").
- **accuracy לבדו לא מספיק:** מחלה של 1% — "תמיד בריא" נותן 99% ומחמיץ הכל (recall=0). שני מסווגים עם acc זהה ← נעדיף את המכסה את **כל** המחלקות.
- **F1-score** — הממוצע ההרמוני של precision ו-recall: F1 = 2·P·R / (P+R). מאזן בין השניים ב**מספר יחיד**; גבוה רק כששניהם גבוהים (אם אחד נמוך — F1 נמוך). שימושי כשהמחלקות לא מאוזנות, במקום להסתכל על accuracy מטעה. טווח [0,1], 1 = מושלם.
- **ROC/AUC:** TPR מול FPR על פני כל הספים. AUC ≈ 1 דירוג מושלם · 0.5 ניחוש · **שינוי סף לא משנה AUC**.

דוגמה מלאה (מהמטריצה למעלה, סה"כ 220):

precision(A)=50/75 · recall(A)=50/70 · precision(C)=50/80 · recall(C)=50/60 · accuracy=150/220 ≈ 0.68 · דוגמאות B בפועל = שורת B = 90.

שני מסווגים, אותו accuracy (5/10): מסווג שלא חוזה C אף פעם (recall(C)=0) נחות מזה שמכסה את כל שלוש המחלקות — **נעדיף כיסוי מלא**. זו הסיבה ש-accuracy לבד מטעה.

10 · מרחקים ו-KNN

מנהטן	Σ|Δᵢ|	אוקלידי	√(ΣΔᵢ²)
hamming	חלק המקומות השונים, מקום מול מקום — תמורה מלאה = 1.0. מחושב: ([1,1,0],[1,0,0]) = 1/3		
jaccard (sets)	גודל החיתוך חלקי גודל האיחוד — **1 = זהים**. מחושב: {1..5} מול {1,2,4,5,6,7,8} ← 4/8 = 0.5		
scipy.jaccard	**אי-דמיון** (0 = זהים): אי-הסכמה רק במקומות שלפחות אחד ≠ 0. מחושב: ([1,1,1,0],[1,0,0,0]) = 2/3		

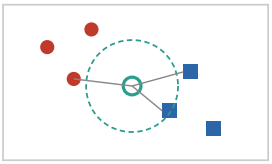
שאלת פיזור — התבנית הקבועה:

K=1 טועה (שכן בודד מטעה)

K=3 מתקן (רוב מקומי)

"קיים R עם דיוק מלא" ← לרוב נכון

מתאים ל-MLP ✓ · KMeans ✗ · לינארית ✗

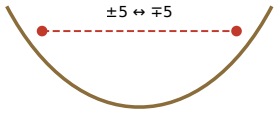


K=3: אדום 1, כחול 2 ← כחול · המעגל = NNR ברדיוס ?

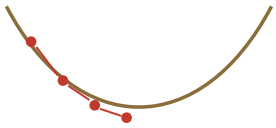
מתכון KNN מחושב (מנהטן, מ-(2,2), K=3):
אל (1,1)W=2 · (3,2)W=1 · B(2,5)=3 · B(1,5)=4 · (4,4)W=4 ← שלושה קרובים: B(3,2),(2,5)W,(1,1)W ← רוב **White**. (עם K=1 היה יוצא אחרת — לשים לב ל-K!)

- **K:** גדול ≠ טוב! |K=|train ← תמיד מחלקת הרוב. K קטן ← רגיש לרעש. בוחרים ב-validation. K אי-זוגי מונע תיקו **רק בבינארי** (1-2-2 בחמש מחלקות!).
- **scaling קריטי**; אבל הכפלה **אחידה** של כל הצירים ← לא משנה כלום (סדר שכנים נשמר). StandardScaler שונה-לכל-feature ← כן משנה. אחרי Standard מרחק **לא חסום** (מעל √N אפשרי); אחרי MinMax — חסום ב-√N.
- עמיד לחריגים · **ל-KNN אין שלב אימון** ← CE loss לא רלוונטי (בחירת K = כיוונון, לא אימון).
- **NNR:** רדיוס ענק ← מחלקת רוב (לא "יותר=מדויק"). features זהים בין datasets ← "ייתכן R1=R2" ✓, "בוודאות" ✗. אחרי scaling הרדיוסים ישתנו.
- **train/val/test:** אימון / כיוונון (K, R) / הערכה סופית נקייה.
- **CE:** $CE = -\log_{10}(p)$, p = הסתברות המחלקה האמיתית. p<0.1 ⇔ CE>1. דוגמה: אמת A, פלט [0.05, 0.90, 0.05] ← −log(0.05) = **1.30**. לא חסום מלמעלה.

11 · רשתות נוירונים ו-Gradient Descent



α גדול (=1 על x²): זגזג נצח



α מתאים: יורדים למינימום

מתכון צעד GD: ① נגזרת בנקודה: f'(x) ② צעד = α·f'(x) ③ צעד − x ← x . מקסימום? מחליפים ל-**+**. מחושב: x = 3 ← f'=3 , α=0.1, x=0 = y=x²+3x = **−0.3**. השלמת קוד: slope=derivative(x) (לא הפונקציה!) ← step=lr·slope ← x=x−step.

α	1 (על x², מ-5)	0.5	0.1	0.1−	0	גדול מאוד
קורה	זגזג ±5 נצחי	0 בצעד אחד	ירידה הדרגתית	מטפס ומתרחק	לא זז	התבדרות

- תלולה (20x²) מסוכנת ממתונה (2x²) עם אותו scaling ← α "מעגל" את נוף ה-loss ועוזר לאימון.
- **GD מול פתרון אנליטי:** יתרון — עובד כשאין פתרון ל-f'=0; חיסרון — מקורב, עלול להיתקע בלוקאלי.
- **Perceptron:** רק הפרדה לינארית — נכשל ב-XOR; מדרגה לא גזירה. הפתרון: MLP + אקטיבציות לא-לינאריות.

softmax · sigmoid

softmax(zᵢ) = e^zᵢ / Σe^zⱼ · σ(z) = 1/(1+e^−z)

- softmax: כולם ב-(0,1), **סכום 1**, **תמיד חיובי** (גם לקלט שליליי!), משמר סדר. מחושב: e^−1=0.368, e^1.5=4.48, e^7=1096.6, סכום 1101.5 ← [0.0003, 0.0041, 0.9956]
- חלופות פסולות: חלוקה בסכום (שליליים!) · ריבועים (הורס סדר: −3 מנצח את 1) · argmax קשיח (לא גזיר).
- **sigmoid: טווח (0,1) — לא [−1,1] (זה tanh)!** בינארי בלבד; לא לרציף. softmax — כל מספר מחלקות (בבינארי שקולה ל-sigmoid).
- **CNN — שיתוף משקלים:** פרמטרים כמספר תאי הפילטר + זיהוי feature בכל מקום. **LLM:** טקסט רהוט ✓, אומן על שיחות ✓, רשת כמסווג token הבא ✓, אחזור עובדות מדויק ✗ (הזיות!).

12 · אשכולות, PCA ו-Scaling

- **K-Means:** עצירה ב**התכנסות** (לא לפי גודל נתונים). רגיש לאתחול ← ריצות שונות = תוצאות שונות; טיפול: n_init (בוחרים inertia מזערי) + ++KMeans. קיים אתחול שמתכנס באיטרציה אחת. באגים בקוד: [0]shape · break · range(k) · previous=copy(current).
- **בחירת K:** לולאה + silhouette (מקסימום) או elbow על **inertia** = Σ‖x−c‖² — יורד תמיד, מחפשים את ה"שבירה".
- **rand index:** זוגות מוסכמים / n(n−1)/2 . שיטה: לספור את הלא-מוסכמים. [A,A,B,B,B] מול [0,0,1,1,2]: רק (3,5),(4,5) ← 8/10 = **0.8**
- **purity:** Σ(רוב בכל אשכול)/N. (3+4+2)/12 = **0.75**. חסם 1; טריוויאלי ב-K=N.
- **silhouette:** s=(b−a)/max(a,b) , ללא תיוגים — הפרדה גיאומטרית. שתי חלוקות נכונות זהות-rand יכולות לקבל s שונה. precision/recall — לא לאשכולות!
- **אלגוריתם סף-דמיון (תרגיל 4, נבחן פעמיים ב-2025!):** MS גבוה ← אשכולות **קטנים ורבים**; NUM=n ייתכן; MS1>MS2 ← בוודאות "NUM1>NUM2" ✗ (מגמה כן, ודאות לא). יתרונות מול KMeans: לא צריך K, סובלני לחריגים.
- **DBSCAN:** eps↑ ← מיזוג, רעש↓. הכל רעש ← eps↑ או min_samples↓; אשכול ענק ← ההפך. אין K.
- **PCA:** מתאם מושלם ← PC1=100%, PC2=0%. מתאם חזק ← כיוון הסכום מסביר את הרוב.
- **StandardScaler** (x−μ)/σ ← ממוצע 0, σ=1, לא חסום; פר-עמודה (axis=0!). מחושב: [1,3,5,7,9] ← [−1.41,−0.71,0,0.71,1.41]. **MinMax** (x−min)/(max−min) ← [0,1]: [−5,10,15,7] ← [0, 0.75, 1, 0.6]. "Standard מעביר ל-[0,1]" ✗!
- **מונחית** = יש תוויות (מחירים, ספרות); **לא מונחית** = מבנה בלי תוויות (פילוח, קיבוץ). **KMeans אינו מסווג!**

סיכום: פונקציית אימון והערכה לכל מודל

מודל	אימון (loss)	scaling חשוב?	לא-לינארי?
רגרסיה לינארית	MSE	לא (רק לפרשנות)	לא
רגרסיה לוגיסטית	log loss	כן (ב-GD)	לא (גבול לינארי)
KNN / NNR	**אין אימון!**	כן (מרחקים)	כן ✓
רשת (MLP)	cross-entropy	כן	כן ✓
KMeans	inertia (לא מונחה)	כן	—

מלכודות-על — לבדוק לפני הגשה: ① שונות נתונה ← שורש. ② אורך CI = 2×מרווח; n מעוגל למעלה. ③ ת"ז/מיקוד = קטגורי. ④ K sigmoid ∈ (0,1). ⑤ אי-זוגי רק בבינארי. ⑥ "ייתכן" כמעט תמיד נכון, "בוודאות" כמעט תמיד לא. ⑦ "הכי משפיע" בלי scaling ← לא ניתן לדעת. ⑧ P(x>a) — להפוך את ערך הטבלה. ⑨ KMeans אינו מסווג; ל-KNN אין אימון. ⑩ merge: count/sum מדלגים על R² ⑪ NaN. לא עולה בהסרה. ⑫ collinearity לא פוגע בחיזוי — רק בפרשנות.

מבנה המבחן (לפי 11 מועדים): ~14 שאלות · שעתיים וחצי · "מה יודפס + מה מבצעת (במשפט אחד)" · "נכון/לא נכון והסבירו" · "סמנו את **כל** הנכונות" · חישוב מלא עם דרך · שאלת פיזור · שאלת OLS. **טכניקה:** נכון/לא-נכון בלי נימוק = חצי נקודות; בחישובים להראות הצבה מלאה — נקודות חלקיות על דרך; "מה מבצעת" — משפט אחד ברמת כוונה.

כל החישובים אומתו בהרצת קוד · בהצלחה!