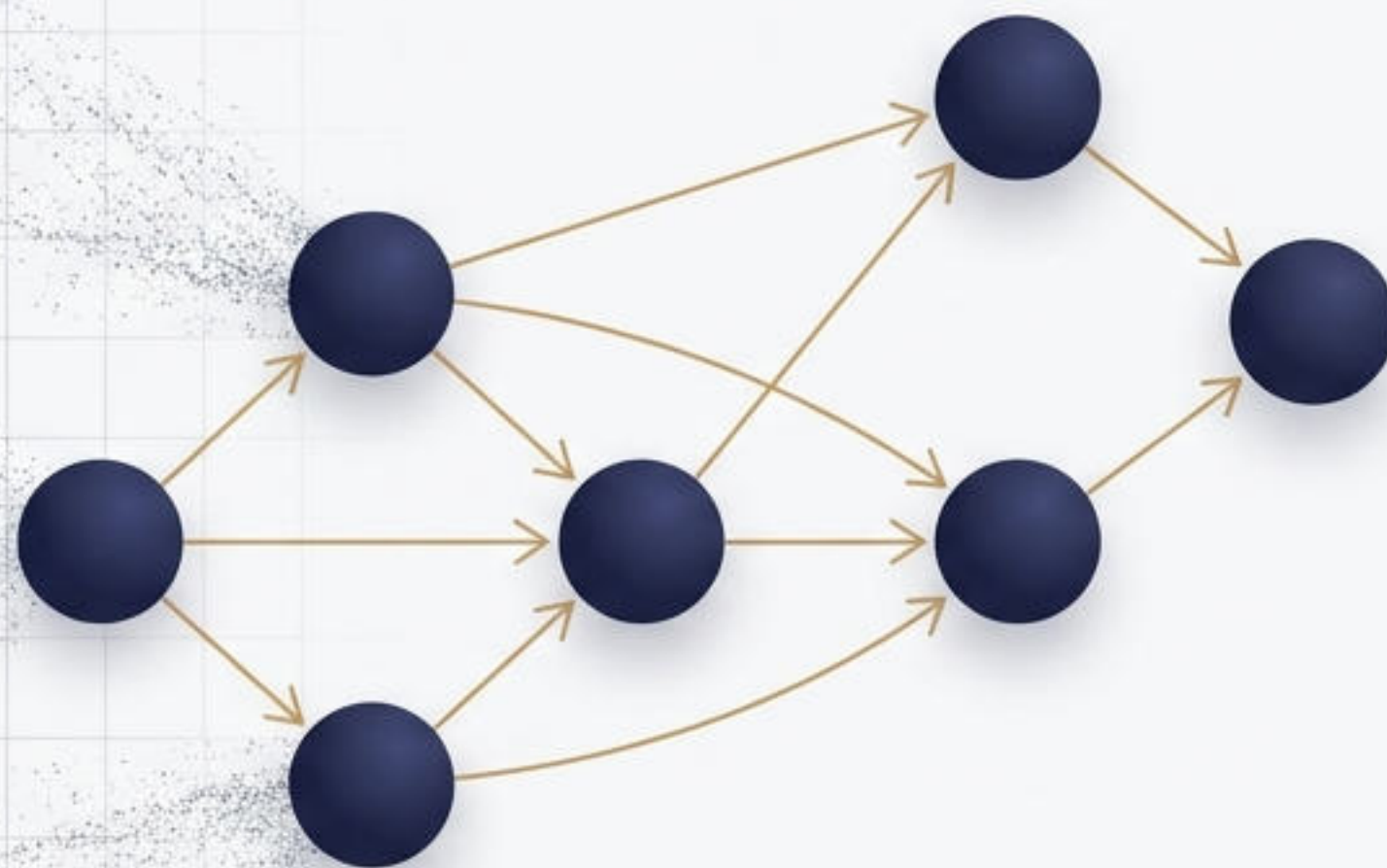




מדריך למידה ויזואלי

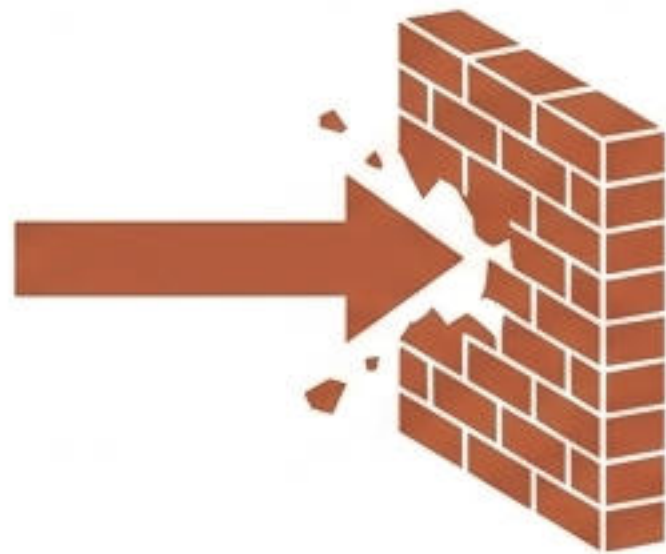
מבוא לבינה מלאכותית: מעבר מלוגיקה להסתברות

מרחב המדגם, חוק בייס, וארכיטקטורה של רשתות בייסיאניות

אשליית הוודאות: למה לוגיקה טהורה נכשלת

בעולם האמיתי, אין לנו את כל הנתונים. חיישנים רועשים, מידע חלקי ותוצאות בלתי בלתי צפויות מחייבים אותנו להתמודד עם אי-ודאות.

הסיכון בשקר (Risks Falsehood)



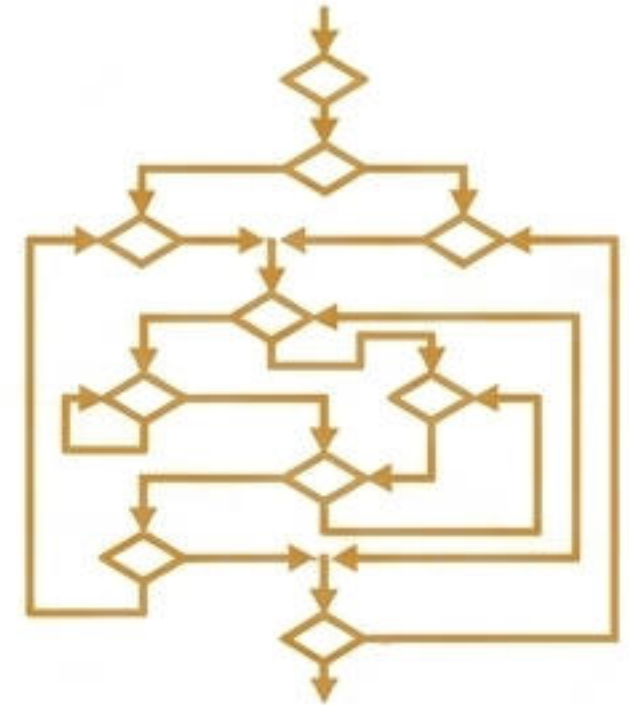
אם נניח שלצאת 25 דקות לפני הטיסה יביא אותנו בזמן, אנו מסתכנים בכך שההצהרה פשוט שגויה (בגלל פקק תנועה פתאומי).

לוגיקה טהורה קורסת כשעובדה אחת משתנה.

מסקנות חלשות (Too Weak)

אם ננסה להיות מדויקים לוגית: 'נגיע בזמן אם אין תאונה, ולא יורד גשם, והצמיגים תקינים...' – המסקנה הופכת לחלשה וארוכה מכדי לקבל החלטות מעשיות.

כדי להיות בטוחים ב-100%, נצטרך לצאת מהבית 24 שעות מראש.



מרחב המדגם והתפלגות משותפת (The Full Joint Distribution)

מרחב הסתברות
(Probability Model)
מקצה ערך בין 0 ל-1 לכל
"עולם אפשרי" (אירוע
אטומי), כך שסכום כל
ההסתברויות שווה ל-1.

	sunny	rain	cloudy	snow
Cavity = true	0.144	0.02	0.016	0.02
Cavity = false	0.576	0.08	0.064	0.08

טבלת ההתפלגות המשותפת מתארת את
ההסתברות של* כל אירוע אטומי* אפשרי במערכת
(למשל, חור בשן יחד עם מזג אוויר שמשי).

טבלת הקסם: כל שאלה על העולם שלנו יכולה להיענות מתוך הטבלה הזו, מכיוון שכל אירוע
הוא פשוט סכום של אירועים אטומיים ספציפיים.

הסתברות מותנית: מה קורה כשהמידע חלקי?

איפה ששני האירועים קורים יחד (החיתוך מתמטי).

המשמעות האמיתית:
'בהינתן ש- b הוא כל מה שאני יודע'.
(זה לא אומר 'אם b אז בוודאות a ').

$$P(a | b) = \frac{P(a \wedge b)}{P(b)}$$

'העולם החדש שלנו' -
אנו מחלקים ב- b כדי
לנרמל את ההסתברות
המדגם המצומצם שבו
אירוע b כבר התרחש.

חוק המכפלה (Product Rule) **נגזר ישירות**
ממשוואה זו: $P(a \wedge b) = P(a|b)P(b)$

הסקה על ידי מנייה (Inference by Enumeration)

שלב 1: חילוף הסתברות שולית (Marginalization)

כדי למצוא את ההסתברות ל- $\neg \text{cavity}$, אנו פשוט סוכמים את כל התאים בשורה הרלוונטית, ללא תלות בשאר המשתנים:

$$0.016 + 0.064 + 0.144 + 0.576$$

	toothache		$\neg \text{toothache}$	
	catch	$\neg \text{catch}$	catch	$\neg \text{catch}$
cavity	.108	.012	.072	.008
$\neg \text{cavity}$	.016	.064	.144	.576

שלב 2: חישוב מותנה

חישוב מותנה $P(\neg \text{cavity} \mid \text{toothache})$ מבוצע על ידי חלוקת החיתוך בסך כל מרחב הנתונים של כאב השיניים.

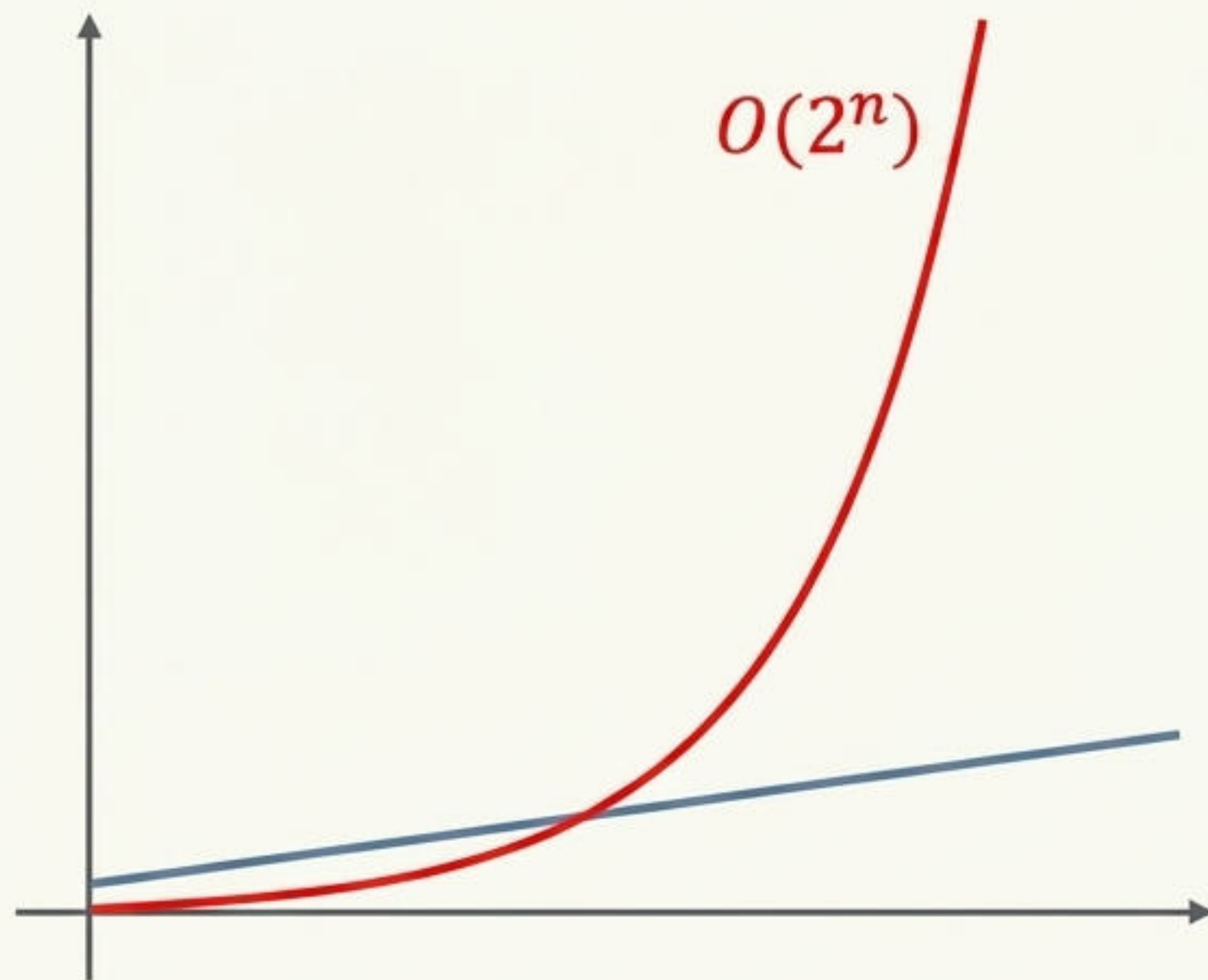
	toothache		$\neg \text{toothache}$	
	catch	$\neg \text{catch}$	catch	$\neg \text{catch}$
cavity	.108	.012	.072	.008
$\neg \text{cavity}$	.016	.064	.144	.576

שלב 3: הנרמול הקבוע (α)

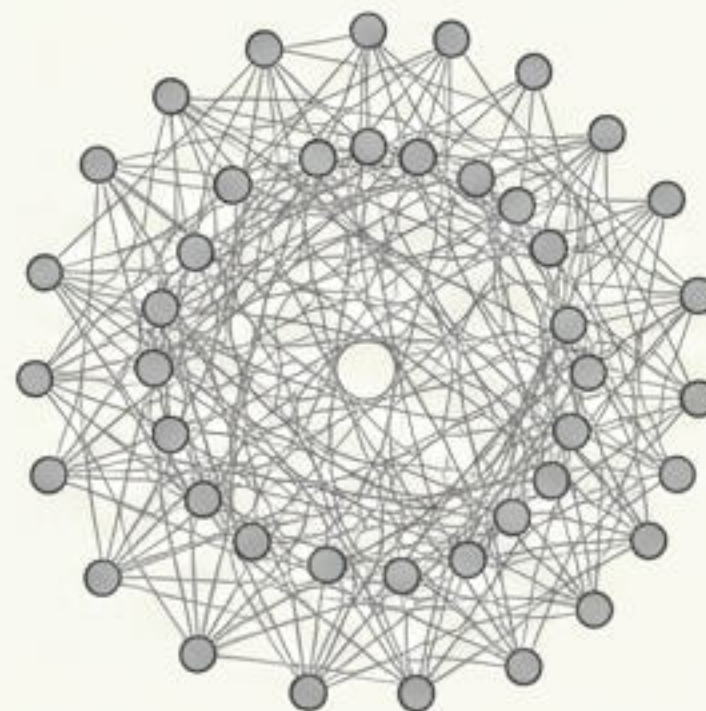
לעיתים קל יותר להתייחס למכנה כקבוע נרמול α .
הרעיון: מקבעים את משתני העדות (Evidence) וסוכמים על פני המשתנים הנסתרים (Hidden variables).

	toothache		$\neg \text{toothache}$	
	catch	$\neg \text{catch}$	catch	$\neg \text{catch}$
cavity	.108	.012	.072	.008
$\neg \text{cavity}$	.016	.064	.144	.576

משבר המורכבות: פיצוץ המידע



שימוש בטבלת ההתפלגות המשותפת הוא מושלם מתמטית, אבל בלתי אפשרי מעשית.
עבור n משתנים בוליאניים, גודל הטבלה הוא 2^n .



מקרה בוחן: רפואת שיניים.
תחום רפואת השיניים מכיל מאות משתנים (כאבים, חורים, דלקות, היסטוריה רפואית).
אף אחד מהם לא בלתי-תלוי לחלוטין בשני.

משבר נתונים: טבלה המייצגת 100 משתנים דורשת 2^{100} שורות. זהו מספר הגדול ממספר האטומים ביקום!
אנחנו חייבים דרך לדחוס את המידע האדיר הזה.

פריצת הדרך: אי-תלות ואי-תלות מותנית

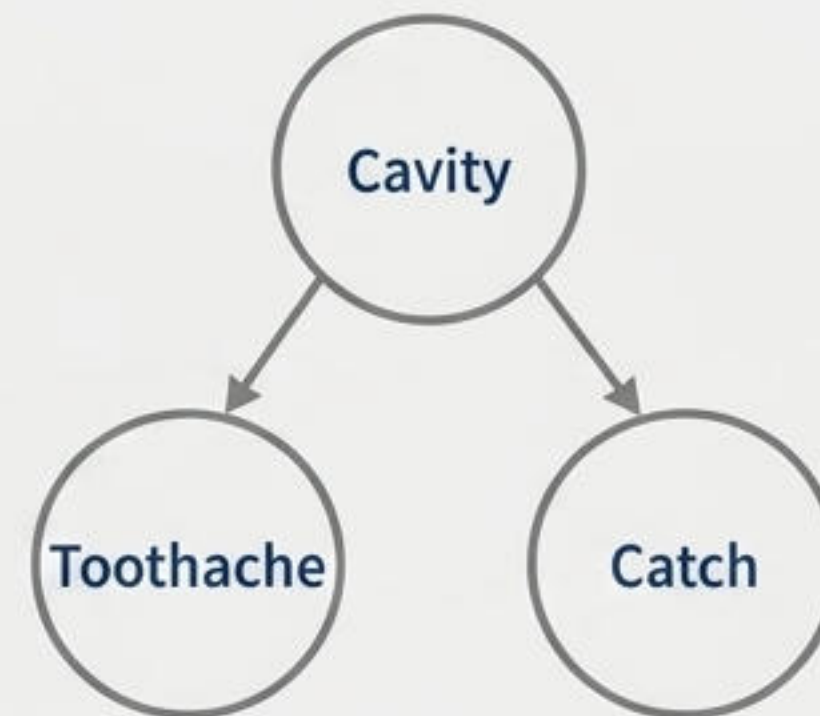
אי-תלות מוחלטת (Absolute Independence)



$$P(A,B) = P(A)P(B)$$

כלי חזק שמקטין דרסטית את כמות הנתונים, אך נדיר מאוד בעולם האמיתי.
(מזג האוויר בחוץ לא באמת קשור לכאב השיניים שלך).

אי-תלות מותנית (Conditional Independence)



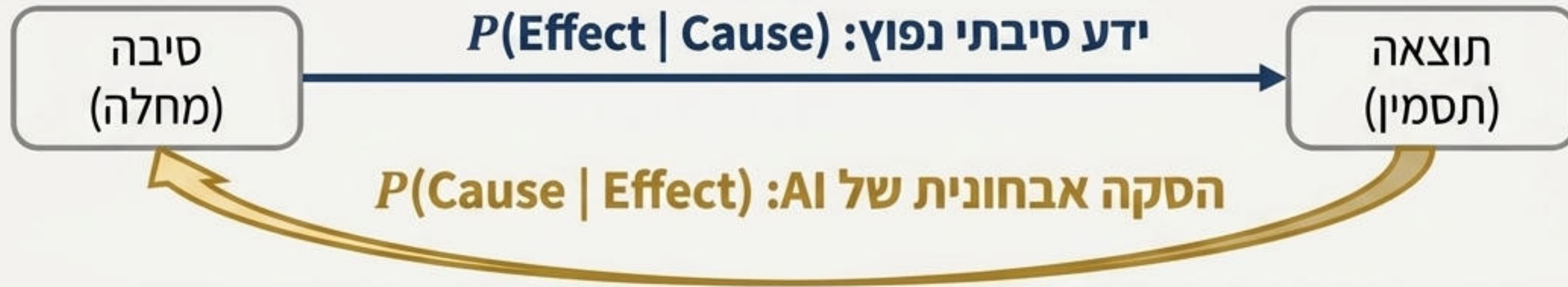
סוס העבודה של הבינה המלאכותית!
אם אני יודע שיש לי חור בשן (Cavity), ההסתברות שהרופא ימצא אותו (Catch) אינה תלויה יותר בשאלה אם כואבת לי השן (Toothache).

$$P(Catch | Toothache, Cavity) = P(Catch | Cavity)$$

החור בשן הוא ה'גורם המשותף' שמנתק את התלות המתמטית בין התסמינים.

חוק בייס (Bayes' Rule): להפוך את כיוון הסיבתיות

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$



דוגמת אבחון: דלקת קרום המוח (Meningitis)

רוב הידע הרפואי מתעד את הסיכוי של מחלה לייצר תסמין (דלקת קרום המוח גורמת לצואר נוקשה ב-80% מהמקרים). אבל רופאים צריכים את הכיוון ההפוך: מה הסיכוי לדלקת בהינתן צוואר נוקשה? באמצעות חוק בייס, אנו מחשבים שהסתברות האחורית (Posterior) נשארת זעירה (0.0008), כי המחלה עצמה עצמה נדירה מאוד באוכלוסייה ($P(m)$ קטן מאוד).

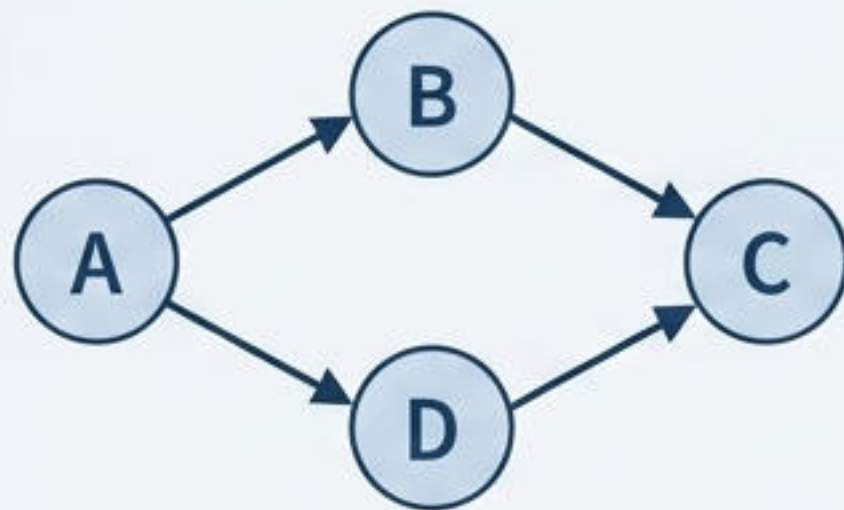
$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008 \quad (m = \text{meningitis}, s = \text{stiff neck})$$

הפתרון האולטימטיבי: רשתות בייסיאניות (Bayesian Networks)

ייצוג גרפי אלגנטי של אי-תלות מותנית, המאפשר הגדרה קומפקטית של התפלגות משותפת מלאה.

התחביר (Syntax) מורכב משני חלקים:

1. גרף מכוון חסר מעגלים (DAG)



קבוצת צמתים המייצגים משתנים, וקשתות המייצגות 'השפעה ישירה' או קשר סיבתי. **אסור** שתהיה לולאה סגורה.

2. טבלאות הסתברות מותנית (CPT)

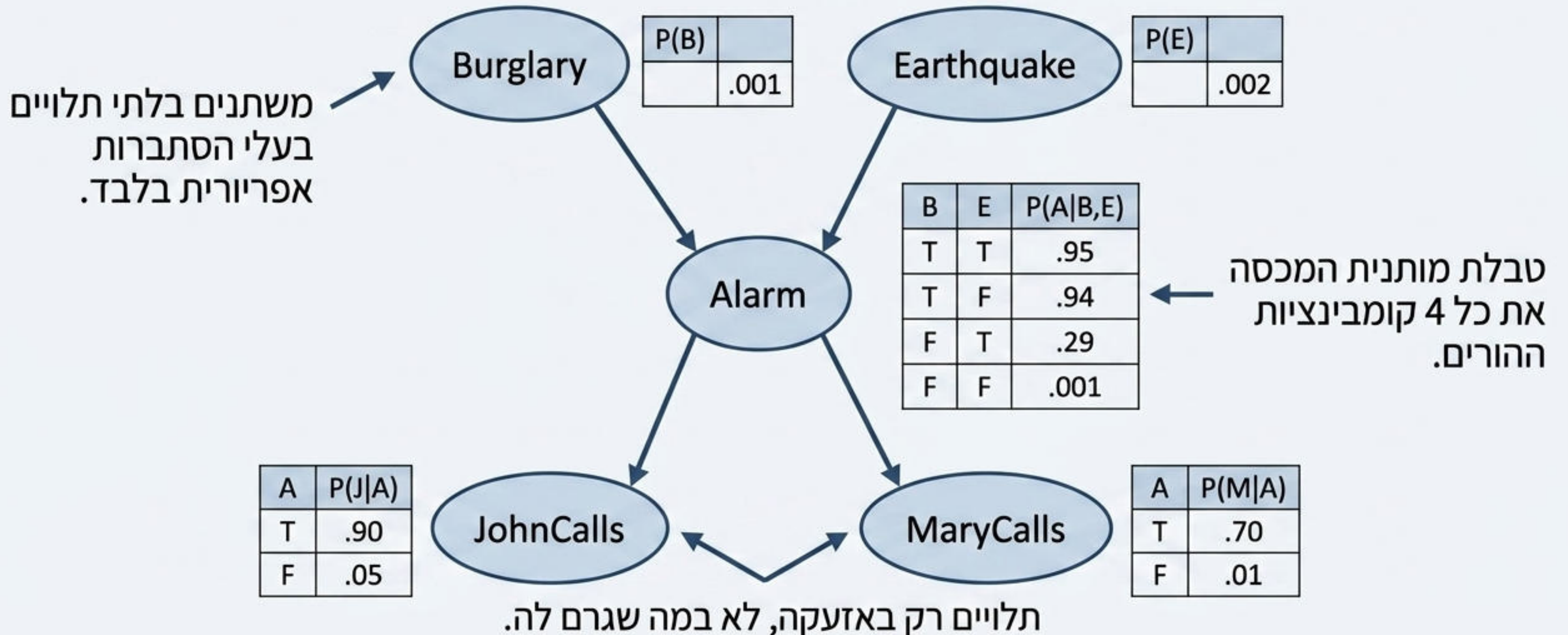
$P(X_i \text{Parents}(X_i))$		
A	C=T	C=F
T	0.8	0.2
F	0.3	0.7

לכל צומת יש טבלה קטנה $P(X_i | \text{Parents}(X_i))$ המפרטת את ההתפלגות שלו עבור כל שילוב אפשרי של ערכי הוריו.

הניצחון על המורכבות: במקום שהטבלה תתפוצץ למורכבות של $O(2^n)$, הרשת הבייסיאנית גדלה ליניארית: $O(n \cdot 2^k)$, כאשר k הוא מספר ההורים המקסימלי לצומת.

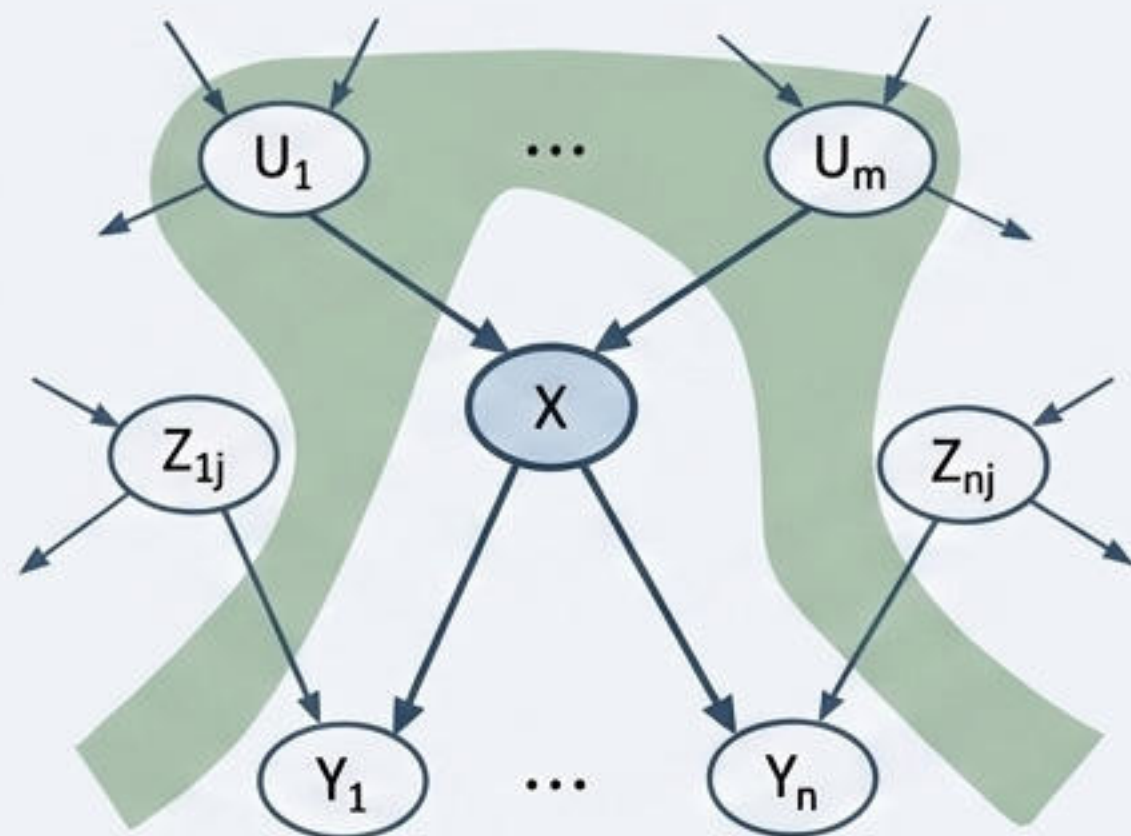
אנטומיה של רשת: דוגמת האזעקה המפורסמת

תרחיש: ג'ון ומרי יתקשרו אלי לעבודה אם האזעקה בביתי תופעל. האזעקה יכולה לפעול בגלל פריצה, אבל גם בגלל **רעידת אדמה קלה**.



כיצד הרשת מחשבת מידע? (סמנטיקה גלובלית ולוקאלית)

סמנטיקה לוקאלית (Local)

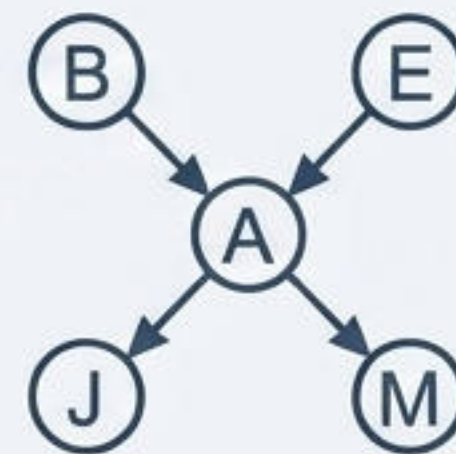


כל צומת בגרף הוא בלתי-תלוי מותנה בכל הצמתים שאינם צאצאיו, בהינתן ההורים שלו. ההורים מהווים '**מגן מתמטי**' החוצץ בין הצומת לשאר הרשת.

סמנטיקה גלובלית (Global)

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{Parents}(x_i))$$

הסמנטיקה הגלובלית מגדירה את ההתפלגות המשותפת המשותפת המלאה כפשוט **מכפלה** של כל ההתפלגויות **המותניות** המקומיות ברשת.



משפט מתמטי יסודי: הסמנטיקה הלוקאלית והגלובלית שקולות לחלוטין זו לזו.

בידוד מוחלט: שמיכת מרקוב (The Markov Blanket)

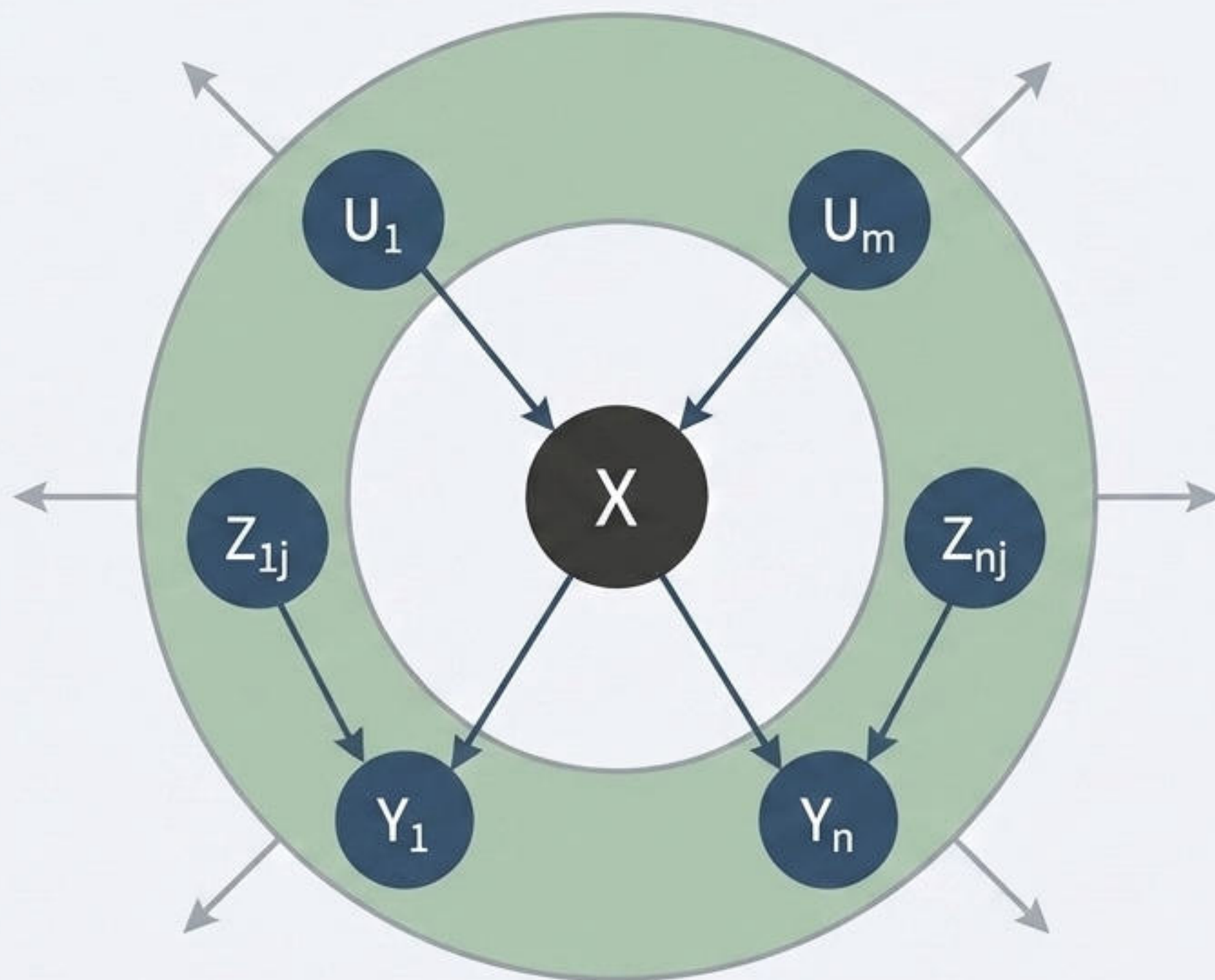
עקרון הבידוד האולטימטיבי:

כל צומת ברשת הוא בלתי-תלוי מותנה בכל שאר הצמתים ברשת, בהינתן 'שמיכת מרקוב' שלו.

אנטומיה של השמיכה:

1. ההורים שלו.
2. הילדים שלו.
3. ההורים של הילדים שלו (בני זוג).

תובנה: אם אנו יודעים בוודאות את הערכים של כל המשתנים בשמיכת מרקוב, שאר הרשת הופכת **למידע למידע 'לא רלוונטי'** לחלוטין עבור אותו צומת.



אלגוריתם: כיצד לבנות רשת בייסיאנית

אנו זקוקים לשיטה מובנית שבה סדרה של טענות **אי-תלות** מקומיות תבטיח לנו שהסמנטיקה **הגלובלית** תהיה מדויקת.

1. בחירת סדר בחר סדר לוגי של משתנים להכנסה לרשת: $X_1, \dots, X_n$.



2. בנייה איטרטיבית

→ עבור כל משתנה i מ-1 עד n :

- הוסף את הצומת X_i לרשת.
- בחר קבוצת הורים מינימלית מתוך המשתנים שכבר קיימים ברשת $(X_1 \dots X_{i-1})$ כך שיתקיים התנאי:

$$[P(X_i | Parents(X_i)) = P(X_i | X_1, \dots, X_{i-1})]$$



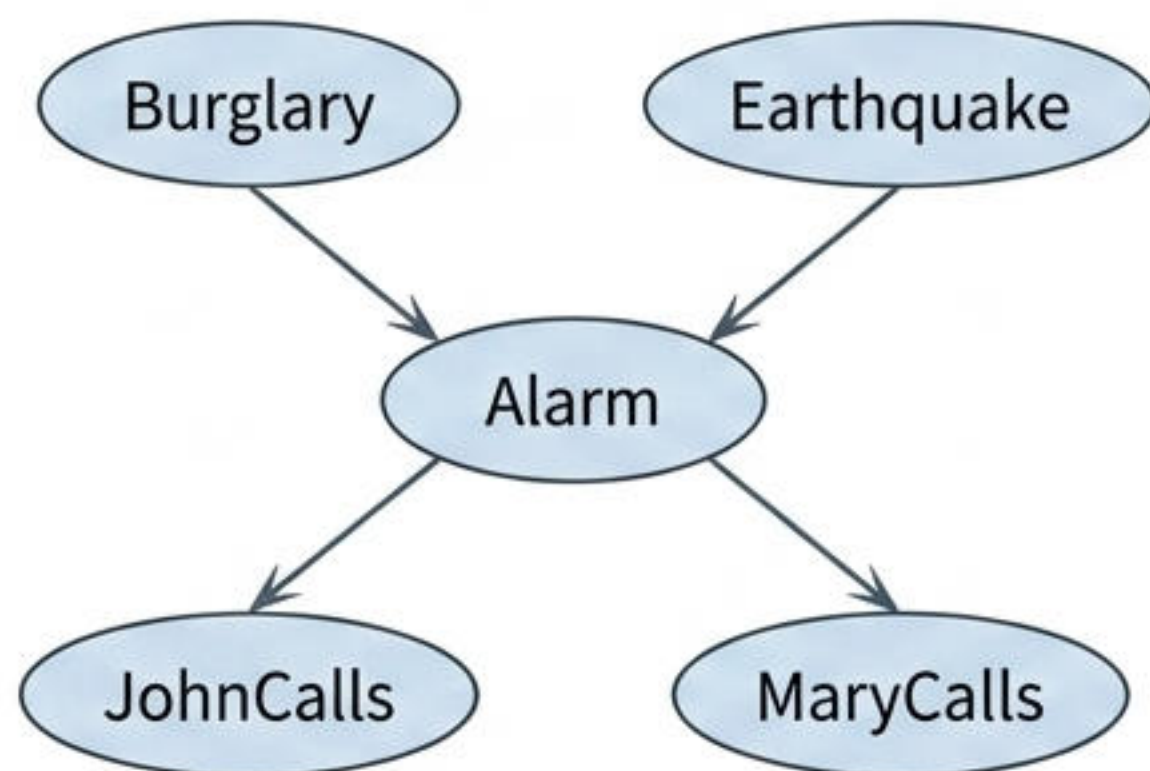
3. הבטחת תקינות

שימוש ב**חוק השרשרת** (Chain Rule) יחד עם בחירת ההורים הקפדנית הזו, מבטיח מתמטית שהרשת שניצר תייצג נכונה את ההתפלגות המשותפת המלאה.

חשיבות הסדר הסיבתי: כשהרשת מסתבכת

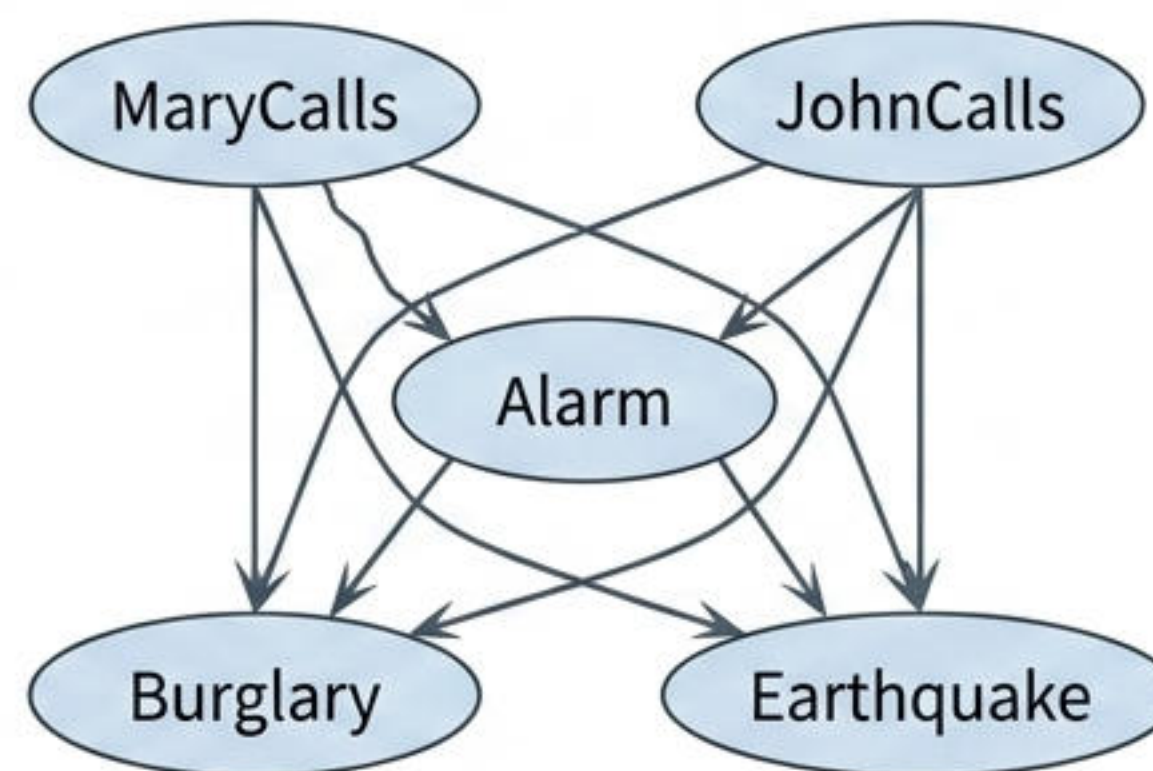
טופולוגיית הרשת תלויה לחלוטין בסדר הוספת המשתנים לאלגוריתם. חובה להוסיף משתנים מהסיבה אל התוצאה.

סדר סיבתי (הטוב)



כאשר בנינו לפי הסדר B, J, A, B, E (מהסיבה – פריצה – לתוצאה – ג'ון מתקשר), קיבלנו רשת אלגנטית הדורשת 10 פרמטרים בלבד. אי-התלויות המותניות נשמרות.

סדר הפוך (הרע)



אם ננסה לבנות בסדר הפוך, נאבד את אי-התלויות. נקבל גרף עמוס בקשתות מיותרות, טבלאות CPT ענקיות, וצורך ביותר מ-30 פרמטרים לאותה בעיה בדיוק!

אלגוריתם D-Separation: המבחן הסופי לאי-תלות

כיצד המחשב בודק אוטומטית האם 'X בלתי-תלוי ב-Y בהינתן קבוצת העדות Z'?
אין נמחשב הגרף.

אין צורך לחשב הסתברויות, רק לקרוא את הגרף.

