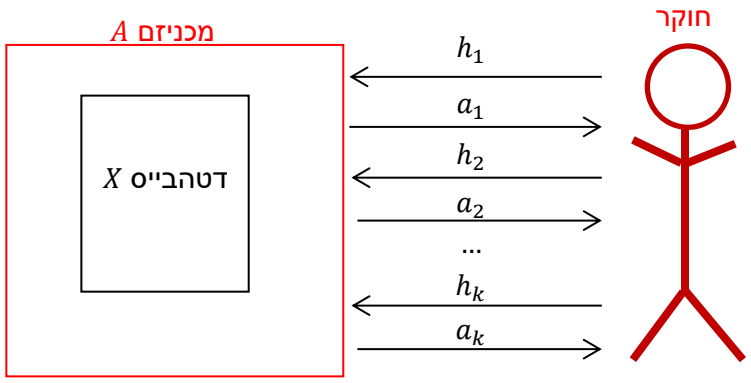
24.11.2022 ***פרטיות מידע***

הרצאה 5: קומפוזיציה

Textbook: Cynthia Dwork and Aaron Roth. The Algorithmic Foundations of Differential Privacy

מרצה: אורי שטמר

דוגמה למוטיבציה: מענה על שאילתות אדפטיביות



- נתון דטהבייס $X \in D^n$
- סדרה של שאילתות ספירה $h_1, h_2, \dots, h_k$ מגיעות <u>אחת אחת</u>
- לאחר כל שאילתא h_i עלינו להחזיר תשובה a_i

<u>המטרה:</u> לכל $i \in [k]$ מתקיים $\underbrace{|h_i(X) - a_i|}_{\substack{\text{הערך של השאילתא } h_i \\ \text{על הדטהבייס } X}} \leq \underbrace{\alpha}_{\substack{\text{פרמטר שגיאה} \\ \text{למשל } \alpha=0.1}}$ $\underbrace{a_i}_{\substack{\text{התשובה שלנו} \\ \text{לשאילתא מספר } i}}$

<u>השאלה:</u> על כמה שאילתות נוכל לענות (כלומר מהו k) כפונקציה של $\alpha, \varepsilon, \delta, n$?

<u>פתרון ראשון (אינטואיציה):</u>

- לכל שאילתא נענה בצורה המשמרת $\frac{\varepsilon}{k}$-פ"ד ולכן לפי קומפוזיציה נקבל בסה"כ ε-פ"ד עבור k שאילתות.
- כלומר לכל שאילתא נוסיף רעש מהתפלגות $\text{Lap}\left(\frac{k}{\varepsilon}\right)$, כלומר רעש מסדר גודל $\approx \frac{k}{\varepsilon}$, ולכן כדי לקבל שגיאה יחסית α נרצה שיתקיים $\frac{k}{\varepsilon} \lesssim \alpha n$, כלומר מספר השאילתות שנוכל לתמוך בהן בצורה הזאת הוא $k \lesssim \alpha \varepsilon n$

<u>פתרון שני:</u>

משפט הקומפוזיציה שלמדנו אומר שאם מריצים k אלגוריתמים המשמרים (ε, δ)-פ"ד אז בסה"כ מקבלים $(k\varepsilon, k\delta)$-פ"ד.

משפט קומפוזיציה חזקה:
יהיו $0 < \varepsilon, \delta \leq 1$. הפעלה אדפטיבית של k מכניזמים המשמרים (ε, δ)-פ"ד כל אחד (ללא גישה נוספת לדטהבייס) משמרת

$$\text{פ"ד}-\left(\sqrt{2k\ln\left(\frac{1}{k\delta}\right)}\cdot\varepsilon + 2k\cdot\varepsilon^2 \;\;,\;\; 2k\delta\right)$$

מסקנה: כדי לקבל אלגוריתם המשמר $(\tilde{\varepsilon}, \tilde{\delta})$-פ"ד, מספיק לדאוג שכ"א מ- k המכניזמים שנריץ ישמר

$$\text{פ"ד}-\left(O\left(\frac{\tilde{\varepsilon}}{\sqrt{k\ln\left(\frac{1}{\tilde{\delta}}\right)}}\right)\;,\;\frac{\tilde{\delta}}{2k}\right)$$

בדוגמה שלנו (מענה על שאילתות) נוכל לענות על כל שאילתא בעזרת הוספת רעש מהתפלגות $\text{Lap}\left(O\left(\frac{\sqrt{k\ln\left(\frac{1}{\delta}\right)}}{\varepsilon}\right)\right)$, כלומר עוצמת הרעש בכל שאילתא היא $\approx \frac{\sqrt{k\ln\left(\frac{1}{\delta}\right)}}{\varepsilon}$, ולכן כדי לקבל שגיאה יחסית α נרצה עכשיו

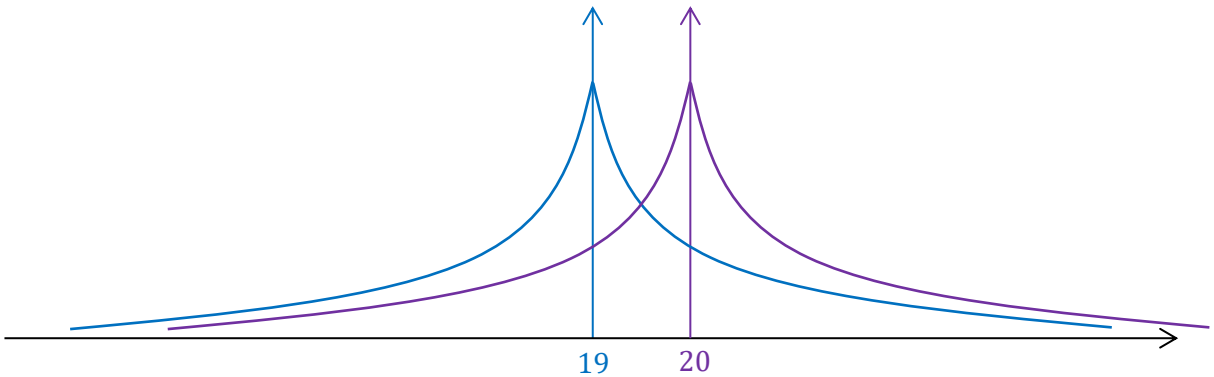
$$\frac{\sqrt{k\ln\left(\frac{1}{\delta}\right)}}{\varepsilon} \lesssim \alpha n$$

כלומר

$$k \lesssim \alpha^2\varepsilon^2 n^2 / \ln\left(\frac{1}{\delta}\right)$$

כלומר עכשיו k גדל כמו n^2 במקום כמו n !!!

אינטואיציה לגבי קומפוזיציה חזקה:
למה שהפרטיות תתדרדר רק כמו $\sqrt{k}$ ולא כמו k ?
נניח שאנחנו מריצים k פעמים את המכניזם הלפלאסי על קלט X או על קלט שכן X'. בכל ריצה נקבל דגימה מאחת מ- 2 ההתפלגויות הבאות:



- נניח שאומרים לי שהריצו את המ.לפלאס על אחד מ- X, X' (נניח הערך האמיתי של X הוא 19 והערך האמיתי של X' הוא 20), אבל אני לא יודע על מי מהם הריצו.
- אני רוצה לנסות לנחש האם הקלט הוא X או X'
- פלט > *19.5* "גורם לי לחשוב" ש- X' יותר סביר
- פלט < *19.5* "גורם לי לחשוב" ש- X יותר סביר
- אבל לפעמים נקבל פלט < *19.5* גם על X', מה שרק מטעה אותי בניסיון להבחין בין X ל- X' (כלומר לפעמים גם "נרוויח פרטיות")

עוד אינטואיציה: אז נניח שאנחנו מריצים את המ.לפלאס על X (כלומר דוגמים מההתפלגות הכחולה בציור הנ"ל). ברור שההסתברות שנקבל פלט קטן מ 19 היא בדיוק $1/2$. בנוסף, ההסתברות לקבל פלט בין 19 ל *19.5* היא בערך ε (אפשר לחשב את זה עם אינטגרל על פונק' הצפיפות). כלומר, ההסתברות שנקבל פלט באיזור בו הגרף הכחול גבוה מהסגול היא בערך $\frac{1}{2}+\varepsilon$, ובהסתברות בערך $\frac{1}{2}-\varepsilon$ נקבל פלט "שמטעה אותנו" שעבורו הגרף הסגול גבוה יותר.
מה יהיה מספר הפעמים שנקבל פלט "שלא מטעה אותנו"? זה משתנה מקרי עם התפלגות בינומית עם הסתברות הצלחה $\frac{1}{2}+\varepsilon$. התוחלת של מספר ההצלחות מתוך k דגימות היא $\frac{k}{2}+\varepsilon k$, כלומר בתוחלת יש באמת ייתרון מסויים לטובת "פלטים שלא מטעים אותנו". אבל סטיית התקן של משתנה מקרי כזה היא בערך $\frac{\sqrt{k}}{2}$. לכן אם $\varepsilon k \gg \frac{\sqrt{k}}{2}$ אז הייתרון שיש לנו "נבלע" בתוך סטיית התקן. כלומר, כל עוד $\varepsilon \ll \frac{1}{2\sqrt{k}}$ אז יהיה לנו מאוד קשה להבחין בין ההתפלגות הסגולה לכחולה.

מה זה אומר בדיוק שפלט הוא "מטעה" או "לא מטעה" אותנו? איך נכמת את זה?

<u>הגדרה:</u> עבור דטהבייסים שכנים X, X', עבור מכניזם M, ועבור פלט y נגדיר את <u>הפסד הפרטיות</u> כ-

$$L_M^{X\to X'}(y) = \ln\left(\frac{\Pr[M(X)=y]}{\Pr[M(X')=y]}\right) = -L_M^{X'\to X}(y)$$

ההגדרה הזאת תופסת את "רמת הביטחון" שיש לי בניסיון שלי לנחש מתוך הפלט y האם הדטהבייס הוא X או X'. אם $L_M^{X\to X'}(y) > 0$ אז הפלט y יותר סביר כשהקלט הוא X ולהיפך.

<u>אבחנה:</u> אלגוריתם $M: D^n \to Y$ משמר $(\varepsilon, 0)$-פ"ד אם ורק אם
$\forall y \in Y,\ \forall X, X'$ שכנים : $\left|L_M^{X\to X'}(y)\right| \le \varepsilon$

מה לגבי אלג' המשמר (ε, δ)-פ"ד ? האם גם עבורו $\left|L_M^{X\to X'}(y)\right|$ תמיד חסום ע"י ε ? לא. למשל אלג' שבהסת' δ פולט את כל הדטהבייס.

אוקיי... אז עם (ε, δ)-פ"ד אנחנו <u>לא</u> יכולים לצפות שהפסד הפרטיות שלנו <u>תמיד</u> יהיה חסום. למה כן אנחנו יכולים לצפות? <u>שבדרך כלל</u> הוא יהיה חסום! מה זה אומר "<u>בדרך כלל</u>"??

<u>טענת עזר 1:</u> יהי $M: D^n \to Y$ מכניזם. אם לכל X, X' שכנים מתקיים

$$\Pr_{y \leftarrow M(X)}\left[L_M^{X \to X'}(y) > \varepsilon\right] < \delta$$

אזי M מקיים (ε, δ)-פ"ד.

* מה שכתוב פה אומר שבשביל להראות ש- M מקיים (ε, δ)-פ"ד מספיק להראות שלכל זוג דטהבייסים שכנים X, X' מתקיים שבהסתברות לפחות $1 - \delta$, הפסד הפרטיות הוא לכל היותר ε.

<u>הוכחה:</u>
נקבע X, X' שכנים וקבוצה $F \subseteq Y$. צ"ל

$$\Pr[M(X) \in F] \leq e^{\varepsilon} \cdot \Pr[M(X') \in F] + \delta$$

<u>נגדיר:</u>

$$B = \left\{y \in Y : L_M^{X \to X'}(y) > \varepsilon\right\}$$

לפי התנאים של טענת העזר מתקיים

$$\Pr[M(X) \in B] < \delta$$

בנוסף, לפי הגדרת $L_M^{X \to X'}(y)$, לכל $y \notin B$ מתקיים:

$$\Pr[M(X) = y] \leq e^{\varepsilon} \cdot \Pr[M(X') = y]$$

ולכן

$$\begin{aligned}\Pr[M(X) \in F] &= \Pr[M(X) \in F \cap B] + \Pr[M(X) \in F \setminus B] \\ &\leq \delta + e^{\varepsilon} \cdot \Pr[M(X') \in F \setminus B] \\ &\leq \delta + e^{\varepsilon} \cdot \Pr[M(X') \in F]\end{aligned}$$

מ.ש.ל. (טענת העזר)

לשם פשטות, אנחנו נוכיח רק את הגרסה מוחלשת הבאה של משפט הקומפיזיציה. יש פה 2 הנחות משפטות: $(\varepsilon, 0)$-פ"ד והקומפוזיציה לא אדפטיבית.

<u>משפט [גרסה מוחלשת של משפט הקומפוזיציה החזקה]</u>
יהיו $0 < \varepsilon, \delta \leq 1$. הפעלה <u>לא-אדפטיבית</u> של k מכניזמים המשמרים <u>$(\varepsilon, 0)$-פ"ד</u> כל אחד (ללא גישה נוספת לדטהבייס) משמרת

$$\text{פ"ד-}\left(\sqrt{2k \ln\left(\frac{1}{\delta}\right)} \cdot \varepsilon + 2k \cdot \varepsilon^2 \ , \ \delta\right)$$

<u>הוכחת המשפט:</u>

יהיו $M_1, M_2, \dots, M_k$ מכניזמים המשמרים $(\varepsilon, 0)$-פ"ד כל אחד. נסמן ב- M את הקומפוזיציה הלא אדפטיבית שלהם, כלומר

$$M(X) = \big(M_1(X), M_2(X), \dots, M_k(X)\big)$$

אנחנו רוצים להראות ש- M מקיים $\big(\tilde{\varepsilon}, \tilde{\delta}\big)$-פ"ד.

לפי טענת עזר 1, מספיק להראות כי לכל X, X' שכנים מתקיים:

$$\Pr_{y \leftarrow M(X)}\left[L_M^{X \to X'}(y) > \tilde{\varepsilon}\right] < \tilde{\delta}$$

אז נקבע $y = (y_1, y_2, \dots, y_k) \in Y$ ונחשב:

$$L_M^{X \to X'}(y) = \ln\left(\frac{\Pr[M(X) = y]}{\Pr[M(X') = y]}\right)$$

$$= \ln\left(\prod_{i=1}^{k} \frac{\Pr[M_i(X) = y_i]}{\Pr[M_i(X') = y_i]}\right)$$

$$= \sum_{i=1}^{k} \ln\left(\frac{\Pr[M_i(X) = y_i]}{\Pr[M_i(X') = y_i]}\right) = \sum_{i=1}^{k} L_{M_i}^{X \to X'}(y_i)$$

- המטרה שלנו היא לנתח את $L_M^{X \to X'}(y)$ כמשתנה מקרי כאשר $y \leftarrow M(X)$ (לפי טענת עזר 1). מה שאנחנו רואים עכשיו זה שאת $L_M^{X \to X'}(y)$ אנחנו יכולים להציג כסכום של k משתנים מקריים <u>בלתי תלויים</u> (לפי ההנחה שהקומפוזיציה לא אדפטיבית).
- אנחנו כבר יודעים לפי האבחנה שהייתה לנו (ולפי ההנחה המפשטת) שלכל i ולכל y_i מתקיים $\left|L_{M_i}^{X \to X'}(y_i)\right| \le \varepsilon$.
- **<u>טענת עזר 2:</u>** מתקיים $\mathbb{E}_{y_i \leftarrow M_i(X)}\left[L_{M_i}^{X \to X'}(y_i)\right] \le 2 \cdot \varepsilon^2$

<u>"הוכחת" טענת עזר 2:</u>

לשם פשטות, נוכיח זאת ספציפית עבור המקרה של שני הלפלסים. כלומר נראה למה התוחלת קטנה מ $\varepsilon^2 \approx$ עבור הציור עם שני הגרפים כאשר אנחנו מניחים שהדטהבייס האמיתי הוא X עם ערך *19*:

- $L_{M_i}^{X \to X'}(y_i) = \varepsilon$ אם דוגמים ערך $y_i \le 19$
- $L_{M_i}^{X \to X'}(y_i) = -\varepsilon$ אם דוגמים ערך $y_i \ge 20$
- באמצע הוא בין לבין... נניח לרעתנו שמתקיים $L_{M_i}^{X \to X'}(y_i) = \varepsilon$ גם עבור $19 \le y_i \le 20$
- נשים לב שההסתברות להיות באמצע היא קטנה:

$$\Pr_{y_i \leftarrow M_i(X)}[19 \le y_i \le 20] = \int_1^2 \frac{\varepsilon}{2} \cdot \exp(-\varepsilon \cdot y)\, dy = \left[\frac{\varepsilon}{2} \cdot \frac{\exp(-\varepsilon \cdot y)}{-\varepsilon}\right]_1^2$$
$$= \frac{1}{2}(e^{-\varepsilon} - e^{-2\varepsilon}) \approx \frac{1}{2}\big((1-\varepsilon) - (1-2\varepsilon)\big) = \frac{\varepsilon}{2}$$

✶ ולכן

$$\mathbb{E}_{y_i \leftarrow M_i(X)}\left[L_{M_i}^{X\to X'}(y_i)\right] \le \Pr[y_i \le 19]\cdot\varepsilon + \Pr[19 < y_i < 20]\cdot\varepsilon + \Pr[y_i > 20]\cdot(-\varepsilon)$$
$$\approx \frac{1}{2}\varepsilon + \frac{\varepsilon}{2}\varepsilon - \left(\frac{1}{2} - \frac{\varepsilon}{2}\right)\varepsilon = \varepsilon^2$$

אז מה קיבלנו?

$$L_M^{X\to X'}(y) = L_{M_1}^{X\to X'}(y_1) + L_{M_2}^{X\to X'}(y_2) + \cdots + L_{M_k}^{X\to X'}(y_k)$$

כאשר כל ה $L_{M_i}^{X\to X'}(y_i)$ הם משתנים מקריים בלתי תלויים עם תוחלת לכל היותר $2\cdot\varepsilon^2$.

לכן התוחלת של $L_M^{X\to X'}(y)$ היא לכל היותר $2k\cdot\varepsilon^2$ (נובע מלינאריות התוחלת).

אז בעצם השאלה שאנחנו שואלים את עצמנו היא:

מה ההסתברות ש $L_M^{X\to X'}(y)$ יסטה ביותר מ $\sqrt{2k\ln\left(\frac{1}{\delta}\right)}\cdot\varepsilon$ מהתוחלת שלו?

כפי שראינו בשיעור שעבר, סכום של משתנים מקריים <u>בלתי תלויים</u> הוא מאוד מרוכז סביב התוחלת שלו. כלומר ההסתברות שסכום של משתנים מקריים יתרחק יותר מדי מהתוחלת שלו היא מאוד מאוד קטנה. ספציפית,

<u>חסם הופדינג:</u>
יהיו $X_1,\ldots,X_k \in [a,b]$ משתנים מקריים בלתי תלויים המקבלים ערכים בקטע $[a,b]$. אזי לכל $t \ge 0$ מתקיים

$$\Pr\left[\left|\sum_{i=1}^{k} X_i - \mathbb{E}\left[\sum_{i=1}^{k} X_i\right]\right| \ge t\right] \le 2\exp\left(-\frac{2t^2}{k\cdot(b-a)^2}\right)$$

במקרה שלנו, נקבל שההסתברות ש- $L_M^{X\to X'}(y)$ יסטה מהתוחלת שלו ביותר מ $\sqrt{2k\ln\left(\frac{1}{\delta}\right)}\cdot\varepsilon$ היא לכל היותר δ. כלומר

$$\Pr_{y\leftarrow M(X)}\left[\left|L_M^{X\to X'}(y) - \mathbb{E}\left[L_M^{X\to X'}(y)\right]\right| > \sqrt{2k\ln\left(\frac{1}{\delta}\right)}\cdot\varepsilon\right] < \delta$$

כלומר בהסתברות לפחות $(1-\delta)$ מתקיים

$$L_M^{X\to X'}(y) - \mathbb{E}\left[L_M^{X\to X'}(y)\right] \le \sqrt{2k\ln\left(\frac{1}{\delta}\right)}\cdot\varepsilon$$

כלומר

$$L_M^{X\to X'}(y) \le \mathbb{E}\left[L_M^{X\to X'}(y)\right] + \sqrt{2k\ln\left(\frac{1}{\delta}\right)}\cdot\varepsilon \le 2k\cdot\varepsilon^2 + \sqrt{2k\ln\left(\frac{1}{\delta}\right)}\cdot\varepsilon$$

לסיכום,

$$\Pr_{y\leftarrow M(X)}\left[L_M^{X\to X'}(y) > 2k\cdot\varepsilon^2 + \sqrt{2k\ln\left(\frac{1}{\delta}\right)}\cdot\varepsilon\right] < \delta$$

ולכן, לפי טענת עזר 1, המכניזם M הוא $(\tilde{\varepsilon},\delta)$-פרטי עבור

$$\tilde{\varepsilon} = 2k\cdot\varepsilon^2 + \sqrt{2k\ln\left(\frac{1}{\delta}\right)}\cdot\varepsilon$$

מ.ש.ל.

==========

סיימנו את ההוכחה של משפט הקומפוזיציה.
עכשיו אנחנו חוזרים לדבר על כלים שימושיים בהקשר של ניתוח פרטי של מידע

Noisy Argmax

<u>נתונים:</u> קבוצה של שאילתות ספירה C (כל $c\in C$ היא פרדיקט הממפה את הדומיין D לקבוצה $\{0,1\}$) ונתון דטהבייס $X\in D^n$.

<u>מטרה:</u> לזהות $c\in C$ עם ערך $c(X)=\sum_{x\in X}c(x)$ גדול ביותר.

<u>אלגוריתם *NoisyArgmax*:</u> נוסיף רעש $\text{Lap}\left(\frac{2}{\varepsilon}\right)$ לערך של כל שאילתא ונחזיר את האינדקס של השאילתא עם הספירה הרועשת הגדולה ביותר.

<u>טענה:</u> אלג' *NoisyAgrmax* משמר $(\varepsilon,0)$-פ"ד

<u>הוכחה:</u> נקבע X,X' שכנים, ונשים לב שמתקיים:

$$\forall c\in C:\quad |c(X)-c(X')|\le 1$$

נקבע $c\in C$. רוצים להראות כי

$$\Pr[A(X)=c]\le e^{\varepsilon}\cdot\Pr[A(X')=c]$$

נזכור כי לכל שאילתא $h\in C$ אנחנו מוסיפים רעש לפלאס, ונסמן $r_h\leftarrow\text{Lap}\left(\frac{2}{\varepsilon}\right)$.
עוד נסמן $\vec{r}_{-h}=$ כל הרעשים <u>חוץ</u> מהרעש עבור h.

מספיק להראות כי לכל קביעה של $\vec{r}_{-c}$ מתקים

$$\Pr[A(X)=c|\vec{r}_{-c}]\le e^{\varepsilon}\cdot\Pr[A(X')=c|\vec{r}_{-c}]$$

כי אז

$$\Pr[A(X)=c]=\sum_{\vec{r}_{-c}}\Pr[\vec{r}_{-c}]\cdot\Pr[A(X)=c|\vec{r}_{-c}]\le\sum_{\vec{r}_{-c}}\Pr[\vec{r}_{-c}]\cdot e^{\varepsilon}\cdot\Pr[A(X')=c|\vec{r}_{-c}]=e^{\varepsilon}\cdot\Pr[A(X')=c]$$

אז נקבע וקטור רעש $\vec{r}_{-c}$. עכשיו מתקיים:

$$\Pr[A(X)=c|\vec{r}_{-c}] = \Pr\left[c(X)+\text{Lap}\left(\frac{2}{\varepsilon}\right) > \max_{h\neq c}\left\{h(X)+r_h\right\}\right]$$

$$= \Pr\left[c(X)+\text{Lap}\left(\frac{2}{\varepsilon}\right)+1 > \max_{h\neq c}\left\{h(X)+1+r_h\right\}\right]$$

$$\leq \Pr\left[c(X)+\text{Lap}\left(\frac{2}{\varepsilon}\right)+1 > \max_{h\neq c}\left\{h(X')+r_h\right\}\right]$$

$$\leq \Pr\left[c(X')+\text{Lap}\left(\frac{2}{\varepsilon}\right)+2 > \max_{h\neq c}\left\{h(X')+r_h\right\}\right]$$

$$\leq e^{\varepsilon}\cdot \Pr\left[c(X')+\text{Lap}\left(\frac{2}{\varepsilon}\right) > \max_{h\neq c}\left\{h(X')+r_h\right\}\right]$$

$$= e^{\varepsilon}\cdot \Pr[A(X')=c|\vec{r}_{-c}]$$

מ.ש.ל.

The Sparse Vector Technique (SVT)

נתון דטהבייס $X\in D^n$ ונתון ערך $t\in\mathbb{R}$. סדרת שאילתות ספירה מגיעות אחת אחת $f_1,f_2,f_3,\dots$, כאשר לכל i מתקיים ש- f_i היא פרדיקט $f_i: D\to\{0,1\}$, והערך של f_i על X הוא $f_i(X)=\sum_{x\in X} f_i(x)$.

<u>המטרה:</u>

- נרצה לקבל הערכה ל- $f_i(X)$ לכל i כך ש- $f_i(X)\gtrsim t$
- עבור שאילתות f_i כך ש- $f_i(X)\lesssim t$ אנחנו לא מעוניינים לקבל תשובה

<u>הרעיון:</u> אם מספר השאילתות "המעניינות" הוא קטן, אז בתקווה הפסד הפרטיות שלנו יהיה תלוי רק במספר השאילתות המעניינות ולא במספר כל השאילתות.

<u>הנחות מקלות:</u>

(1) במקום לתת הערכה ל- $f_i(X)$ עבור f_i "מעניינות", רק נציין <u>מיהן</u> השאילתות המעניינות. ההנחה הזאת היא בה"כ כי אם נדע לזהות מיהן השאילתות המעניינות אז נוכל לחשב עבורן הערכה עם המ.לפלאס (ואז הפסד הפרטיות יהיה תלוי רק במס' השאילתות המעניינות)

(2) נניח שיש רק שאילתא מעניינית <u>אחת</u>. זה גם בה"כ כי אחרי שנזהה שאילתא מעניינת נוכל להריץ שוב את האלג'

כלומר, אנחנו רוצים אלגוריתם המחזיק דטהבייס $X\in D^n$ ומקבל פרמטר t וסדרת שאילתות $f_1,f_2,\dots$.
לאורך הריצה, כאשר מקבלים שאילתא f_i:

- אם $f_i(X)\lesssim t$ אזי מחזירים פלט $\perp$ וממשיכים לשלב הבא
- אם $f_i(X)\gtrsim t$ אזי <u>עוצרים</u> ומחזירים $\top$.

(כלומר זהו תהליך שמזהה את השאילתא הראשונה עם ערך $\lesssim t$)

***Algorithm* AboveThreshold**$(X \in D^n, \{f_i\}, t, \varepsilon)$
Let $\hat{t} \leftarrow t + \text{Lap}\left(\frac{2}{\varepsilon}\right)$ *For each query* i *do* *Let* $v_i \leftarrow \text{Lap}\left(\frac{4}{\varepsilon}\right)$ *If* $f_i(X) + v_i \geq \hat{t}$ *then output* $a_i = \top$ *and halt* *Else output* $a_i = \perp$ *and continue*

<u>משפט:</u> אלגוריתם AboveThreshold מקיים $(\varepsilon, 0)$-פ"ד
(לא משנה כמה שאילתות הוא קיבל לפני שהוא עצר...)

<u>הוכחה:</u>
נקבע שני דטהבייסים שכנים X, X' ונקבע סדרת שאילתות $f_1, f_2, \ldots$.
נסמן $A(X) = \text{AboveThreshold}(X, \{f_i\}, t, \varepsilon)$ ונשים לב שפלט האלגוריתם הוא תמיד מהצורה

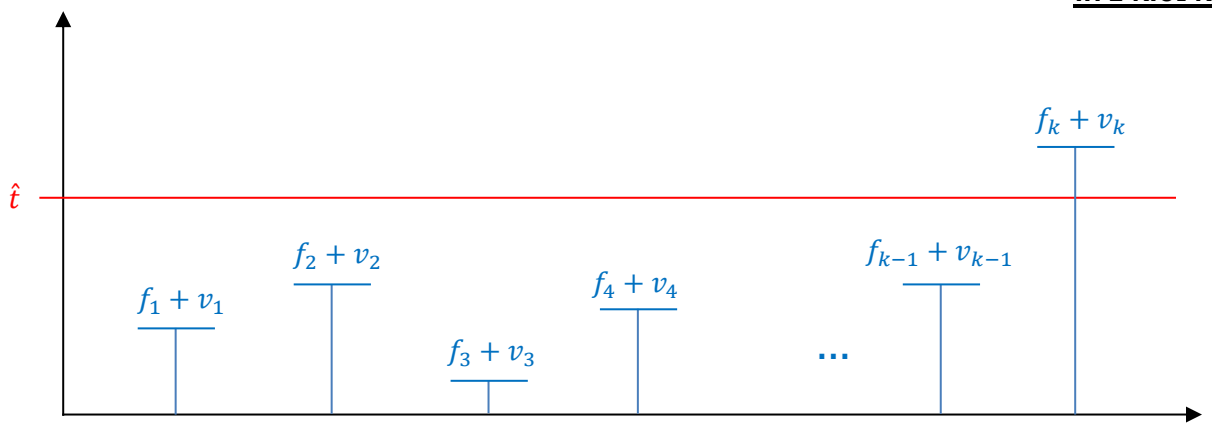
$$a_1 = \perp,\ a_2 = \perp,\ a_3 = \perp,\ \ldots,\ a_{k-1} = \perp,\ a_k = \top$$

(כל השאלה היא מהו k)

נקבע פלט אפשרי $\vec{a} = (a_1 = \perp, \ldots, a_{k-1} = \perp, a_k = \top)$.
עלינו להראות כי

$$\Pr[A(X) = \vec{a}] \leq e^{\varepsilon} \cdot \Pr[A(X') = \vec{a}]$$

<u>אינטואיציה:</u>



<u>הערה:</u> מספיק להראות כי לכל קביעה של $v_1, v_2, \ldots, v_{k-1}$ מתקיים

$$\Pr_{\hat{t}, v_k}[A(X) = \vec{a}] \leq e^{\varepsilon} \cdot \Pr_{\hat{t}, v_k}[A(X') = \vec{a}]$$

כי אז

$$\Pr[A(X) = \vec{a}] = \sum_{v_1, \ldots, v_{k-1}} \Pr[A(X) = \vec{a} | v_1, \ldots, v_k] \cdot \Pr[v_1, \ldots, v_k] \leq \cdots$$

אז נקבע $v_1, v_2, \ldots, v_{k-1}$ ונחשב:

$$\Pr_{\hat{t},v_k}[A(X)=\vec{a}] = \Pr_{\hat{t},v_k}\left[\hat{t} > \max_{i<k}\left\{f_i(X)+v_i\right\} \ \text{ AND } \ f_k(X)+v_k \geq \hat{t}\right] = \big((1)\big)$$

נזכור כי $v_k \sim \text{Lap}\left(\frac{4}{\varepsilon}\right)$ ולכן מתכונות התפלגות לפלאס נקבל

$$\big((1)\big) \ \leq \ e^{\varepsilon/2}\cdot \Pr_{\hat{t},v_k}\left[\hat{t} > \max_{i<k}\left\{f_i(X)+v_i\right\} \ \text{ AND } \ f_k(X)+v_k-2 \geq \hat{t}\right]$$

$$= \ e^{\varepsilon/2}\cdot \Pr_{\hat{t},v_k}\left[\ \hat{t} \in \left(\max_{i<k}\left\{f_i(X)+v_i\right\} \ , \ \ f_k(X)+v_k-2\right)\right] = \big((2)\big)$$

נזכור כי $\hat{t} \leftarrow t + \text{Lap}\left(\frac{2}{\varepsilon}\right)$ ולכן מתכונות התפלגות לפלאס נקבל

$$\big((2)\big) \ \leq \ e^{\varepsilon/2}\cdot e^{\varepsilon/2}\cdot \Pr_{\hat{t},v_k}\left[\ \hat{t} \in \left(\max_{i<k}\left\{f_i(X)+1+v_i\right\} \ , \ \ f_k(X)+v_k-1\right)\right]$$

$$\leq e^{\varepsilon}\cdot \Pr_{\hat{t},v_k}\left[\ \hat{t} \in \left(\max_{i<k}\left\{f_i(X')+v_i\right\} \ , \ \ f_k(X')+v_k\right)\right] = e^{\varepsilon}\cdot \Pr_{\hat{t},v_k}[A(X')=\vec{a}]$$

כאשר אי-השוויון האחרון נובע מכך ש- $|f_i(X)-f_i(X')| \leq 1$ ולכן הקטע האחרון <u>מכיל</u> את הקטע שלפניו ולכן נכנסים אליו בהסתברות גבוהה יותר.

מ.ש.ל.

<u>ניתוח הדיוק של *AboveThreshold*:</u>

- נניח שמבצעים k שאילתות לכל היותר: $f_1, f_2, \ldots, f_k$
- אזי לאורך הריצה אנחנו מגרילים (לכל היותר) $k+1$ רעשים לפלאסיים עם פרמטר $\frac{4}{\varepsilon}$ או $\frac{2}{\varepsilon}$. כפי שלמדנו, לכל אחד מהרעשים האלה מתקיים

$$\Pr\left[\left|\text{Lap}\left(\frac{4}{\varepsilon}\right)\right| > \frac{4}{\varepsilon}\ln\left(\frac{k+1}{\beta}\right)\right] \leq \frac{\beta}{k+1}$$

ולכן, בהסתברות לפחות $1-\beta$ מתקיים <u>שכל</u> הרעשים הם בגודל $\frac{4}{\varepsilon}\ln\left(\frac{k+1}{\beta}\right)$ לכל היותר.

במקרה כזה, אם האלגוריתם החזיר $\perp$ עבור f_i אזי מתקיים

$$f_i(X) < t + \frac{8}{\varepsilon}\ln\left(\frac{k+1}{\beta}\right)$$

ואם האלגוריתם החזיר $\top$ אזי

$$f_i(X) \geq t - \frac{8}{\varepsilon}\ln\left(\frac{k+1}{\beta}\right)$$