אוניברסיטת חיפה
החוג למדעי המחשב

אלגוריתמים מתקדמים

סיכומי ההרצאות של ד"ר עמית לוי

נכתב ע"י בר וייסמן

סמסטר אביב תשפ"ד

תוכן העניינים

הקדמה

דוגמה 1. בעיות אופטימיזציה.

1. (MST) עץ פורש במשקל מינימלי.
   - קלט: גרף ממושקל.
   - פלט: עץ פורש מינימלי.
2. קליקה בגודל מקסימלי.
   - קלט: גרף לא מכוון.
   - פלט: תת-גרף מלא בגודל מקסימלי.
3. (3-SAT) סיפוק נוסחת CNF.
   - קלט: נוסחת 3-CNF.
   - פלט: האם קיימת ל-φ השמה מספקת?

$\textbf{SAT} \leq_m \textbf{CLIQUE}$

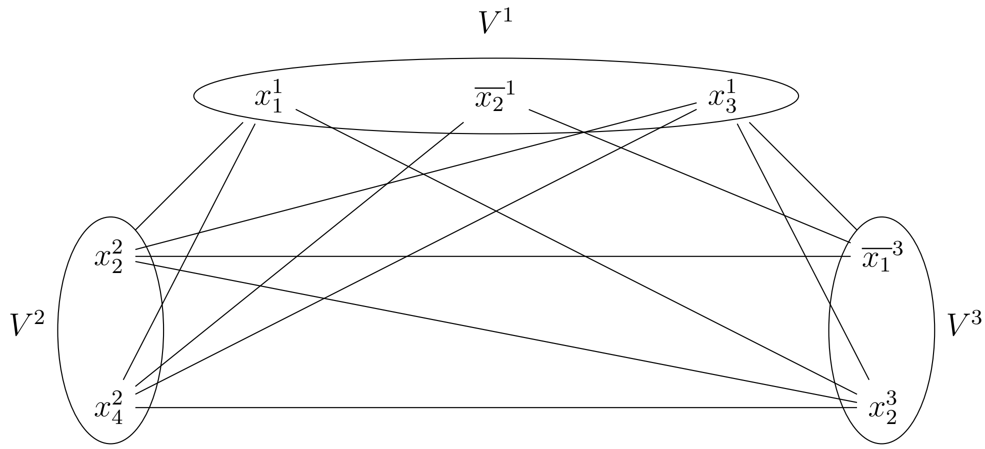
נראה רדוקציה מ-SAT ל-CLIQUE: נניח שקיים אלגוריתם שרץ בזמן פולינומי בגודל הקלט עבור עבור k-CLIQUE. בהינתן CNF $\varphi = \bigcap_{i=1}^{k} c^i$, נבנה גרף G_φ באופן הבא:

- קודקודי הגרף מחולקים ל-k קבוצות: $V^1, \ldots, V^k$.
- לכל קבוצה V^j, נתאים פסוקית C^j כך שלכל ליטרל x_i או $\overline{x_i}$ ב-c^j יש קודקוד ב-V^j, באופן דומה לאיור 0.0.1.
- נחבר קשת בין כל שני קודקודים שלא נמצאים באותהה קבוצה, וגם אין סתירה ביניהם.

למה 1. קיימת ל-φ השמה מספקת $\iff$ קיימת ב-G_φ קליקה בגודל k.

הוכחה. (למה 1) נראה את שני הכיוונים.

- נניח שיש ל-φ השמה מספקת $\tau : \{x_1, \ldots, x_n\} \to \{0, 1\}$, ונראה שב-$G_\varphi$ קליקה בגודל k.
- מאחר ו-τ מספקת, אזי לכל פסוקית c^j יש לפחות ליטרל אחד שמסתפק (כלומר $\tau(x_i) = 0$ אם מופיע $\overline{x_i}$, ואחרת $\tau(x_i) = 1$).
  - מכל c^j נבחר משתנה, ונתבונן בקודקודים המתאימים ב-G_φ.



איור 0.0.1: הגרף G_φ עבור $\varphi = (x_1 \vee \overline{x_2} \vee x_3) \wedge (x_2 \vee x_4) \wedge (\overline{x_1} \vee x_2)$.

- מאחר ו-τ השמה מספקת, אין סתירה בין הקודקודים השונים, ולכן יש צלע בין כל שני קודקודים: זו קליקה בגודל k.

- נניח שב-G_φ יש קליקה בגודל k, ונראה השמה מספקת.
  - מאחר ואין צלעות בין קודקודים באותו חלק, הקליקה מורכבת מקודקוד אחד בכל V^j.
  - לכל קודקוד x_i^j, נציב $\tau(x_i) = 1$ לכל קודקוד $\overline{x_i}^j$ נציב $\tau(x_i) = 0$, ולשאר המשתנים נבחר שרירותית.
  - קיבלנו השמה, מאחר ואין סתירות.
  - לפי הגדרת τ, לכל פסוקית יש משתנה x_i כך ש-$\tau(x_i) = 1$ או שיש $\overline{x_i} = 0$ ו-$\tau(x_i) = 0$: כל פסוקית מסתפקת.

$\square$

התמודדות עם בעיות קשות

גישה 1. למצוא משפחת קלטים אשר יש בהם עניין ועבורן יש אלג' פולינומיים (למשל: גרף מישוריים, גרפים עם דרגה קטנה).

גישה 2. לתכנן אלגוריתם קירוב. כלומר, עבור בעיות אופטימיזציה אלו אלגוריתמים שמוצאים פתרון תת-אופטימלי, אבל קרוב בערכו לערך האופטימלי.

סימונים

- עבור קלט x, נסמן ב-$\mathrm{OPT}(x)$ את הפתרון האופטימלי ל-x.
- עבור פתרון כלשהו y, נסמן ב-$\mathrm{Val}(y)$ את ערך הפתרון.

- אלגוריתם קירוב $\mathcal{A}$ מבטיח מציאת פתרון y כך ש-$y=\mathcal{A}(x)$, עבורו:
  אם הבעיה היא מקסימיזציה,

$$\text{Val}(\mathcal{A}(x)) \leq \frac{\text{Val}(\text{OPT}(x))}{B},$$

ואם הבעיה היא מינימיזציה:

$$\text{Val}(\mathcal{A}(x)) \leq B \cdot \text{Val}(\text{OPT}(x)),$$

עבור $B \geq 1$. למשל $B = 2, \log n$.

הערה. היינו רוצים אלגוריתם שמקבל כקלט פרמטר $0 < \varepsilon < 1$ ונותן $B = 1 + \varepsilon$, וזמן הריצה של האלגוריתם הוא פולינומי ב-${}^{1}/_{\varepsilon}$ (FPTAS: Fully Polynomial Time Approximation Scheme).

1 אלגוריתמים דטרמיניסטיים

1.1 בעיית כיסוי הקודקודים

<u>קלט:</u> גרף לא מכוון ולא ממושקל G.
<u>פלט:</u> תת-קבוצה C של קודקודים <u>בגודל מינימלי</u> כך שכל צלע ב-G נוגעת בקודקוד ב-C.

- זו בעיה NP-קשה.
- אלגוריתם 1 מציג גישה חמדנית לבעיה. האלגוריתם לעיל עלול להחזיר תשובה C שגודלה (בערך) $\log n$ מהאופטימלי.

- נראה חסם תחתון - נבנה גרף דו-צדדי $G=(L,R,E)$, בדומה לאיור 1.1.1.
  - ב-L יש k קודקודים.
  - R מורכב מקבוצות $R_2,\ldots,R_k$ כאשר $|R_i|=\left\lfloor\frac{k}{i}\right\rfloor$.
  - לכל $i\in[2,k]$, לכל קודקוד ב-R_i יש i קודקודים שכנים ב-L, כאשר לשני קודקודים ב-R_i אין שכנים משותפים.

- נשים לב:

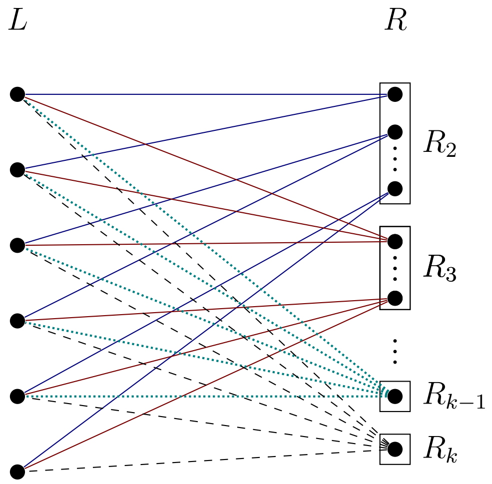
$$\Theta\left(k\log k\right)=\sum_{i=1}^{k}\left(\frac{k}{i}-1\right)\leq\sum_{i=1}^{k}\left\lfloor\frac{k}{i}\right\rfloor\leq k\sum_{i=1}^{k}\frac{1}{i}=\Theta\left(k\log k\right)$$

ולכן $n=\Theta\left(k\log k\right)$.

- נשים לב ש-L הוא כיסוי בגודל k.
- כל קודקוד ב-L, דרגתו לכל היותר $k-1$.
- לכן, בשלב הראשון האלגוריתם יבחר את הקודקוד היחיד ב-R_k, שדרגתו k.
- לאחר בחירתו, דרגת כל הקודקודים ב-L יורדת ב-1, ולכן לכל היותר $k-2$.
- בנקודה זו הקודקוד ב-R_{k-1} ייבחר (דרגתו $k-1$).
- באופן כללי, בשלב ה-j לכל קודקוד ב-L יש דרגה $\leq k-j$, ולקודקודים ב-R_{k-j+1} יש דרגה $k-j+1$.

Algorithm 1: $\texttt{VC} - \texttt{greedy}\,(G)$

אתחל $E' \leftarrow E, C \leftarrow \emptyset$.
כל עוד $E' \neq \emptyset$:
  יהי v קודקוד בעל דרגה מקסימלית ב-$G' = (V, E')$.
  עדכן $C \leftarrow C \cup \{v\}\,, E' \leftarrow E' \setminus \{(u,v)\}$.
החזר את C.



איור 1.1.1: חסם תחתון לאלגוריתם 1.

- לכן, האלגוריתם יבחר $C = R$, וכך $|C| = \Omega\left(k \log k\right)$.

אלגוריתם 2 מקרב את ה-VC על ידי מציאת זיווג מקסימלי (לא בהכרח מקסימום).

שים לב:

- האלגוריתם אכן מוצא כיסוי, מאחר ומסיימים כאשר E' ריקה.
- נתבונן בקבוצת הצלעות שנבחרות ע"י האלגוריתם (נסמנה ב-M): קבוצה זו היא זיווג, אזי $|C| = 2\,|M|$.
- בנוסף, גודל כל כיסוי הוא לפחות M: עבור כל צלע ב-M חייבים לבחור לכיסוי לפחות קודקוד אחד, והקודקודים זרים. לכן:

$$|\mathrm{OPT}\,(G)| \geq |M| \implies \boxed{|C| \leq 2\,|\mathrm{OPT}\,(G)|}$$

1.2 בעיית כיסוי הקבוצות

קלט: קבוצה סופית X ואוסף תתי-קבוצות $\mathcal{F} = \{S_1, \ldots, S_k\}$ כך ש-X-$\bigcup_{i=1}^{k} S_i = X$.

פלט: תת-קבוצה $C \subseteq \mathcal{F}$ שמהווה כיסוי קבוצות של X בגודל מינימלי.

Algorithm 2: $\texttt{VC}-\texttt{match}\,(G)$

אתחל $E' \leftarrow E, C \leftarrow \emptyset$.
כל עוד $E' \neq \emptyset$:
  תהי (u,v) צלע כלשהי ב-E'.
  $C \leftarrow C \cup \{u,v\}$.
  הסר מ-E' את כל הצלעות שנוגעות ב-u או ב-v.
החזר את C.

Algorithm 3: $\texttt{SetCoverGreedy}\,(X,\mathcal{F})$

אתחל $C \leftarrow \emptyset, U \leftarrow X$.
כל עוד $U \neq \emptyset$:
  תהי $S \in \mathcal{F}$ קבוצה שממקסמת את $|S \cap U|$.
  $C \leftarrow C \cup S, U \leftarrow U \setminus S$.
החזר את C.

הגדרה. נגיד ש-$C \subseteq \mathcal{F}$ היא כיסוי קבוצות (Set Cover) של X אם

$$\bigcup_{S\in C} S = X\,.$$

אלגוריתם 3 מקרב את הפתרון בגישה חמדנית: בכל איטרציה תיבחר הקבוצה שמכסה כמה שיותר איברים שטרם כוסו.

משפט 1. אלגוריתם 3 משיג פקטור $\mathcal{O}(\log n)$.

סימונים:

- U_i הוא סט האיברים שטרם כוסו בסוף האיטרציה ה-i של האלגוריתם.
- $u_i := |U_i|$. לכן $u_0 = n$, ואם t האיטרציה האחרונה אז $u_t = 0$.
- OPT הוא הגודל האופטימלי של הכיסוי, ו-$C^\star$ הוא כיסוי אופטימלי כלשהו.
- S_i הוא הסט שנבחר באיטרציה ה-i של האלגוריתם.

למה 2. לכל $i \geq 0$, בתחילת האיטרציה ה-$i+1$ קיים קבוצה $S \in \mathcal{F}$ שטרם נבחרה כך ש-S מכסה לפחות u_i/OPT איברים ב-U_i:

$$|S \cap U_i| \geq \frac{|U_i|}{\mathrm{OPT}}$$

הוכחה. (למה 2)

- מתוך הקבוצות ב-$\mathcal{F}$ שלא נבחרו ב-i האיטרציה הראשונות, נסתכל על הקבוצות ששייכות ל-$C^\star$ ונסמנם ב-$C_i^\star$: $C_i^\star = C^\star \setminus \{S_1,\ldots,S_i\}$.
- $C^\star$ מכסה את X (לכן גם את U_i), ו-U_i מכיל את כל האיברים שטרם כוסו בסוף האיטרציה ה-i (כלומר $U_i \cap \bigcup_{k=1}^{i} S_k = \emptyset$). לכן $C_i^\star$ מכסה את U_i, וכך $\bigcup_{S\in C_i^\star} S \cap U_i = U_i$.

- נניח בשלילה שלא קיימת $S \in C_i^\star$ שמכסה לפחות u_i/OPT איברים. אזי, מאחר ו-$\mathrm{OPT} = |C^\star| \geq |C_i^\star|$:

$$|U_i| = \left|\bigcup_{S\in C_i^\star} S \cap U_i\right| \leq \sum_{S\in C_i^\star} |S \cap U_i| < \sum_{S\in C_i^\star} \frac{|U_i|}{\mathrm{OPT}} \leq \sum_{S\in C_i^\star} \frac{|U_i|}{|C_i^\star|} = |U_i|$$

הגענו לסתירה, ולכן קיימת $S \in C_i^\star$ כזו.

□

הוכחה. (משפט 1)
יהי $i > 0$. לפי הלמה, קיבלנו כי

$$\begin{aligned}
u_i &= |U_i| \\
&= |U_{i-1}| - |S_i \cap U_{i-1}| \\
&\leq |U_{i-1}| - \frac{|U_{i-1}|}{\mathrm{OPT}} \\
&= u_{i-1}\left(1 - \frac{1}{\mathrm{OPT}}\right) \\
&\leq u_{i-2}\left(1 - \frac{1}{\mathrm{OPT}}\right)^2 \\
&\vdots \\
&\leq u_0\left(1 - \frac{1}{\mathrm{OPT}}\right)^i \\
&= n\left(1 - \frac{1}{\mathrm{OPT}}\right)^i
\end{aligned}$$

באיטרציה האחרונה t, ברור כי $u_t = 0 < 1$. נמצא את ה-t שמקיים זאת:

$$u_t = n\left(1 - \frac{1}{\mathrm{OPT}}\right)^t < 1$$

ידוע כי לכל $x \geq -1$ $\alpha > 0,$ מתקיים $(1+x)^\alpha \leq e^{x\alpha}$. לכן, עבור $x = -\frac{1}{\mathrm{OPT}}$, נקבל:

$$n\left(1 - \frac{1}{\mathrm{OPT}}\right)^t \leq ne^{-\frac{t}{\mathrm{OPT}}} < 1 \implies n < e^{\frac{t}{\mathrm{OPT}}} \implies \boxed{t = \log n \cdot \mathrm{OPT} + 1}$$

□

טענה 1. אלגוריתם 3 משיג יחס קירוב של $\mathcal{O}(\log S)$ כאשר $S = \max_{S\in\mathcal{F}}\{|S|\}$.

הוכחה. (טענה 1) נשתמש באותם הסימונים עבור U_i ו-S_i. נסמן ב-S_i' תת קבוצה של S_i שכוסו בפעם הראשונה על ידי S_i.

- למשל, $S'_1 = S_1$ ו-$S'_1 = S_2 \setminus S_1 = S_2 \cap U_1$. באופן כללי, $S'_i = S_i \cap U_{i-1}$.
- אוסף הקבוצות S'_i מהווה חלוקה של X: S'_i זרות ואיחודן הוא X.
- לכל $i \geq 0$ ולכל $x \in S'_i$, נגדיר מחיר $c_x := \frac{1}{|S'_i|}$.
  - נשים לב כי לכל i מתקיים

$$\sum_{x \in S'_i} c_x = |S'_i| \cdot \frac{1}{|S'_i|} = 1$$

$$\sum_{x \in X} c_x = \sum_{i=1}^{t} \sum_{x \in S'_i} c_x = t .$$

- יהי $C^\star$ כיסוי אופטימלי. מאחר וכל x מופיע לפחות באחת מהקבוצות ב-$C^\star$:

$$\sum_{x \in X} c_x \leq \sum_{S \in C^\star} \sum_{x \in S} c_x .$$

□

טענה 2. (הוכחה בתרגיל) לכל $S \in \mathcal{F}$, מתקיים $\sum_{x \in S} c_x \leq H(|S|) = \sum_{i=1}^{|S|} \frac{1}{i}$.

מטענה 2, קיבלנו

$$\begin{aligned} |C| &= t \\ &= \sum_{x \in X} c_x \\ &\leq \sum_{S \in C^\star} \sum_{x \in S} c_x \\ &\leq \sum_{S \in C^\star} H(|S|) \\ &\leq |C^\star| \cdot \max_{S \in \mathcal{F}} \log |S| \\ &= \mathcal{O}(\log S) \cdot \text{OPT}. \end{aligned}$$

תהייה. האם ניתן לעשות יותר טוב?

תשובה. כנראה שלא.

1.3 בעיית הסוכן הנוסע

<u>קלט:</u> גרף מלא G עם פונקציית משקולות חיובית w.

<u>פלט:</u> מסלול סגור שעובר בכל קודקוד בגרף פעם אחד בדיוק עם <u>משקל מינימלי</u>.

Algorithm 4: $\mathtt{TSP-v1}\,(G,w)$

1. מצא MST T בגרף G.
2. יהי $\hat{T}$ מולטיגרף שמתקבל ע"י הכפלת כל קשת ב-T (עם משקל זהה).
3. מצא מסלול אוילר ב-$\hat{T}$ (קיים כזה מאחר והדרגות זוגיות), נסמנו ב-P.
4. יהי H מסלול סגור שמתקבל מ-P ע"י השארת המופעים הראשונים של כל קודקוד ב-P (הגרף מלא ולכן H מסלול תקין).
5. החזר את H.

- הבעיה היא NP-קשה, ואפילו קשה לקרב אותה.
- נסתכל על מקרה יותר מאולץ, בו המשקלים מקיימים את אי-שוויון המשולש:

$$\forall u,v,t\in V: w\left(u,v\right)\leq w\left(u,t\right)+w\left(t,v\right).$$

- נראה קירוב 2, ואז נשפרו לקירוב $3/2$.

הגדרה 1. מולטי-גרף הוא גרף עם אפשרות לקשתות מרובות בין זוג קודקודים.

הגדרה 2. מסלול אוילר הוא מסלול שעובר בכל קשת בגרף פעם אחת בדיוק.

הגדרה 3. גרף אוילריאני הוא גרף שמכיל מסלול אוילר.

משפט 2. גרף קשיר הוא אוילריאני $\iff$ כל הדרגות שלו זוגיות, או שיש בדיוק שני קודקודים שדרגתם אי-זוגית.

הערה 1. קיים אלגוריתם פולינומי שמוצא מסלול אוילרי. בנוסף, במידה וכל הדרגות זוגיות המסלול הוא סגור.

אלגוריתם 4 מקרב את הפתרון בגישה חמדנית (מבוסס MST).

משפט 3. אם w מקיימת את אי-שוויון המשולש, אז אלגוריתם 4 הוא 2-קירוב ל-TSP.

הוכחה. (משפט 3) עבור תת-קבוצה של צלעות y, נסמן $w\left(y\right)=\sum_{e\in y}w\left(e\right)$, ויהי $H^{\star}$ מסלול במשקל מינימלי (פתרון אופטימלי).

- נשים לב ש-$w\left(H^{\star}\right)\geq w\left(T^{\star}\right)\geq w\left(T\right)$ ($T^{\star}$ הוא איזשהו עץ פורש שמתקבל מהורדת צלע מ-$H^{\star}$).
- בנוסף, $w\left(\hat{T}\right)=2w\left(T\right)$ ו-$w\left(P\right)=w\left(\hat{T}\right)$.
- לבסוף, $w\left(H\right)\leq w\left(P\right)$ (נובע מאי-שוויון המשולש). בסך הכל, קיבלנו ש-

$$w\left(H\right)\leq w\left(P\right)=w\left(\hat{T}\right)=2w\left(T\right)\leq w\left(H^{\star}\right)$$

□

Algorithm 5: $\mathrm{TSP-v2}\left(G,w\right)$

1. מצא MST T בגרף G.
2. מצא זיווג מושלם במשקל מינימלי M בין הקודקודים שדרגתם אי-זוגית ב-T.
3. מצא מסלול אוילר במולטיגרף $T\cup M$, נסמנו ב-P.
4. יהי H מסלול סגור שמתקבל מ-P ע"י מחיקת קודקודים שכבר ראינו.
5. החזר את H.

רעיון:

- במקום להכפיל צלעות ב-MST, כל קודקוד עם דרגה זוגית נשאר כשהיה (לא נוגעים בו).
- אם לקודקוד דרגה אי-זוגית, נהפוך את דרגתו לזוגית ע"י מציאת זיווג מושלם עם משקל מינימלי.
  - אפשרי כי הגרף מלא, ויש מספר זוגי של קודקודים עם דרגה אי-זוגית.
  - ניתן למצוא זאת בזמן פולינומי.

אלגוריתם 5 מתאר את האלגוריתם המשופר.

משפט 4. אלגוריתם 5 הוא $3/2$-קירוב ל-TSP.

הוכחה. (משפט 4) באופן דומה להוכחה של משפט 4.

- $w\left(H\right)\leq w\left(P\right)$.
- $w\left(P\right)\leq w\left(T\right)+w\left(M\right)$.
- כמו קודם, $H^{\star}$ הוא מסלול אופטימלי, ולכן $w\left(T\right)\leq w\left(H^{\star}\right)$.
- נותר להראות ש-$w\left(M\right)\leq\frac{w\left(H^{\star}\right)}{2}$ - נובע מלמה ?.
- בסך הכל,

$$w\left(H\right)\leq w\left(P\right)\leq w\left(T\right)+w\left(M\right)\leq w\left(H^{\star}\right)+\frac{w\left(H^{\star}\right)}{2}=\frac{3}{2}w\left(H^{\star}\right).$$

□

למה 3. $w\left(M\right)\leq\frac{w\left(H^{\star}\right)}{2}$.

הוכחה. (למה 3) נסמן ב-S את קבוצת הקודקודים בעלי דרגה אי-זוגית ב-T.

- כזכור, $H^{\star}$ הוא מסלול סגור במשקל מינימלי (שעובר בכל הקודקודים), נסמנו ב-

$$\left\{v_{1},v_{2},\ldots,v_{n},v_{1}\right\}.$$

- נגדיר את $H^{\star}\left(S\right)=\left\{v_{i},v_{i_{2}},\ldots,v_{i_{k}},v_{i_{1}}\right\}$ להיות מסלול סגור שמתקבל מהורדה של כל הקודקודים שלא ב-S מ-$H^{\star}$.

דוגמה 2. עבור $n=7$, $S=\{v_2,v_4,v_5,v_7\}$, $H^\star=v_1,v_2,\dots,v_7,v_1$, נקבל

$$H^\star(S)=v_2v_4v_5v_7v_2.$$

- מאי-שוויון המשולש, מתקיים $w(H^\star(S))\le w(H^\star)$. נגדיר שני זיווגים:

$$M_1=\left\{(v_{i_1},v_{i_2}),(v_{i_3},v_{i_4}),\dots,\left(v_{i_{k-1}},v_{i_k}\right)\right\}$$

$$M_2=\{(v_{i_2},v_{i_3}),(v_{i_4},v_{i_5}),\dots,(v_{i_k},v_{i_1})\}$$

- נשים לב ש-$w(H^\star(S))=w(M_1)+w(M_2)$, וגם $w(M_1),w(M_2)\ge w(M)$. לכן,

$$w(H^\star)\ge w(H^\star(S))\ge w(M_1)+w(M_2)\ge 2w(M).$$

□

1.4 עץ שטיינר

קלט: גרף לא מכוון וממושקל G עם פונקציית משקל חיובית w (אם $(u,v)\notin E$ אז $w(u,v)=\infty$).

פלט: תת-גרף של G שהוא עץ ומכיל את כל קודקודי R (ואולי עוד מ-$V\setminus R$) עם משקל מינימלי.

- הבעיה היא NP-קשה.
- נראה שאפשר לעשות רדוקציה משמרת יחס קירוב של המקרה הכללי (כל w חיובית) למקרה שבו w מקיימת את אי-שוויון המשולש (אם קיים אלגוריתם עם ρ-קירוב למקרה בו w מקיימת את אי-שוויון המשולש, אז קיים אלגוריתם ρ-קירוב למקרה הכללי).

רדוקציה. בהינתן גרף $G=(V,E)$ ופונקציית משקל w שלא בהכרח מקיימת את אי-שוויון המשולש, נבנה גרף $G'=(V,E')$ ו-w' שמקיימת את אי-שוויון המשולש, באופן שישמר את פתרון הבעיה. G' יהיה גרף מלא, ו-$w'(u,v)=w(P_w(u,v))$ הוא משקל המסלול הקצר ביותר $u\rightsquigarrow v$ ב-G.

שים לב: w' מקיימת את אי-שוויון המשולש.

נסמן ב-$\mathrm{ST}(G,w,R)$ את עץ השטיינר של G עם פונקציית משקל w וקבוצת טרמינלים R, ו-$\mathrm{ST}(G',w',R)$ באותו האופן.

אבחנה. לכל $u,v\in V$, $w'(u,v)\le w(u,v)$, ולכן $\mathrm{ST}(G',w',R)\le\mathrm{ST}(G,w,R)$.

- נניח שיש לנו אלגוריתם ρ-קירוב לקלטים שמקיימים את אי-שוויון המשולש.
  - יהי T' הפלט של האלגוריתם על G',w',R, אזי $w'(T')\le\rho\cdot\mathrm{ST}(G',w',R)$.
- נגדיר עץ שטיינר T עבור G,w,R באופן הבא:
  - נחליף כל קשת (u,v) ב-T' במסלול $P_w(u,v)$.

- נסמן ב-K את תת-הגרף של G שמכיל את כל הקודקודים ב-T' ואת כל הקשתות שמופיעות על מסלולים $P_w\left(u,v\right)$ לכל $\left(u,v\right)$ ב-T'.

$$\implies w\left(P_w\left(u,v\right)\right)=w'\left(u,v\right)$$

- נגדיר את T להיות עץ פורש של K.

$$\implies w\left(T\right)\leq w\left(K\right)$$

- בסך הכל:

$$w\left(T\right)\leq w\left(K\right)\leq w'\left(T'\right)\leq\rho\cdot\mathrm{ST}\left(G',w',R\right)\leq\rho\cdot\mathrm{ST}\left(G,w,R\right).$$

- נותר להראות שקיים אלגוריתם 2-קירוב לקלטים שמקיימים את אי-שוויון המשולש (תרגיל).

1.5 בעיית תרמיל הגב

קלט: n אובייקטים המיוצגים ע"י מחירים $\left(s_i\right)_{i=1}^n$ ורווחים $\left(p_i\right)_{i=1}^n$, וקיבול תרמיל הגב B.

פלט: תת-קבוצה $U\subseteq\left\{1,\ldots,n\right\}$ כך ש-$\sum_{i\in U}s_i\leq B$ ו-$\sum_{i\in U}p_i$ מקסימלי.

- הבעיה היא NP-קשה.
- נראה FPTAS לבעיה (קירוב $\left(1-\varepsilon\right)$ בהינתן פרמטר קירוב $\varepsilon\in\left(0,1\right)$, זמן הריצה פולינומי בגודל הקלט: $\mathrm{poly}\left(1/\varepsilon\right)$).

1.5.1 תכנות דינמי

נתחיל באלגוריתם 6, שפותר את הבעיה באופן מדויק בזמן פולינומי ב-n ו-$P=\sum_{i=1}^n p_i$. גודל הקלט (בביטים) הוא

$$\sum_{i=1}^n\log\left(s_i\right)+\log\left(p_i\right)+\log B.$$

- נרצה ש-$A\left[i\right]\left[j\right]$ יהיה הגודל המינימלי של תת-קבוצה של $\left\{1,\ldots,i\right\}$ שהרווח שלהם שווה ל-j. אם לא קיימת תת-קבוצה כזאת, אז $A\left[i\right]\left[j\right]=\infty$.
- הפתרון יהיה $\max\left\{j\mid A\left[n\right]\left[j\right]\leq B\right\}$.
- זמן הריצה של האלגוריתם הוא $\mathcal{O}\left(nP\right)$.
  - אם P הוא פולינומי ב-n, סיימנו.
  - אחרת, נעשה scaling למשקלים.

Algorithm 6: $\texttt{KnapsackExact}\left(\left(s_i\right)_{i=1}^{n},\left(p_i\right)_{i=1}^{n}\right)$

תהי $A \in \mathbb{R}^{n\times(P+1)}$.
עבור $i=1,\ldots,n$:
    $A[i][0] \leftarrow 0$.
עבור $j=1,\ldots,P$:
    אם $j=p_1$ אז $A[i][0] \leftarrow s_1$, אחרת $A[i][0] \leftarrow \infty$.
עבור $i=2,\ldots,n$:
    עבור $j=1,\ldots,P$:
        $A[i][j] \leftarrow A[i-1][j]$
        אם $p_i \le j$ אז $A[i][j] \leftarrow \min\left\{A[i][j], s_i + A[i-1][j-p_i]\right\}$.
אתחל $found \leftarrow \text{FALSE}$.
כל עוד $found = \text{FALSE}$:
    אם $A[n][j] \le B$ אז $found \leftarrow \text{TRUE}$.
    אחרת $j \leftarrow j-1$.
עבור $i=2,\ldots,n$:
    אם $A[i][j] < A[i-1][j]$ אז הדפס i, $j \leftarrow j-p_i$.
אם $j=p_1$ הדפס 1.

Algorithm 7: $\texttt{KnapsackFPTAS}\left(\left(s_i\right)_{i=1}^{n},\left(p_i\right)_{i=1}^{n},\varepsilon\right)$

$k \leftarrow p_{max}\cdot\frac{\varepsilon}{n}$.
לכל i, הגדר $p'_i := \left\lfloor \frac{p_i}{k} \right\rfloor$.
הרץ את אלגוריתם 6 עבור s ו-p', וקבל קבוצה U'.
החזר את U'.

1.5.2 FPTAS

נראה אלגוריתם שעושה scaling למשקלים.

- נגדיר $p_{max} = \max_i p_i$.
  - אם $p_{max} \le \frac{n}{\varepsilon}$, אזי $P \le n\cdot p_{max} \le \frac{n^2}{\varepsilon}$, וכך זמן הריצה הוא $\mathcal{O}\left(\frac{n^3}{\varepsilon}\right)$.
  - אחרת, צריך לעשות scaling.
- נניח בה"כ כי לכל i, $s_i \le B$, ונפעיל את אלגוריתם 7.
- זמן הריצה הוא $\mathcal{O}\left(nP'\right) = \mathcal{O}\left(\frac{n^3}{\varepsilon}\right)$:

$$P' = \sum_{i=1}^{n} p'_i = \sum_{i=1}^{n}\left\lfloor \frac{p_i}{k}\right\rfloor \le \frac{P}{k} = \frac{P}{p_{max}}\cdot\frac{n}{\varepsilon} \le \frac{n^2}{\varepsilon}.$$

משפט 5. $P\left(U'\right) \ge (1-\varepsilon)\,\text{OPT}$.

הוכחה. (משפט 5) נסמן ב-$U^\star$ את הפתרון האופטימלי ביחס ל-p_i ($\sum_{i\in U^\star} s_i \le B$).

$$P\left(U^\star\right) = \sum_{i\in U^\star} p_i = \mathrm{OPT}$$

נשים לב ש-$p_{max} \le \mathrm{OPT}$. מאחר ו-U' אופטימלי ביחס ל-p', $P'\left(U'\right) \ge P'\left(U^*\right)$.

$$\begin{aligned}
P\left(U'\right) &= \sum_{i\in U'} p_i \\
&\ge \sum_{i\in U'} k\left\lfloor \frac{p_i}{k}\right\rfloor \\
&= k\sum_{i\in U'} p'_i \\
&= kP'\left(U'\right) \\
&\ge kP'\left(U^\star\right) \\
&= \sum_{i\in U^\star} k\left\lfloor \frac{p_i}{k}\right\rfloor \\
&\ge \sum_{i\in U^\star} k\left(\frac{p_i}{k}-1\right) \\
&\ge \sum_{i\in U^\star} p_i - nk \\
&= \mathrm{OPT} - p_{max}\varepsilon \\
&\ge \left(1-\varepsilon\right)\mathrm{OPT}.
\end{aligned}$$

□

1.6 בעיית k המרכזים

קלט: מספר $k\in\mathbb{N}$ ומרחב מטרי (P,d).

פלט: תת-קבוצה $S\subseteq P$ בגודל k כך ש-$r\left(P,S\right)$ מינימלי.

הגדרה 4. עבור תת-קבוצה $S\subseteq P$, נגדיר את המחיר של ה-clustering לפי S להיות

$$r\left(P,S\right) := \max_{p\in P} d\left(p,S\right),$$

כאשר $d\left(p,S\right) := \min_{p_j\in S} d\left(p,p_j\right)$.

דוגמה 3. עבור נקודות במישור: מציאת כיסוי P עם k דיסקים עם רדיוס מינימלי.

הבעיה היא NP-קשה. נראה אלגוריתם 2-קירוב לבעיה (אלגוריתם 8).

- נשים לב שבאיטרציה ה-i מחיר ה-clustering הוא בדיוק $\max_{p\in P} d\left(p,S\right)$.

Algorithm 8: $\texttt{CentersApprox}\left(P,d,k\right)$

אתחל $S\leftarrow\left\{p_1\right\}$ שרירותית.
עבור $i=2,\ldots,k$:
  $p\leftarrow\arg\max_{p\in P}d\left(p,S\right)$.
  $S\leftarrow S\cup\left\{p\right\}$.
החזר את S.

- בנוסף, המחיר של הפתרונות לאורך הריצה לא עולה.
- זמן הריצה של האלגוריתם הוא $\mathcal{O}\left(nk\right)$.

משפט 6. אלגוריתם 8 משיג פקטור קירוב 2.

הוכחה. (משפט 6) נסמן ב-$S^\star$ פתרון אופטימלי ו-$r^\star=r\left(P,S^\star\right)$.
נתאים לכל נקודה $p\in S$ נקודה (מרכז) $c\left(p_i\right)$ הקרוב ביותר ב-$S^\star$:

$$c\left(p\right):=\arg\min_{c\in S^\star}d\left(p,c\right).$$

נפריד למקרים:

1. כל נקודה ב-S ממופה לנקודה אחרת ב-$S^\star$: $c\left(p_i\right)\neq c\left(p_j\right)\implies p_i\neq p_j$.
   - נשים לב ש-$\left|S^\star\right|=\left|S\right|=k$, ולכל $p_j\in S^\star$ קיימת $p_i\in S$ שממופה אליה.
   - נסתכל על נקודה כלשהי $p_l\in P$, אזי קיימת $p_j\in S^\star$ כך ש-$d\left(p_l,p_j\right)\leq r^\star$.
   - יהי $p_i\in S$ כך ש-$c\left(p_j\right)=p_i$, אזי

$$d\left(p_l,p_i\right)\leq d\left(p_l,p_j\right)+d\left(p_j,p_i\right)\leq 2r^\star.$$

$$\implies r\left(P,S\right)\leq 2r^\star$$

2. קיימות שתי נקודות $p_i\neq p_{i'}$ כך ש-$c\left(p_i\right)=c\left(p_{i'}\right)$.
   - נניח בה"כ ש-$p_{i'}$ נוספה ל-S מייד אחרי p_i, ונסמן ב-S' את הסט רגע לפני ש-$p_{i'}$ נוספה.
   <u>שים לב:</u> $p_i\in S'$.
   - מאחר והמחירים של הפתרונות לאורך הריצה לא עולים, מתקיים

$$r\left(P,S'\right)\geq r\left(P,S\right).$$

   - מפעולת האלגוריתם, $p_{i'}$ היא נקודה ב-P שממקסמת את המרחק מ-S'.

$$r\left(P,S'\right)=d\left(p_{i'},S'\right)$$

   מאחר ו-$p_i\in S'$, ידוע כי $d\left(p_{i'},S'\right)\leq d\left(p_{i'},p_i\right)$.
   - לבסוף, נסמן $c=c\left(p_{i'}\right)=c\left(p_i\right)$, ונקבל

$$r\left(P,S\right)\leq r\left(P,S'\right)=d\left(p_{i'},S'\right)\leq d\left(p_{i'},p_i\right)\leq d\left(p_{i'},c\right)+d\left(c,p_i\right)\leq 2r^\star.$$

□

2 אלגוריתמים הסתברותיים

2.1 הסתברות

הגדרה 5. מרחב הסתברות בדיד הוא קבוצה סופית Ω כך שלכל $\omega \in \Omega$ יש משקל אי-שלילי (ההסברות של ω), ונסמנו $\Pr[w]$, כך שמתקיים

$$\sum_{\omega\in\Omega}\Pr[w]=1.$$

לעיתים נגיד ש-Ω (יחד עם פונקציית ההסתברות) היא מרחב המדגם.

דוגמה 4. מספר דוגמאות למרחבי הסתברות.

$$\Omega=\left\{\underbrace{H}_{\text{heads}},\underbrace{T}_{\text{tails}}\right\},\Pr[H]=\Pr[T]={}^{1}/_{2}$$

$$\Omega=\{H,T\}^{n},\forall\omega\in\Omega:\Pr[\omega]={}^{1}/_{2^{n}}$$

$$\Omega=\{H,T\}^{n},\forall\omega\in\Omega:\Pr[\omega]=p^{n_{1}}(1-p)^{n-n_{1}};n_{1}=\#_{H}\in\omega$$

הגדרה 6. תת-קבוצה $W\subseteq\Omega$ נקראת מאורע. אם $|W|=1$ אז W הוא <u>מאורע בסיסי</u>.

$$\Pr[W]=\sum_{\omega\in\Omega}\Pr[\omega]$$

דוגמה 5. תוצאה ספציפית של הטלת 6 מטבעות, למשל $HHTTHT$ הוא מאורע בסיסי. אוסף המחרוזות שיש להן אותו מספר H,T אינו מאורע בסיסי.

הגדרה 7. שני מאורעות $A,B\in\Omega$ הם זרים אם $A\cap B=\emptyset$.

דוגמה 6. יהיו A,B מאורעות כך ש-

$$A=\{\omega\in\{H,T\}^{n}\mid\#_{H}\in\omega={}^{n}/_{2}\}$$

$$B=\{\omega\in\{H,T\}^{n}\mid\#_{H}\in\omega={}^{n}/_{3}\}$$

אזי A,B זרים. לעומת זאת, A' ו-B' אינם זרים:

$$A'=\{\omega\in\{H,T\}^{n}\mid\#_{H}\in\omega\leq{}^{n}/_{2}\}$$

$$B'=\{\omega\in\{H,T\}^{n}\mid\#_{H}\in\omega\geq{}^{n}/_{3}\}$$

הגדרה 8. שני מאורעות $A,B\in\Omega$ הם בלתי תלויים אם

$$\Pr\left[A\cap B\right]=\Pr\left[A\right]\cdot\Pr\left[B\right].$$

דוגמה 7. A,B הם מאורעות בלתי תלויים.

$$A=\left\{H\right\}\times\left\{H,T\right\}^{n-1},B=\left\{H,T\right\}\times\left\{H\right\}\times\left\{H,T\right\}^{n-2}$$

$$\Pr\left[A\right]=\Pr\left[B\right]=\frac{1}{2}$$

$$\Pr\left[A\cap B\right]=\frac{1}{4}=\Pr\left[A\right]\cdot\Pr\left[B\right]$$

הגדרה 9. פונקציה $\chi:\Omega\to\mathbb{R}$ היא משתנה מקרי מעל Ω.

דוגמה 8. עבור A מדוגמא 7,

$$\chi\left(\omega\right)=\begin{cases}1 & \omega\in A\\ 0 & \omega\notin A\end{cases}.$$

זהו משתנה מקרי ברנולי, או למשל $\chi\left(\omega\right)$ הוא מספר ה-H ב-ω.

הגדרה 10. התוחלת של מ"מ χ הוא ממוצע משוקלל לפי Pr:

$$\mathbb{E}\left[\chi\right]\coloneqq\sum_{\omega\in\Omega}\chi\left(\omega\right)\cdot\Pr\left[\omega\right].$$

דוגמה 9. עבור $\chi:\Omega\to\left\{0,1\right\}$, נקבל

$$\mathbb{E}\left[\chi\right]=\Pr\left[\chi=1\right].$$

טענה 3. עבור מאורעות זרים A,B,

$$\Pr\left[A\cup B\right]=\Pr\left[A\right]+\Pr\left[B\right]$$

טענה 4. (חסם האיחוד) יהיו $W_1,\ldots,W_k$ מאורעות מעל Ω. אזי,

$$\Pr\left[\bigcup_{i=1}^{k}W_i\right]\leq\sum_{i=1}^{k}\Pr\left[W_i\right].$$

טענה 5. (לינאריות התוחלת) יהיו $\chi_1,\ldots,\chi_k$ מ"מ מעל Ω. אז עבור משתנה מקרי $X=\sum_{i=1}^{k}\chi_i$ מתקיים

$$\mathbb{E}\left[X\right]=\sum_{i=1}^{k}\mathbb{E}\left[\chi_i\right].$$

2.2 חסמי ריכוז מידה

- אי-שוויונים אלה נותנים חסמים על ההסתברות שמ"מ יהיה רחוק מהתוחלת.
  - עוזר להראות שפלט של אלגוריתם רנדומי עובד כמצופה.
- נראה את האי-שוויונים הבסיסיים והשימושיים ביותר.

2.2.1 אי-שוויון מרקוב (Markov)

משפט 7. (אי-שוויון מרקוב) יהי X מ"מ חיובי. אזי,

$$\forall a>0:\Pr\left[X\geq a\right]\leq\frac{\mathbb{E}\left[X\right]}{a}.$$

הוכחה. (משפט 7)

$$\mathbb{E}\left[X\right]=\sum_{i}i\cdot\Pr\left[X=i\right]\geq\sum_{i\geq a}i\cdot\Pr\left[X=i\right]\geq\sum_{i\geq a}a\cdot\Pr\left[X=i\right]=a\cdot\Pr\left[X\geq a\right]$$

□

דוגמה 10. (שימושים)

1. ידוע של-RandomQuickSort יש תוחלת זמן ריצה של $2n\ln\left(n\right)$. לפי מרקוב, ההסתברות שזמן הריצה הוא יותר מ-$2c\cdot n\ln n$ היא לכל היותר $\frac{1}{c}$.
2. הטלת מטבעות: אם נטיל n מטבעות הוגנים ($Pr\left[H\right]=\frac{1}{2}$), אז תוחלת מספר ה-$H$-ים היא $n/2$. לפי מרקוב, ההסתברות שיהיו יותר מ-$\frac{3n}{4}$ H-ים היא לכל היותר $\frac{2}{3}$.

הערה. אי-שוויון מרקוב שימושי כאשר אין לנו מידע על המשתנה המקרי מלבד התוחלת (או שקשה להשיג עוד מידע עליו). בדוגמאות לעיל ניתן להשיג חסמים טובים יותר.

כדי לקבל חסמים טובים יותר, נצטרך יותר מידע על המ"מ. מדד נפוץ הוא השונות של מ"מ: מודד את המרחק הטיפוסי מהתוחלת.

הגדרה 11. (מומנט k) המומנט ה-k של מ"מ X מוגדר להיות $\mathbb{E}\left[X^{k}\right]$.

הגדרה 12. (שונות) השונות של מ"מ X מוגדר להיות

$$Var\left[X\right]\coloneqq\mathbb{E}\left[\left(X-\mathbb{E}\left[X\right]\right)^{2}\right]=\mathbb{E}\left[X^{2}\right]-\mathbb{E}\left[X\right]^{2}$$

סטיית התקן מוגדרת להיות

$$\sigma\coloneqq\sqrt{Var\left[X\right]}$$

השונות המשותפת של מ"מ X,Y מוגדרת להיות

$$Cov\left(X,Y\right)=\mathbb{E}\left[\left(X-\mathbb{E}\left[X\right]\right)\left(Y-\mathbb{E}\left[Y\right]\right)\right]$$

אומרים ש-X,Y בעלי קורלציה חיובית (שלילית) אם $Cov\left(X,Y\right)>0$ (<0).

עובדה 1.

$$Var\left[X+Y\right]=Var\left[X\right]+Var\left[Y\right]+2Cov\left(X,Y\right)$$

בנוסף, אם X,Y בלתי תלויים נקבל

$$Var\left[X+Y\right]=Var\left[X\right]+Var\left[Y\right].$$

2.2.2 אי-שוויון צ'בישב (Chebyshev)

משפט 8. (אי-שוויון צ'בישב) יהי X מ"מ ו-$a>0$, אזי,

$$\Pr\left[|X-\mathbb{E}\left[X\right]|\geq a\right]\leq\frac{Var\left[X\right]}{a^2}.$$

הוכחה. (משפט 8) נשתמש באי-שוויון מרקוב:

$$\Pr\left[|X-\mathbb{E}\left[X\right]|\geq a\right]=\Pr\left[\left(X-\mathbb{E}\left[X\right]\right)^2\geq a^2\right]\underset{(\star)}{\leq}\frac{\mathbb{E}\left[\left(X-\mathbb{E}\left[X\right]\right)^2\right]}{a^2}=\frac{Var\left(X\right)}{a^2}$$

מעבר $(\star)$ משתמש במרקוב, מאחר ו-$\left(X-\mathbb{E}\left[X\right]\right)^2$ הוא חיובי. □

דוגמה 11. (שימוש) הטלת מטבעות:

- נסמן ב-X את מספר הטלות ה-H בהטלת n מטבעות הוגנים:

$$X=\sum_{i=1}^{n}X_i,\qquad X_i=\begin{cases}1 & H\\ 0 & \text{אחרת}\end{cases}$$

- נרצה לחסום את ההסתברות ש-$X\geq\frac{3}{4}n$:

$$Var\left[X\right]=\sum_{i=1}^{n}Var\left[X_i\right]=\sum_{i=1}^{n}\left(\frac{1}{2}\left(1-\frac{1}{2}\right)^2+\frac{1}{2}\left(0-\frac{1}{2}\right)^2\right)=\frac{n}{4}$$

$$\implies\Pr\left[X\geq\frac{3}{4}n\right]\leq\Pr\left[|X-\mathbb{E}\left[X\right]|\geq\frac{n}{4}\right]\leq\frac{Var\left[X\right]}{\left(\frac{n}{4}\right)^2}=\boxed{\frac{4}{n}}$$

חסם הרבה יותר טוב ממרקוב.

הערה 2. נוח להשתמש בצ'בישב כאשר השונות קלה לחישוב.

- לעיתים נרצה לחסום את הזנב העליון והתחתון של התפלגות המ"מ:

$$\Pr\left[X>(1+\varepsilon)\,\mathbb{E}\left[X\right]\right],\quad\Pr\left[X<(1-\varepsilon)\,\mathbb{E}\left[X\right]\right]$$

- נתעניין במקרה בו X הוא סכום של מ"מ ב"ת, (נפוץ בניתוח של אלגוריתמים רנדומיים).
- מחוק המספרים הגדולים, יודעים שסכום של n מ"מ ב"ת מאותה התפלגות הוא בערך $n\mu$ כאשר μ היא התוחלת.
- ממשפט הגבול המרכזי, ידוע ש-$\frac{X-n\mu}{\sqrt{n\cdot\sigma^2}}\xrightarrow[n\to\infty]{}\mathcal{N}\left(0,1\right)$, ולכן הסטייה מהתוחלת היא $\mathcal{O}\left(\sqrt{n}\right)$.

2.2.3 חסמי צ׳רנוף (Chernoff)

חסמי צ׳רנוף נותנים לנו תוצאה כמותית להסתברות שמ"מ X רחוק מהתוחלת (לערכים גדולים מספיק של n).

דוגמה 12. (הטלת מטבעות) נניח שיש לנו n מטבעות כך ש-$\Pr\left[X\right]=p\in\left(0,1\right)$.

- בתוחלת מספר ה-H-ים הוא $p\cdot n$.
- כדי לחסום את הזנב (עליון), בעיקרון צריך לחשב את

$$\Pr\left[X\geq k\right]=\sum_{i\geq k}\binom{n}{i}p^{i}\left(1-p\right)^{n-i}.$$

- צריך להראות שההסתברות הנ"ל קטנה כאשר $k\gg np$ (נניח $\left(1+\varepsilon\right)n\cdot p$), אבל זה קשה מדי לחישוב.
- לחילופין, נוכל להרחיב את השיטה בה השתמשנו באי-שוויון מרקוב, אבל נסתכל על מומנטים גבוהים יותר. נסתכל על מומנטים זוגיים:

$$\Pr\left[\left|X-\mathbb{E}\left[X\right]\right|>a\right]=\Pr\left[\left(X-\mathbb{E}\left[X\right]\right)^{2k}>a^{2k}\right]\leq\frac{\mathbb{E}\left[\left(X-\mathbb{E}\left[X\right]\right)^{2k}\right]}{a^{2k}}$$

- באי-שוויון צ׳רנוף, עבור $t>0$, נסתכל על

$$\Pr\left[X\geq a\right]=\Pr\left[e^{tX}\geq e^{ta}\right]\leq\frac{\mathbb{E}\left[e^{tX}\right]}{e^{ta}}$$

- למה דווקא e^{tX}?

1. נסמן:

$$M_{X}\left(t\right)=\mathbb{E}\left[e^{tX}\right]=\mathbb{E}\left[\sum_{i\geq0}\frac{t^{i}}{i!}X^{i}\right]=\sum_{i\geq0}\frac{t^{i}}{i!}\mathbb{E}\left[X^{i}\right]$$

אם יש לנו את $M_{X}\left(t\right)$, נוכל למצוא בקלות את המומנט ה-j ע"י $M_{X}^{(j)}\left(t\right)$ (הנגזרת ה-j לפי x, ולהציב $t=0$). $M_{X}\left(t\right)$ נקראת פונקציה יוצרת מומנטים של X.

2. אם $X=X_{1}+X_{2}$ כאשר X_{1},X_{2} ב"ת, אזי

$$\mathbb{E}\left[e^{tX}\right]=\mathbb{E}\left[e^{t\left(X_{1}+X_{2}\right)}\right]=\mathbb{E}\left[e^{tX_{1}}\right]\mathbb{E}\left[e^{tX_{2}}\right].$$

באופן כללי, חסמי צ׳רנוף נגזרים מ-

$$\Pr\left[X\geq a\right]\leq\frac{\mathbb{E}\left[e^{tX}\right]}{e^{ta}}.$$

דוגמה 13. נסתכל על מקרה שימושי של הטלת מטבעות לא זהים. נסמן ב-$x_1,\ldots,x_n$ מ"מ ב"ת כך ש-$x_i=1$ בהסתברות p_i, ו-$x_i=0$ אחרת. נסמן $X=\sum_{i=1}^n x_i$.

$$\mu=\mathbb{E}\left[X\right]=\sum_i p_i$$

לכן,

$$\begin{aligned}
\mathbb{E}\left[e^{tX}\right] &= \prod_{i=1}^{n}\mathbb{E}\left[e^{tx_i}\right]\\
&= \prod_{i=1}^{n}\left(p_ie^{1\cdot t}+\left(1-p_i\right)e^{0\cdot t}\right)\\
&= \prod_{i=1}^{n}\left(1+p_i\left(e^t-1\right)\right)\\
(1+x\leq e^x) &\leq \prod_{i=1}^{n}e^{p_i\left(e^t-1\right)}\\
&= e^{\mu\left(e^t-1\right)}.
\end{aligned}$$

משפט 9. (צ'רנוף - זנב עליון) במקרה וההטלות ב"ת,

1. לכל $\delta>0$,

$$\Pr\left[X\geq\left(1+\delta\right)\mathbb{E}\left[X\right]\right]\leq\left(\frac{e^\delta}{\left(1+\delta\right)^{1+\delta}}\right)^\mu$$

2. אם $0<\delta<1$,

$$\Pr\left[X\geq\left(1+\delta\right)\mathbb{E}\left[X\right]\right]\leq e^{-\frac{\delta^2\mu}{3}}$$

3. עבור $R>6\mu$,

$$\Pr\left[X\geq R\right]\leq 2^{-R}$$

הוכחה. (משפט 9) נשתמש במרקוב.

1.

$$\Pr\left[X\geq\left(1+\delta\right)\mu\right]\leq\frac{\mathbb{E}\left[e^{tX}\right]}{e^{t(1+\delta)\mu}}\leq\frac{e^{\mu\left(e^t-1\right)}}{e^{t(1+\delta)\mu}}$$

מחדו"א יודעים שהביטוי לעיל מינימלי כאשר $t=\ln\left(1+\delta\right)$. לכן,

$$\Pr\left[X\geq\left(1+\delta\right)\mu\right]\leq\frac{e^{\mu\delta}}{\left(1+\delta\right)^{(1+\delta)\mu}}$$

2. כאשר $0<\delta<1$, מתקיים
$$\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\le e^{-\frac{\delta^2}{3}}$$
ניתן להראות זאת ע"י לקיחת $\ln$ משני הצדדים ולהגדיר
$$f(\delta)=\delta-(1+\delta)\ln(1+\delta)+\frac{\delta^2}{3}.$$
נשים לב ש-$f'(\delta)<0$ ב-$[0,1]$, ומכיוון ש-$f(0)=0$ נקבל ש-$f(\delta)\le 0$ ב-$[0,1]$.

3. $R=(1+\delta)\mu$, אז כאשר $R>6\mu$ נקבל $\delta\ge 5$, ואז
$$\Pr[X\ge(1+\delta)\mu]\le\left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}\le\left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu}\le\left(\frac{e}{6}\right)^{R}\le 2^{-R}$$

□

הערה 3. חסם על הזנב התחתון מתקבל באותו אופן (עם $t<0$).

משפט 10. (צ'רנוף - זנב תחתון) עבור הטלת מטבעות לא הומוגניים:

1. לכל $\delta>0$,
$$\Pr[X\le(1-\delta)\mathbb{E}[X]]\le\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}$$

2. אם $0<\delta<1$,
$$\Pr[X\le(1-\delta)\mathbb{E}[X]]\le e^{-\frac{\delta^2\mu}{3}}$$

מסקנה 1. לכל $0<\delta<1$,
$$\Pr[|X-\mu|>\delta\mu]\le 2e^{-\frac{\delta^2\mu}{3}}$$

2.2.4 חסמי הופדינג (Hoeffding)

אותן תוצאות מתקבלות גם עבור מ"מ ב-$[0,1]$ (כלומר $X_i\in[0,1]$) עם תוחלת p_i. הפונקציה e^{tx} היא קמורה, ולכן תמיד נמצאת מתחת לישר שמחבר בין $(0,1)$ ל-$(1,e^t)$. קו ישר זה מתואר ע"י $y=\alpha x+\beta$, $\beta=1$, $\alpha=e^t-1$. אזי,
$$\mathbb{E}\left[e^{tX_i}\right]\le\mathbb{E}[\alpha X_i+\beta]=p_i(\alpha+\beta)+(1-p_i)\beta=1+p_i\left(e^t-1\right)$$

כלומר אותם החישובים עובדים כמו קודם.

הערה 4.

1. השיטה ניתנת ליישום עבור מ"מ אחרים (גאוסיאן, פואסון).

2. לעיתים יותר קל לחשב מומנטים ע"י פונקציה יוצרת מומנטית.

3. חסמי צ׳רנוף עובדים גם עם משתנים בעלי קורלציה שלילית:

$$\mathbb{E}\left[e^{t(X_1+X_2)}\right] \leq \mathbb{E}\left[e^{tX_1}\right]\mathbb{E}\left[e^{tX_2}\right]$$

דוגמה 14. הטלת מטבעות הוגנים: $\mu = \frac{n}{2}$,

$$\Pr\left[\left|\#H - \mu\right| > \delta\mu\right] \leq 2e^{-\frac{\delta^2\mu}{3}} \leq 2e^{-\frac{\delta^2 n}{6}}$$

ניקח $\delta = \sqrt{\frac{60}{n}}$ ונקבל שההסתברות היא לכל היותר $2e^{-10}$. בהסתברות גבוהה מספר ה-H-ים מתרכז ברדיוס $\mathcal{O}\left(\sqrt{n}\right)$.

2.3 הגברת ביטחון (Probability Amplification)

דוגמה 15. נניח שיש לנו אלגוריתם שבסיכוי 0.6 מחזיר תשובה נכונה.

- כדי להקטין את השגיאה, נריץ את האלגוריתם k פעמים ונחזיר את החלטת הרוב.
  - נניח שעל קלט מסוים האלגוריתם צריך להחזיר YES.
  - צריך במקרה זה שה"רוב" יטעו, כלומר יחזירו NO (לפחות $k/2$ פעמים).
  - תוחלת מספר ה-NO היא $0.4k$, נשתמש בצ׳רנוף:

$$\Pr\left[\#\mathrm{NO} > \left(1+\frac{1}{4}\right)\mathbb{E}\left[\#\mathrm{NO}\right]\right] \leq e^{-\frac{\mu\delta^2}{3}} = e^{-\frac{0.4k}{16\cdot 3}} \leq e^{-\frac{k}{120}},$$

לכן אם $k = \mathcal{O}\left(\log n\right)$ נקבל הסתברות כישלון של $\mathcal{O}\left(1/n\right)$.

2.4 דילול גרפים

קלט: גרף לא מכוון וממושקל $G = (V, E)$ עם פונקציית משקלים w.

פלט: גרף דליל H שהוא $(1 \pm \varepsilon)$ מקרב חתך של G.

עבור תת-קבוצה $S \subseteq V$, נסמן את החתך המושרה מ-S ב-

$$\delta_G\left(S\right) := \left\{(u,v) \in E \mid u \in S, v \notin S\right\}.$$

$$w\left(\delta_G\left(S\right)\right) = \sum_{e \in \delta_G(S)} w\left(e\right)$$

הגדרה 13. גרף H הוא $(1 \pm \varepsilon)$ מקרב חתך של G אם

$$\forall S \subseteq V: (1-\varepsilon)\, w\left(\delta_G\left(S\right)\right) \leq w\left(\delta_H\left(S\right)\right) \leq (1+\varepsilon)\, w\left(\delta_G\left(S\right)\right)$$

הנחה: נסתכל על מקרה פרטי בו גודל החתך המינימלי הוא $c = \Omega\left(\log n\right)$, וכי G הוא לא ממושקל.

Algorithm 9: GraphSparsification (G, w)

1 אתחל גרף H: $V(H) \leftarrow V(G), E(H) = \emptyset$.
2 לכל $e \in E(G)$, הוסף את e ל-H בהסתברות p.
3 החזר את H.

נסמן ב-p הסתברות דגימה כלשהי.

משפט 11. אם $p = \frac{9\ln(n)}{\varepsilon^2 \cdot c}$, אז H הוא $(1 \pm \varepsilon)$ מקרב חתך של G עם $\mathcal{O}(p|E(G)|)$ צלעות בהסתברות לפחות $1 - \frac{4}{n}$.

הוכחה. (משפט 11; ניסיון ראשון) נקבע $S \subseteq V$ ונניח ש-$\delta_G(S)$ מכיל k קשתות ($k \geq c$). מלינאריות התוחלת,

$$\mathbb{E}[\delta_H(S)] = pk \implies \mathbb{E}[w(\delta_H(S))] = pk \cdot \frac{1}{p} = |\delta_G(S)|$$

מכיוון שכל קשת היא מ"מ ב"ת, נפעיל צ'רנוף:

$$\begin{aligned}\Pr[|\delta_H(S) - pk| > \varepsilon pk] &\leq 2e^{-\frac{pk\varepsilon^2}{3}} \\ &= 2e^{-\frac{9\ln(n)}{\varepsilon^2 c} \cdot \frac{k\varepsilon^2}{3}} \\ &= 2e^{-\frac{3k\ln(n)}{c}} \\ &\leq \frac{2}{n^3}.\end{aligned}$$

□

אבחנה: ההסתברות שחתך גדול מופר (כלומר יש סטייה מהתוחלת) הרבה יותר קטנה מ-$\frac{2}{n^3}$, וגם אין הרבה חתכים קטנים.

למה 4. מספר החתכים עם לכל היותר $\alpha \cdot c$ קשתות ($\alpha \geq 1$) הוא לכל היותר $\frac{n^{2\alpha}}{2}$.

הוכחה. (משפט 11) באמצעות הלמה (שתוכח בתרגיל הבית):

$$\begin{aligned}\Pr[\text{קיימת } S \text{ שמפרה}] &\leq \sum_{S \subseteq V} \Pr[\text{S מפרה}] \\ &\leq \sum_{\alpha \geq 1} \sum_{S \subseteq V : |\delta_G(S)| \leq \alpha C} \Pr[\text{S מפרה}] \\ &\leq \sum_{\alpha \geq 1} n^{2\alpha} \cdot \Pr[\text{S מפרה} \mid |\delta_G(S)| \leq \alpha c] \\ &\leq \sum_{\alpha \geq 1} n^{2\alpha} \cdot 2e^{-3\alpha\ln(n)} \\ &\leq \sum_{\alpha \geq 1} 2 \cdot n^{-\alpha} \\ &\leq \frac{4}{n}.\end{aligned}$$

בנוסף, קל למצוא שמספר הקשתות ב-H הוא $\mathcal{O}(p|E(G)|)$ בהסתברות גבוהה. □

הערה 5. מדוע היינו צריכים להניח ש-$c = \Omega(\log n)$? קשה להתמודד עם חתכים קטנים.

2.5 כדורים ותאים

- תהליך אקראי פשוט שמתאים תופעות בסיסיות.
- נתונים m כדורים ו-n תאים, וכל כדור נזרק לתא באופן אחיד ב"ת.
- יש הרבה סיטואציות שאפשר לנתח.

דוגמה 16. תוחלת ממספר הכדורים בתא מסוים.
נסמן ב-B_{ij} את האינדיקטור למאורע שכדור j נמצא בתא i. אזי:

$$\mathbb{E}\left[\text{מספר הכדורים ב-}i\right] = \mathbb{E}\left[\sum_{j=1}^{m} B_{ij}\right] = \sum_{j=1}^{m} \mathbb{E}\left[B_{ij}\right] = \sum_{j=1}^{m} \Pr\left[\text{כדור } j \text{ בתא } i\right] = \sum_{j=1}^{m} \frac{1}{n} = \frac{m}{n}$$

מכאן, אם $n = m$ אז תוחלת מספר הכדורים בתא ה-i הוא 1.

שאלה: האם נכון לצפות שרוב התאים יכילו לפחות כדור 1?

דוגמה 17. תוחלת מספר התאים הריקים.
לכל i, נגדיר את y_i להיות האינדיקטור למאורע שהתא ה-i ריק. אזי,

$$\mathbb{E}\left[y_i\right] = \Pr\left[\text{התא ה-}i\text{ הריק}\right] = \left(1 - \frac{1}{n}\right)^m \approx e^{-\frac{m}{n}}.$$

$$\Longrightarrow \mathbb{E}\left[\text{מספר התאים הריקים}\right] = \mathbb{E}\left[\sum_{i=1}^{n} y_i\right] = \sum_{i=1}^{n} \mathbb{E}\left[y_i\right] = n \cdot e^{-\frac{m}{n}}.$$

אם $m = n$ אז בתוחלת יהיו $\frac{1}{e} \cdot n$ תאים ריקים.

2.5.1 פרדוקס יום ההולדת

שאלה: מהו המספר המקסימלי של כדורים בתא באופן טיפוסי?

- נשאל שאלה פשוטה יותר: עבור אילו m אנחנו מצפים לראות התנגשות?
- מה ההסתברות שאין התנגשויות ב-m הכדורים?

$$\begin{aligned}1 \cdot \left(1 - \frac{1}{n}\right) \cdot \cdots \cdot \left(1 - \frac{m-1}{n}\right) &\le e^{-\frac{1}{n}} \cdot e^{-\frac{2}{n}} \cdot \cdots \cdot e^{-\frac{m-1}{n}} \\ &= e^{-\frac{(m-1)m}{2n}} \\ &\approx e^{-\frac{m^2}{2n}}.\end{aligned}$$

ההסתברות קטנה מ-$\frac{1}{2}$ עבור $m = \sqrt{2n \ln 2}$.

"פרדוקס" יום ההולדת: עבור $n = 365$, אם $m > 22.5$ ההסתברות שמקסימום הכדורים בתא הוא לפחות 2 היא לפחות $\frac{1}{2}$.

לסיכום, מצפים לראות התנגשות עבור $m = \Theta\left(\sqrt{n}\right)$.

שאלה: מה ההסתברות שבתא מסוים יהיו לפחות k כדורים?

$$\binom{n}{k} \cdot \left(\frac{1}{n}\right)^k$$

נשים לב ש-

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \frac{n^k}{k!} \leq \left(\frac{en}{k}\right)^k$$

$$\Longrightarrow \binom{n}{k} \cdot \left(\frac{1}{n}\right)^k \leq \left(\frac{en}{k}\right)^k \cdot \left(\frac{1}{n}\right)^k = \frac{e^k}{k^k}$$

נשתמש בחסם האיחוד:

$$\Pr\left[\text{קיים תא עם יותר מ-}k\text{ כדורים}\right] \leq \frac{ne^k}{k^k} = e^{\ln(n)+k-k\ln(k)}$$

נחפש את ה-k המינימלי עבורו ההסתברות קטנה מספיק:

$$\arg\min_k \left\{k \ln k \geq \ln n\right\}$$

אם ניקח

$$k = \frac{3\ln n}{\ln\ln n},$$

נקבל שבהסתברות גבוהה (קבועה) מספר הכדורים המקסימלי בתא הוא

$$\mathcal{O}\left(\frac{\log n}{\log\log n}\right).$$

2.5.2 איסוף קלפי סופרגול

- עבור איזה m נצפה שלא יהיו תאים ריקים?
  - אינטואיציה: כמה קלפי סופרגול צריך לקנות, כדי שיהיה לנו קלף מכל סוג?
- נגדיר את X להיות מספר הכדורים שהוטלו עד שאין תאים ריקים.
- לכל i, נגדיר את X_i להיות מספר הכדורים שנזרקו בהינתן שיש בדיוק i תאים ריקים (בין i כדורים ריקים ועד $i+1$).
  - נשים לב ש-X_i הוא מ"מ גיאומטרי. לכן,

$$\mathbb{E}\left[X\right] = \mathbb{E}\left[\sum_i X_i\right] = \sum_i \mathbb{E}\left[X_i\right] = \sum_i \frac{n}{i} \approx n\log n.$$

2.6 השיטה ההסתברותית

- המונח "שיטה הסתברותית" משמש בדרך כלל להוכחות קיום של אובייקט כלשהו (לדוגמא גרף עם תכונות כלשהן).
- נרצה להראות שעבור מרחב מדגם כלשהו, המאורע קורה בהסתברות לא 0.
- לעיתים ניתן בנוסף להוכחת הקיום גם להשיג את האובייקט.
- נסתכל על דוגמא פשוטה אך שימושית - חתך מקסימלי.

עובדה 2. (עיקרון הממוצע) יהי χ מ"מ מעל Ω, אז קיים מאורע בסיסי $\omega \in \Omega$ כך ש-

$$\chi\left(\omega\right) \geq \mathbb{E}\left[\chi\right].$$

אחרת, התוחלת הייתה קטנה יותר.

2.6.1 Max Cut

קלט: גרף לא מכוון ולא ממושקל $G = (V, E)$.

פלט: חתך C בגודל מקסימלי.

תזכורת: עבור גרף $G = (V, E)$, חתך הוא חלוקה של הקודקודים ל-(V_1, V_2) כך ש-$V_1 \sqcup V_2 = V$. גודל החתך מוגדר להיות מספר הקשתות שחוצות את החתך.

- נסמן $m = |E|$, נראה שתמיד קיים חתך בגודל $m/2$, ואז איך למצוא אותו.
- נסתכל על בחירה מקרית של חתך (V_1, V_2) כך שכל קודקוד $v \in V$ נבחר ל-V_1 בהסתברות $\frac{1}{2}$ ול-V_2 בהסתברות $\frac{1}{2}$.
- מכאן מושרה מרחב הסתברות $\{1, 2\}^n$, בו לכל חתך יש הסתברות $1/2^n$.
- לכל קשת $e \in E$ נגדיר מ"מ χ_e שהוא $1 \iff e$ חוצה את החתך:

$$\forall e = (u, v) \in E : \chi_e\left(V_1, V_2\right) = 1 \iff u \in V_1 \wedge v \in V_2 \vee u \in V_2 \wedge v \in V_1$$

מכאן, לכל $e \in E$ מתקיים

$$\Pr\left[\chi_e = 1\right] = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}.$$

ולכן גם $\mathbb{E}\left[\chi_e\right] = \frac{1}{2}$.

- גודל החתך הוא $\chi = \sum_{e \in E} \chi_e$.

$$\implies \mathbb{E}\left[\chi\right] = \sum_{e \in E} \mathbb{E}\left[\chi_e\right] = \frac{m}{2}$$

מעובדה 2, קיים חתך בגודל לפחות $m/2$.

Algorithm 10: MaxCut (G, δ)

1 חזור $S = m\log\left(\frac{1}{\delta}\right)$ פעמים:
2 בחר חלוקה אקראית (V_1, V_2).
3 חשב את את $|E(V_1, V_2)|$.
4 החזר את החתך המקסימלי שנמצא.

- כדי לקבל אלגוריתם רנדומי, נרצה להראות חסם תחתון על ההסתברות שגודל החתך הוא לפחות $m/2$.
- לפשטות, נניח ש-m זוגי. אזי,

$$\begin{aligned}\frac{m}{2} = \mathbb{E}[\chi] &= \sum_{k=0}^{m} k\cdot\Pr[\chi = k]\\ &= \sum_{k<\frac{m}{2}} k\cdot\Pr[\chi=k] + \sum_{k\geq\frac{m}{2}} k\cdot\Pr[\chi=k]\\ &\leq \left(\frac{m}{2}-1\right)\sum_{k<\frac{m}{2}}\Pr[\chi=k] + m\sum_{k\geq\frac{m}{2}}\Pr[\chi=k]\\ &= \left(\frac{m}{2}-1\right)\cdot\Pr\left[\chi<\frac{m}{2}\right] + m\cdot\Pr\left[\chi\geq\frac{m}{2}\right]\\ &< \left(\frac{m}{2}-1\right) + m\cdot\Pr\left[\chi\geq\frac{m}{2}\right].\end{aligned}$$

$$\implies \Pr\left[\chi\geq\frac{m}{2}\right] > \frac{1}{m} \tag{2.1}$$

משפט 12. בהסתברות לפחות $1-\delta$, גודל החתך המוחזר מאלגוריתם 10 הוא לפחות $m/2$.

הוכחה. (משפט 12) ממשוואה (2.1), בכל איטרציה בהסתברות לפחות $1/m$ נקבל חתך שגודלו לפחות $m/2$. לכן, ההסתברות לקבל חתך בגודל פחות מ-$m/2$ היא

$$\geq 1-\frac{1}{m}.$$

כיוון שיש S איטרציות ב"ת, ההסתברות שבכל האיטרציות נקבל חתך קטן היא

$$\geq \left(1-\frac{1}{m}\right)^S = \left(1-\frac{1}{m}\right)^{m\cdot\log\left(\frac{1}{\delta}\right)} = e^{-\log\left(\frac{1}{\delta}\right)} = \delta.$$

□

Algorithm 11: $\texttt{MaxSAT} - \texttt{V1}\left(\varphi, m, n, \delta \in (0, 1/2)\right)$

1. חזור $S = m \log\left(\frac{1}{\delta}\right)$ פעמים:
2. בחר השמה אקראית וחשב את מספר הפסוקיות שמסתפקות.
3. החזר את ההשמה שסיפקה את המספר המקסימלי של פסוקיות.

2.6.2 Max SAT

קלט: נוסחת CNF φ.

פלט: השמה המספקת מספר מקסימלי של פסוקיות.

תזכורת: נוסחת CNF מעל משתנים $x_1, \ldots, x_n$ היא מהצורה

$$\varphi = \bigwedge_{j=1}^{m} c_i,$$

כאשר c_i היא פסוקית שהיא $\bigvee$ של ליטרלים (x או $\neg x$).

למה 5. בהינתן נוסחת CNF φ, שבה כל פסוקית מכילה לפחות k ליטרלים. אזי, תוחלת מספר הפסוקיות שמספקות (עבור השמה מקרית) היא לפחות $m\left(1 - 2^{-k}\right)$.

הוכחה. (למה 5) מרחב ההסתברות הוא כל ההשמות ל-n משתנים: $\{0, 1\}^n$.

- עבור פסוקית כלשהי c_i, נגדיר מ"מ χ_i שהוא $1 \iff c_i$ מסתפקת.
- כדי ש-c_i לא תסתפק, כל הליטרלים ב-c_i צריכים לקבל 0, ולכן ההסתברות היא $\leq 2^{-k}$.
- לכן, ההסתברות ש-c_i מסתפקת היא לפחות $1 - 2^{-k}$, וכך $\mathbb{E}\left[\chi_i\right] \geq 1 - 2^{-k}$.
- כמו קודם, נסמן ב-χ את מספר הפסוקיות שמסתפקות: $\chi = \sum_i \chi_i$. אזי,

$$\mathbb{E}\left[\chi\right] = \sum_{i=1}^{m} \mathbb{E}\left[\chi_i\right] \geq m\left(1 - 2^{-k}\right).$$

משום ש-$k \geq 1$, נקבל ש-$\mathbb{E}\left[\chi\right] \geq m/2$.

□

משפט 13. בהסתברות לפחות $1 - \delta$, פלט אלגוריתם 11 מספק לפחות $m/2$ פסוקיות.

הערה 6. הוכחת משפט 13 ברורה. בנוסף - תוצאה זו אינה משמעותית: נוכל לבדוק השמה של 0-ים לעומת השמה של 1-ים, והטובה מבין השתיים תספק לפחות $m/2$ פסוקיות.

2.6.3 קודים לינאריים

- קיום של קודים טובים לתיקון שגיאות.
- נתונים לנו פרמטרים n, k, d (כאשר $k, d < n$).
- נרצה למצוא פונקציה $E : \{0,1\}^k \to \{0,1\}^n$, כך שלכל $x, y \in \{0,1\}^k$ מתקיים

$$\text{dist}\left(E\left(x\right), E\left(y\right)\right) > d,$$

כאשר המרחק הוא hamming distance:

$$\text{dist}\left(x, y\right) = \left|\left\{i \in [n] \mid x_i \neq y_i\right\}\right|$$

- $\{0,1\}^k$ הוא מרחב ההודעות, ו-$\{E\left(x\right)\}_{x \in \{0,1\}^k} \subseteq \{0,1\}^n$ מרחב מילות הקוד.
- עבור מילה x, המשקל של x מוגדר להיות

$$w\left(x\right) \coloneqq \left|\left\{i \in [n] \mid x_i \neq 0\right\}\right|.$$

שים לב: $\text{dist}\left(x, y\right) = w\left(x \oplus y\right)$.

- נגיד שקוד הוא <u>טוב</u> אם $\frac{k}{n}$ (קצב הקוד) ו-$\frac{d}{n}$ (מרחק יחסי) הם גבוהים (קבוע גלובלי כלשהו).

אבחנה:

- אם $E\left(x\right)$ הוא תוצאה של הכפלה של x במטריצה G מעל $\{0,1\}^{k \times n}$ (מטריצה יוצרת), אז מקבלים קוד לינארי.
- אוסף מילות הקוד הוא תת-מרחב לינארי של $\{0,1\}^n$ ממימד k.
- לכן, סכום כל שתי מילות קוד הוא גם מילת קוד.
- אבל, משקל של וקטור הסכום של מילות קוד שווה למרחק ביניהן.
- לכן, חסם תחתון על המרחק המינימלי שקול לחסם תחתון על מילה עם משקל מינימלי.

נראה שאם נבחר מטריצה G מעל $\{0,1\}^{k \times n}$ באופן אחיד, אז בהסתברות חיובית נקבל קוד טוב.

טענה 6. אם $d < n/2$ ניתן לקבל קוד עם קצב

$$\frac{k}{n} = 1 - H\left(\frac{d}{n}\right),$$

כאשר H היא פונקציית האנטרופיה:

$$H\left(p\right) = -p \log p - \left(1 - p\right) \log\left(1 - p\right).$$

הוכחה. (טענה 6) נבחר את k השורות של G אחת אחרי השנייה, מ-$i = 1$ ל-$i = k$.

- נסמן ב-G_i את תת המטריצה של G שמתקבלת מ-i השורות הראשונות.

- נראה שלכל i, אם G_{i-1} מקיימת שלכל $x \in \{0,1\}^{i-1}$, $x \neq \vec{0}$, $w(xG_{i-1}) \geq d$, אז בהסתברות לא 0 מעל בחירת השורה ה-i מתקיים

$$\forall x \in \{0,1\}^i, x \neq \vec{0} : w(xG_i) \geq d.$$

- נסתכל על $i \leq k$, ונסמן את השורה החדשה ב-R_i.
- לכל וקטור $x \in \{0,1\}^i$ מתקיים

$$xG_i = x'G_{i-1} \oplus x_i R_i, \quad x' = \begin{pmatrix} x_1 & \cdots & x_{i-1} \end{pmatrix}.$$

- ידוע כי לכל $x \in \{0,1\}^{i-1}$ שאינו 0, המשקל של הוקטור $y' = x'G_{i-1}$ הוא לפחות d.
- אם $x_i = 0$, אז $w(xG_i) = w(x'G_{i-1}) \geq d$ לכל $x \in \{0,1\}^i$.
- אחרת, נגדיר לכל $j \in \{1,\ldots,n\}$ מ"מ $\chi_j = y'_j \oplus R_i[j]$. אזי,

$$\sum_{j=1}^{n} \chi_S = w(xG_i).$$

נשים לב ש-$\Pr[\chi_j = 1] = \frac{1}{2}$, ואז

$$\Pr\left[\sum_j \chi_j < d\right] = \frac{1}{2^n} \sum_{t=0}^{d-1} \binom{n}{t}.$$

- בנוסף, עבור $d \leq n/2$ מתקיים $2^{n \cdot H\left(\frac{d}{n}\right)}$, ולכן

$$\Pr\left[\sum_j \chi_j < d\right] \leq 2^{-n\left(1-H\left(\frac{d}{n}\right)\right)}.$$

מחסם איחוד, נקבל

$$2^{i-1} 2^{-n\left(1-H\left(\frac{d}{n}\right)\right)} < 1,$$

שמתקיים עבור $i \leq n(1 - H(d/n))$.

מכאן, קיבלנו כי עבור $k \leq n(1 - H(d/n))$, בהסתברות לא אפס נקבל קוד טוב. $\square$

2.7 תכנות לינארי

תוכנית לינארית נותנת דרך לפרמל הרבה בעיות אופטימיזציה, ולפתור אותן (אולי באופן מקורב). תוכנית לינארית מוגדרת ע"י אוסף אילוצים לינאריים מעל סט משתנים, יחד עם פונקציית מטרה לינארית מעל המשתנים.

המטרה: למקסם (או למנמם) את פונקציית המטרה תחת האילוצים.

- באופן פורמלי, נסמן ב-$\{x_i\}_{i=1}^n$ את המשתנים מעליהם התוכנית מוגדרת.
- פונקציית המטרה היא מהצורה

$$\text{maximize} \quad \sum_{i=1}^n c_i x_i \quad = \vec{c} \cdot \vec{x}$$

עבור $c_1, \ldots, c_n$ כלשהם.

- האילוצים בדרך כלל נכתבים בצורת מטריצה $Ax \leq b$ עבור $A \in \mathbb{R}^{m \times n}$.
  - כלומר, יש m אילוצים כך שהאילוץ ה-i הוא מהצורה

$$\sum_{j=1}^n A_{ij} x_j \leq b_j.$$

- פתרון לתוכנית לינארית הוא השמה למשתני התוכנית שעומדת באילוצים ומאפטמת את פונקציית המטרה.

דוגמה 18. לחוואי יש פיסת אדמה בגודל S. בה הוא רוצה לשתול שני סוגים של יבולים: חיטה ושעורה.

- הרווח של חיטה הוא p_1 ליחידת שטח, ושל שעורה p_2.
- כל סוג יבול צריך שני סוגים של דשנים:
  - חיטה זקוקה ל-F_{11} מהדשן הראשון, ו-F_{21} מהדשן השני.
  - שעורה זקוקה ל-F_{12} מהדשן הראשון, ו-F_{22} מהדשן השני.
- יש לחוואי כמות של B_1 מהדשן הראשון ו-B_2 מהשני.

נפרמל את הבעיה בתוכנית לינארית:

- נגדיר את x_1 כשטח עליו נשתול חיטה, ו-x_2 כשטח עליו נשתול שעורה. אזי:

$$\begin{array}{lrl}
\text{maximize} & p_1 x_1 + p_2 x_2 & \\
\text{subject to} & x_1 + x_2 \leq S & (\text{אילוץ שטח}) \\
& F_{11} \cdot x_1 + F_{12} \cdot x_2 \leq B_1 & (\text{כמות החיטה}) \\
& F_{21} \cdot x_1 + F_{22} \cdot x_2 \leq B_2 & (\text{כמות השעורה}) \\
& x_1, x_2 \geq 0 &
\end{array}$$

- ניתן להסתכל על הבעיה כבעיה גיאומטרית: כל אילוץ מייצג על-מישור במרחב, ואזור הפתרונות ה-feasible הם בחיתוך של ה-hyperplanes.
- איך פותרים תוכנית לינארית?
  - Simplex - מדויק אך לא פולינומיאלי.
  - אלגוריתמים פולינומיאלים אחרים.
- להרבה בעיות מעניינות נרצה פתרונות בהם המשתנים מקבלים ערכים שלמים (Integer Linear Program או ILP).

2.7.1 כיסוי קודקודים

תזכורת:

קלט: הוא גרף $G = (V, E)$, $|V| = n$.

פלט: תת-קבוצה $C \subseteq V$ כך שלכל $(u, v) \in E$ מתקיים $\{u, v\} \cap C \neq \emptyset$.

נבנה תוכנית לינארית שמתארת את הבעיה:

- לכל קודקוד $v \in V$ נגדיר משתנה בינארי $x_v \in \{0, 1\}$.

$$\begin{array}{lll} \text{minimize} & \sum_{v \in V} x_v & \\ \text{subject to} & x_u + x_v \geq 1 & \forall (u, v) \in E \\ & x_v \in \{0, 1\} & \forall v \in V \end{array}$$

למה 6. לכל גרף G, בהינתן פתרון ל-VC-ILP (G) ניתן לקבל כיסוי קודקודים בגודל מינימלי.

הוכחה. (למה 6) יהי $\{x_v^\star\}_{v \in V}$ פתרון ל-VC-ILP (G). נגדיר $C = \{v \in V \mid x_v^\star = 1\}$.

- נראה תחילה ש-C היא כיסוי:
  - לכל קשת $(u, v) \in E$ מתקיים $x_u^\star + x_v^\star \geq 1$, ולכן אחד מ-$\{u, v\}$ ב-C והקשת מכוסה ע"י C.
  - מכיוון שהטענה נכונה לכל קשת, C הוא כיסוי.
- נראה ש-C היא בגודל מינימלי:
  - נניח בשלילה שקיים כיסוי C' כך ש-$|C'| < |C|$.
  - נגדיר השמה x' למשתנים כך שלכל $v \in V$, $x'_v = \delta_{v \in C'}$.
  - מכיוון ש-C' הוא כיסוי, לכל קשת $(u, v) \in E$ מתקיים $x'_u + x'_v \geq 1$. עם זאת,

$$\sum_{v \in V} x'_v = |C'| < |C| = \sum_{v \in V} x_v^\star,$$

בסתירה לאופטימליות של $x^\star$.

□

- כעת נחליש (relax) את האילוצים הבינאריים:

$$\forall v \in V : x_v \in [0, 1] .$$

- כעת נותרו עם LP, VC-LP (G), ואותה ניתן לפתור באופן יעיל.
- נסמן $\text{OPT}_{LP}(G)$ להיות הפתרון האופטימלי של VC-LP (G), ו-$\text{OPT}_{ILP}(G)$ עבור VC-ILP (G). נשים לב ש-

$$\text{OPT}_{LP}(G) \leq \text{OPT}_{ILP}(G) .$$

- נסמן ב-$\{x'_v\}_{v\in V}$ פתרון ל-$\text{VC-LP}\,(G)$. אזי,

$$\sum_{v\in V} x'_v = \text{OPT}_{LP}\,(G) \le \text{OPT}_{ILP}\,(G)\,.$$

- עם זאת, $\{x'_v\}_{v\in V}$ לא בהכרח פתרון ל-$\text{VC-ILP}\,(G)$.

רעיון: נעגל את הפתרון. נגדיר את הכיסוי להיות

$$C = \{v \in V \mid x'_v \ge 1/2\}\,.$$

- נשים לב ש-C הוא כיסוי חוקי: לכל $(u,v)\in E$ מתקיים $x'_u + x'_v \ge 1$, ולכן אחד מהם הוא לפחות חצי. מכאן, הקשת מכוסה ע"י C.
- נראה ש-C הוא קירוב 2 ל-$\text{OPT}_{ILP}\,(G)$:
  - לכל $v\in V$ נגדיר השמה $\tilde{x}_v$:

$$\tilde{x}_v = \begin{cases} 1 & x'_v \ge 1/2 \\ 0 & \text{אחרת} \end{cases}\,.$$

$$\tilde{x}_v \le 2x'_v \implies \sum_{v\in V}\tilde{x}_v \le 2\sum_{v\in V} x'_v$$

$$|C| = \sum_{v\in V}\tilde{x}_v \le 2\sum_{v\in V} x'_v = 2\cdot \text{OPT}_{LP}\,(G) \boxed{\le 2\cdot \text{OPT}_{ILP}\,(G)}$$

הערה 7. הניסוח של הבעיה כתוכנית לינארית מקל על ההכללה לבעיה הממושקלת: לכל קודקוד $v\in V$ יש משקל w_v, והמטרה למצוא כיסוי במשקל מינימלי ($\text{WVC-ILP}\,(G)$):

$$\begin{array}{ll} \text{minimize} & \sum_{v\in V} w_v x_v \\ \text{subject to} & x_u + x_v \ge 1 \quad \forall\,(u,v)\in E \\ & x_v \in \{0,1\} \quad \forall v\in V \end{array}$$

ועבור $\text{WVC-LP}\,(G)$, הדרישה תתחלף ל-$x_v \in [0,1]$. ניתן להשיג 2-קירוב באותו האופן.

2.7.2 Max SAT

תזכורת:

קלט: נוסחת CNF φ.

פלט: השמה שמספקת מספר מקסימלי של פסוקיות.

נגדיר את התוכנית הלינארית בשלמים ל-Max-SAT, $\text{SAT-ILP}\,(\varphi)$:

- לכל משתנה x_i ב-φ נגדיר משתנה y_i.
- לכל פסוקית c_j בנוסחא נגדיר משתנה z_j בתוכנית.

- לכל פסוקית c_j נגדיר S_j^+ להיות אוסף האינדקסים של המשתנים שמופיעים בפסוקית ללא שלילה, ובאופן דומה עבור S_j^- ומשתנים המופיעים בשלילה.
  - למשל, עבור $c_j = x_1 \vee \neg x_3 \vee x_7$, $S_j^+ = \{1, 7\}$ ו-$S_j^- = \{3\}$.
- כעת נגדיר את התוכנית:

$$
\begin{array}{lll}
\text{maximize} & \sum_{j=1}^{m} z_j & \\
\text{subject to} & \sum_{i \in S_j^+} y_i + \sum_{i \in S_j^-} (1 - y_i) \geq z_j & \forall j \in [m] \\
& y_i \in \{0,1\} & \forall i \in [n] \\
& z_j \in \{0,1\} & \forall j \in [m]
\end{array}
$$

דוגמה 19. נסתכל על הנוסחא

$$\varphi = (x_1 \vee x_2) \wedge (x_1 \vee \neg x_2 \vee \neg x_3) \wedge (x_2 \vee x_4)\,.$$

התוכנית היא

$$
\begin{array}{ll}
\text{maximize} & z_1 + z_2 + z_3 \\
\text{subject to} & y_1 + y_2 \geq z_1 \\
& y_1 + (1 - y_2) + (1 - y_3) \geq z_2 \\
& y_2 + y_4 \geq z_3 \\
& y_1, \ldots, y_4, z_1, \ldots, z_3 \in \{0,1\}
\end{array}
$$

למה 7. לכל φ, בהינתן פתרון ל-$\text{SAT-ILP}(\varphi)$, נוכל לקבל השמה $x_1, \ldots, x_n$ כך שמספר הפסוקיות שמסתפקות הוא מקסימלי.

הוכחה. (למה 7) יהי $\left(z_j^\star\right)_{j=1}^m, \left(y_i^\star\right)_{i=1}^n$ פתרון ל-$\text{SAT-ILP}(\varphi)$.

- נגדיר השמה $\alpha : \{x_1, \ldots, x_n\} \to \{0,1\}$ כך ש-$\alpha(x_i) = y_i^\star$. נרצה להוכיח ש-α מספקת מקסימום פסוקיות.
- נראה שמספר הפסוקיות שמסתפקות ע"י ההשמה α שווה בדיוק ל-$\sum_{j=1}^m z_j^\star$:
- נסתכל על פסוקית c_j:

1. אם c_j מסתפקת ע"י α, קיים i' כך ש-$i' \in S_j^+$ ו-$\alpha(x_{i'}) = 1$, או ש-$i' \in S_j^-$ ו-$\alpha(x_{i'}) = 0$.
   - מכיוון שלכל i מתקיים $\alpha(x_i) = y_i^\star$, אם $y_{i'}^\star = 1$ אז $\sum_{i \in S_j^+} y_i^\star \geq 1$ ואחרת $\sum_{i \in S_j^-} (1 - y_i^\star) \geq 1$.
   - בשני המקרים, כיוון ש-$z_j^\star$ הוא פתרון אופטימלי חייב להתקיים כי $z_j^\star = 1$ (אחרת היינו יכולים לשפר).
2. אם c_j לא מסתפקת ע"י α, בהכרח $\alpha(x_i) = 0$ לכל $i \in S_j^+$ ו-$\alpha(x_i) = 1$ לכל $i \in S_j^-$. מכאן,

$$\sum_{i \in S_j^+} y_i + \sum_{i \in S_j^-} (1 - y_i) = 0 \implies z_j^\star = 0$$

- בסך הכל, קיבלנו כי $\sum_j z_j^\star$ הוא מספר הפסוקיות שמסופקות.
- נניח בשלילה שקיימת α' שמספקת יותר פסוקיות מ-α.
  - נגדיר השמה למשתנים באופן הבא:

$$y_i' = \alpha'(x_i)$$
$$z_j' = \begin{cases} 1 & \sum_{i\in S_j^+} y_i' + \sum_{i\in S_j^-}(1-y_i') \geq 1 \\ 0 & \text{אחרת} \end{cases}$$

נשים לב שההשמה היא פתרון פיזבילי ל-ILP: מספקת את כל האילוצים.

  - בנוסף, $\sum z_j'$ הוא מספר הפסוקיות שמסתפקות ע"י ההשמה, בסתירה לאופטימליות.

□

באופן דומה ל-VC, נבצע רלקסציה כדי לקבל LP: נחליף את האילוצים הבינאריים באילוצים

$$\forall i,j : y_i, z_j \in [0,1],$$

לקבלת התוכנית $\text{SAT-LP}(\varphi)$.

- נניח שיש לנו פתרון ל-$\text{SAT-LP}(\varphi)$, $y_1',\ldots,y_n',z_1',\ldots,z_m'$.
- נבצע <u>עיגול אקראי</u>: נקבע $\alpha'(x_i)=1$ בהסתברות y_i'.

משפט 14. תוחלת מספר הפסוקיות שההשמה α' מספקת היא לפחות $(1-1/e)\,\text{OPT}$.

למה 8. נניח ש-c_j היא פסוקית עם k ליטרלים. אז, ההסתברות ש-c_j מסתפקת עם α' היא לפחות $\beta(k)\cdot z_j'$, כאשר

$$\beta(k) = 1-\left(1-\frac{1}{k}\right)^k.$$

הוכחה. (משפט 14) נובעת ישירות מלמה 8: מכיוון ש-$\text{SAT-LP}(\varphi)$ היא רלקסציה של $\text{SAT-ILP}(\varphi)$, מתקיים $\sum_{j=1}^m z_j' \geq \text{OPT}$.

- יהיו $\chi_1,\ldots,\chi_m \in \{0,1\}$ מ"מ כך ש-$\chi_j=1 \iff \alpha'$ מספקת את c_j.
- נסמן ב-k_j את מספר הליטרלים ב-c_j.
- לפי למה 8,

$$\mathbb{E}[\chi_j] = \Pr[\chi_j=1] \geq \beta(k_j)\cdot z_j^\star = \left(1-\left(1-\frac{1}{k_j}\right)^{k_j}\right) z_j' \geq \left(1-\frac{1}{e}\right) z_j'.$$

לבסוף, מלינאריות התוחלת,

$$\sum_{j=1}^m \mathbb{E}[\chi_j] \geq \left(1-\frac{1}{e}\right)\sum_{j=1}^m z_j' \geq \left(1-\frac{1}{e}\right)\text{OPT}.$$

Algorithm 12: $\texttt{MaxSAT}-\texttt{V2}\left(\varphi,m,n,\delta\in\left(0,1/2\right)\right)$

1 הרץ את $\texttt{MaxSAT-V1}\left(\varphi,m,n,\delta\right)$.
2 מצא פתרון עבור $\texttt{SAT-LP}\left(\varphi\right)$ ובצע עיגול אקראי.
3 החזר את הפתרון הטוב יותר.

□

הוכחה. (למה 8) נניח בה"כ ש-$c_j=x_1\vee x_2\vee\cdots\vee x_k$ (ניתן להגדיר משתנים אחרים והשמה אחרת כך שנקבל את אותם האילוצים, ו-c_j תהיה מהצורה הזו), אזי

$$S_j^+=\left\{1,\ldots,k\right\},S_j^-=\emptyset.$$

$$\implies z_j'\leq\sum_{i\in S_j^+}y_i'+\sum_{i\in S_j^-}\left(1-y_i'\right)=\sum_{i\in S_j^+}y_i'$$

מכאן, ההסתברות ש-c_j לא מסתפקת היא

$$\prod_{i=1}^{k}\left(1-y_i'\right).$$

מאי-שוויון הממוצעים, נקבל

$$\left(\prod_{i=1}^{k}\left(1-y_i'\right)\right)^{1/k}\leq\frac{1}{k}\sum_{i=1}^{k}\left(1-y_i'\right).$$

$$\begin{aligned}\implies \Pr\left[\text{מסתפקת } c_j\right]&=1-\prod_{i=1}^{k}\left(1-y_i'\right)\\&\geq1-\left(\frac{1}{k}\sum_{i=1}^{k}\left(1-y_i'\right)\right)^k\\&=1-\left(1-\frac{1}{k}\sum_{i=1}^{k}y_i'\right)^k\\&\geq1-\left(1-\frac{1}{k}z_j'\right)^k\\&\geq\beta\left(k\right)\cdot z_j'.\end{aligned}$$

כאשר המעבר האחרון נובע מטיעון קמירות. □

טענה 7. אלגוריתם 12 משיג קירוב $3/4$.

הוכחה. (טענה 7)

- נסמן ב-n_1 את המשתנה המקרי ששווה למספר הפסוקיות המסופקות שבוחר Max-SAT-V1, וב-n_2 את מספר הפסוקיות המסופקות ע"י עיגול אקראי.
- נראה ש-
$$\mathbb{E}\left[\max\left\{n_1,n_2\right\}\right] \geq \frac{3}{4}\sum_{j=1}^{m} z'_j \geq \frac{3}{4}\text{OPT}.$$
- מכיוון ש-$\max\left\{n_1,n_2\right\} \geq \left(n_1+n_2\right)/2$, מספיק להראות כי
$$\mathbb{E}\left[n_1\right]+\mathbb{E}\left[n_2\right] \geq \frac{3}{2}\text{OPT}.$$
- מלמה 8, ידוע כי
$$\mathbb{E}\left[n_2\right] = \sum_{k=1}^{n}\sum_{j:|c_j|=k}\beta\left(k\right)z'_j.$$
- עבור השמה מקרית, ההסתברות ש-c_j בגודל k מסופקת הוא $1-2^{-k}$. מכאן,
$$\mathbb{E}\left[n_1\right] = \sum_{k=1}^{n}\sum_{j:|c_j|=k}\left(1-2^{-k}\right) \geq \sum_{k=1}^{n}\sum_{j:|c_j|=k}\left(1-2^{-k}\right)z'_j.$$
$$\begin{aligned}\implies \mathbb{E}\left[n_1\right]+\mathbb{E}\left[n_2\right] &\geq \sum_{k=1}^{n}\sum_{j:|c_j|=k}\left(\beta\left(k\right)+\left(1-2^{-k}\right)\right)z'_j \\ &\geq \sum_{k=1}^{n}\sum_{j:|c_j|=k}\frac{3}{2}z'_j \\ &\geq \frac{3}{2}\text{OPT}.\end{aligned}$$
- ניתן להראות ש-$\forall k\geq 1:\beta\left(k\right)+\left(1-2^{-k}\right)\leq 3/2$ ע"י הפרדה למקרים.

□

2.7.3 Metric Facility Location

קלט: אוסף ערים $C=[n]$, מחירי מפעלים $\left(f_j\right)_{j=1}^{m}$ ומרחקים $\left(d_{ij}\right)_{i,j\in C\times[m]}$.
פלט: אוסף מפעלים $O\subseteq[m]$ ושיוך $j:C\to[m]$ כך שהמחיר הכולל מינימלי.

- המחיר של לפתוח מפעל $j\in\{1,\ldots,m\}$ הוא f_j.
- המרחק בין עיר i למפעל j הוא d_{ij}.
- המפעלים והערים נמצאים במרחב מטרי (d_{ij} מקיים את אי-שוויון המשולש).

- נרצה לפתוח אוסף מפעלים $O \subseteq [m]$, ולשייך כל עיר i למפעל $j(i) \in O$, כך שהמחיר הכולל,

$$\sum_{j \in O} f_j + \sum_{i \in C} d_{i,j(i)} \tag{2.1}$$

יהיה מינימלי.

- נשים לב ש-$\sum_{i \in C} d_{i,j(i)}$ מקבל מינימום כאשר לכל i,

$$\tilde{j}(i) = \arg\min_{j \in O} d_{ij}.$$

ננסח את הבעיה כ-ILP:

- לכל מפעל $j \in [m]$ נגדיר משתנה $y_j \in \{0,1\}$ ($y_j = 1 \iff j$ נפתח).
- לכל $i \in C, j \in [m]$ נגדיר משתנה $x_{i,j} \in \{0,1\}$ ($x_{ij} = 1 \iff$ עיר i משויכת למפעל j).
- מחיר פתיחת המפעלים הוא

$$F(y) = \sum_{j=1}^{m} f_j y_j.$$

- מחיר השירות הכולל הוא

$$D(x) = \sum_{i=1}^{n} \sum_{j=1}^{m} d_{ij} x_{ij}.$$

- התוכנית הלינארית ($\mathrm{MFL\text{-}ILP}(\{f_j\}, \{d_{ij}\})$) היא:

$$\begin{array}{rll} \text{minimize} & F(y) + D(x) & \\ \text{subject to} & \sum_{j=1}^{m} x_{ij} = 1 & \forall i \in C \\ & y_j \geq x_{ij} & \forall i \in C, j \in [m] \\ & x_{ij}, y_j \in \{0,1\} & \forall i \in C, j \in [m] \end{array}$$

למה 9. לכל $(f_j)_{j=1}^{m}$ ו-$(d_{ij})_{i,j \in [n] \times [m]}$, בהינתן פתרון ל-$\mathrm{MFL\text{-}ILP}((f_j), (d_{ij}))$ ניתן לקבל פתרון אופטימלי ל-MFL.

הוכחה. (למה 9) יהיו $x^\star$ ו-$y^\star$ פתרונות ל-$\mathrm{MFL\text{-}ILP}((f_j), (d_{ij}))$.

- נגדיר

$$O^\star = \left\{ j \in [m] \mid y_j^\star = 1 \right\},$$

וניקח את $j^\star(i)$ להיות j כך ש-$x_{i,j}^\star = 1$. מכיוון ש-$x_{ij} \in \{0,1\}$ ו-$\sum_{j \in [m]} x_{ij} = 1$ אז קיים $j^\star(i)$ יחיד לכל i.

- בנוסף, בגלל האילוץ $y_j^\star \geq x_{i,j^\star(i)}^\star$ ולכן $j^\star(i) \in O^\star$.

- מהגדרת הפונקציות F, D, $F\left(y^{\star}\right)+D\left(x^{\star}\right)$ הוא בדיוק המחיר ממשוואה (2.1).
- נותר להראות שהפתרון שהגדרנו אופטימלי.
  - נניח בשלילה שקיים O' והשמה $j'\left(i\right)$ כך שמקבלים מחיר קטן יותר.
  - נגדיר את ההשמה הבאה למשתני MFL-ILP:

$$y'_j = 1 \iff j \in O', \quad x'_{ij} = 1 \iff j'\left(i\right) = j.$$

  - השמה זו פיזבילית, ו-$F\left(y'\right)+D\left(x'\right)$ הוא המחיר של O' ו-j', ולכן מקבלים סתירה לאופטימליות של $x^{\star}$ ו-$y^{\star}$ ב-MFL-ILP.

□

- נבצע רלקסציה לבעיה לקבלת התוכנית MFL-LP ע"י החלפת האילוצים ל-

$$\forall i, j : y_j, x_{ij} \in [0, 1] .$$

- נסמן ב-$x^{\star}, y^{\star}$ את הפתרון (אופטימלי) ל-MFL-LP, שמחירו הוא $F\left(y^{\star}\right)+D\left(x^{\star}\right)$.
- לפני שנבצע עיגול, נעשה מסאז' לפתרון $x^{\star}, y^{\star}$ כדי לקבל פתרון אחר, x', y' שהוא תת-אופטימלי אך פיזבילי.

סימונים:

- עבור $i \in C$ נסמן ב-$D_i^{\star} = \sum_{j=1}^{m} d_{ij} x_{ij}^{\star}$ את התרומה של i ל-$D\left(x^{\star}\right)$.
- נסמן ב-

$$N_i = \left\{ j \in [m] \mid y_j^{\star} > 0 \wedge d_{ij} \leq 2D_i^{\star} \right\}$$

אוסף מפעלים שקרובים במידה מה ל-i (בכדור שמרכזו ב-i עם רדיוס $2D_i^{\star}$).

טענה 8. לכל $i \in C$,

$$\sum_{j \in N_i} x_{ij}^{\star} \geq \frac{1}{2}.$$

הוכחה. (טענה 8)

$$D_i^{\star} = \sum_{j=1}^{m} d_{ij} x_{ij}^{\star} \geq \sum_{j \notin N_i} d_{ij} x_{ij}^{\star} > 2D_i^{\star} \sum_{j \notin N_i} x_{ij}^{\star} = 2D_i^{\star} \left(1 - \sum_{j \in N_i} x_{ij}^{\star}\right)$$

$$\implies \sum_{j \in N_i} x_{ij}^{\star} \geq \frac{1}{2}$$

□

Algorithm 13: $\texttt{MFL}\left(\left(f_j\right),\left(d_{ij}\right),\left(x^\star_{ij}\right),\left(y^\star_j\right)\right)$

אתחל $\hat{y}_j=0,\hat{x}_{ij}=0$ לכל i,j.
כל עוד יש עיר שלא שויכה למפעל:
  בחר את העיר i עבורה $D^\star_i$ מינימלי.
  בחר את $j\left(i\right)\in N_i$ עם מינימום $f_{j(i)}$.
  הגדר $\hat{y}_{j(i)}=1$ ו-$\hat{x}_{i,j(i)}=1$.
  לכל עיר i' עבורה $N_i\cap N_{i'}\neq\emptyset$:
    הגדר $j\left(i'\right)\leftarrow j\left(i\right)$, $\hat{x}_{i',j(i)}=1$ ו-$\hat{y}_{j(i')}=\hat{y}_{j(i)}=1$.

- לכל עיר i ומפעל j נגדיר

$$x'_{ij}=\begin{cases}\frac{x^\star_{ij}}{\sum_{j\in N_i}x^\star_{ij}} & j\in N_i\\ 0 & j\notin N_i\end{cases}.$$

נשים לב ש-$\sum_{j=1}^m x'_{ij}=1$, ומטענה 8 מתקיים $x'_{ij}\leq 2x^\star_{ij}$.

- לכל מפעל j נגדיר $y'_j=\min\left\{1,2y^\star_j\right\}$, ונקבל $y'_j\geq x'_{ij}$ (מכיוון ש-$x^\star_{ij}\leq y^\star_{ij}$).
- בסך הכל, קיבלנו ש-

$$D\left(x'\right)+F\left(y'\right)\leq 2\left(D\left(x^\star\right)+F\left(y^\star\right)\right).$$

- כעת נעגל את הפתרון - נסמן את הפתרון לאחר עיגול ב-$\left(\hat{x},\hat{y}\right)$.
- בהתחלה אף מפעל לא פתוח, ואף עיר לא משויכת למפעל.
- נתבונן באלגוריתם 13 לעיגול ההשמה.

טענה 9. לכל עיר i' כך ש-$N_i\cap N_{i'}\neq\emptyset$ מתקיים כי $d_{i',j(i)}\leq 6D^\star_{i'}$.

הוכחה. (טענה 9) נסמן ב-k-מפעל כך ש-$k\in N_i\cap N_{i'}$. מאי-שוויון המשולש, יחד עם העובדה ש-$D^\star_i\leq D^\star_{i'}$, נקבל

$$d_{i',j(i)}\leq d_{i',k}+d_{k,i}+d_{i,j(i)}\leq 2D^\star_{i'}+2D^\star_i+2D^\star_i=6D^\star_{i'}.$$

□

מסקנה 2. $D\left(\hat{x}\right)\leq 6D\left(x^\star\right)$.

- נותר לחסום את $F\left(\hat{y}\right)$.
- נסתכל על הערים שנבחרו בעת פעולת האלגוריתם, $i_1,\ldots,i_t$.

- לכל עיר i_r מתקיים

$$\forall j' \in N_{i_r} : f_{j(i_r)} \le f_{j'}.$$

מכיוון ש-$\hat{y}_{j(i_r)} = 1$ ו-$\hat{y}_j = 0$ לכל $j \in N_{i_r} \setminus \{j(i_r)\}$, נקבל

$$\begin{aligned}\sum_{j \in N_{i_r}} f_j \hat{y}_j &= f_{j(i_r)} \\ &= f_{j(i_r)} \cdot \sum_{j \in N_{i_r}} x'_{i_r,j} \\ &\le \sum_{j \in N_{i_r}} f_j \cdot y'_j \\ &\le 2 \sum_{j \in N_{i_r}} f_j y^\star_j.\end{aligned}$$

$$\implies F(\hat{y}) \le 2F(y^\star).$$

- בסך הכל קיבלנו קירוב 6 (בתרגיל הבית - להוריד ל-4).

2.8 בעיות ספירה

- עד כה דנו בעיקר על בעיות אופטימיזציה.
  - סוג נוסף של בעיות שנרצה להתעניין בו הוא בעיות ספירה.
  - במקום לקבוע אם פתרון קיים, או לחפש פתרון עם ערך אופטימלי, נרצה לספור את מספר הפתרונות לבעיה כלשהי.
- לחלק מבעיות הספירה קיים אלגוריתם פולינומי (כמו ספירת מספר העצים הפורשים).
- נרצה לתכנן אלגוריתם הסתברותי כך שלכל קלט x ופרמטרים ε, δ יחשב ערך y כך שבהסתברות לפחות $1-\delta$ מתקיים

$$(1-\varepsilon)\#x \le y \le (1+\varepsilon)\#x,$$

כאשר $\#x$ הוא מספר הפתרונות של x.

2.8.1 השמות מספקות של נוסחת DNF

קלט: נוסחת DNF φ מעל n משתנים.
פלט: קירוב של מספר ההשמות המספקות של φ.

- ספירה מקורבת של מספר ההשמות המספקות של נוסחת DNF.
  - נוסחת DNF φ היא מהצורה

$$\varphi = \bigvee_{j=1}^{m} T_i,$$

כאשר כל T_i היא $\bigwedge$ של ליטרלים.

- בהשוואה ל-CNF, קל למצוא השמה מספקת α: נבחר T_i כלשהו, ו-$\alpha(x_i)=1$ אם x_i מופיע בעצמו, ו-0 אם בשלילתו.
- עם זאת, ספירה מדויקת היא קשה: נניח שיש לנו אלגוריתם פולינומי שסופר את מספר ההשמות המספקות - נראה שניתן לפתור את SAT: כלומר להכריע אם ל-CNF קיימת השמה מספקת:

$$\varphi=\bigwedge_{j=1}^{m}c_j, \qquad \varphi'=\neg\varphi=\bigvee_{j=1}^{m}\neg c_j=\bigvee_{j=1}^{m}T_j$$

כך שאם $c_j=\bigvee_i z_i$ אז $T_j=\bigvee_i \neg z_i$.

  - φ' היא נוסחת DNF.
  - מכיוון ש-$\varphi'(x)=\neg\varphi(x)$, אם φ לא ספיקה, כלומר $\varphi(x)=0$ לכל x, אז $\forall x:\varphi'(x)=1$ - כלומר יש 2^n השמות מספקות.
  - אחרת, קיימות פחות מ-2^n השמות מספקות ל-φ'.
  - לכן, אם יכלנו לספור במדויק את מספר ההשמות ל-DNF, אז היינו פותרים SAT בזמן פולינומי.

- נסמן ב-$\#\varphi$ את מספר ההשמות המספקות של φ.
- נסמן ב-$P_{\text{SAT}}(\varphi)$ את ההסתברות שהשמה מקרית אחידה מספקת את φ:

$$P_{\text{SAT}}(\varphi)=\frac{\#\varphi}{2^n}\implies\#\varphi=2^n\cdot P_{\text{SAT}}(\varphi)\,.$$

  - נניח שנבחר S השמות מקריות באופן אחיד וב"ת.
  - נסמן ב-P' מ"מ שמייצג את אחוז ההשמות במדגם שמספקות את φ. אזי,

$$\mathbb{E}[P']=P_{\text{SAT}}(\varphi)\,.$$

שאלה: כמה גדול צריך להיות S כדי שבהסתברות גבוהה מתקיים

$$(1-\varepsilon)P_{\text{SAT}}(\varphi)\le P'<(1+\varepsilon)P_{\text{SAT}}(\varphi)$$

תשובה: נשתמש בצ'רנוף! נרצה להגדיר מ"מ אינדיקטורים $\chi_1,\ldots,\chi_S$ כך ש-

$$\forall i\in[S]:\Pr[\chi_i=1]=P,$$

ו-$\gamma\in(0,1)$, ואז נקבל:

$$\Pr\left[\frac{1}{S}\sum_{i=1}^{S}\chi_i>(1+\gamma)P\right]\le e^{-\frac{\gamma^2PS}{3}}$$

$$\Pr\left[\frac{1}{S}\sum_{i=1}^{S}\chi_i<(1-\gamma)P\right]\le e^{-\frac{\gamma^2PS}{3}}$$

לכל i נגדיר $\chi_i = 1$ אם ההשמה ה-i מספקת את φ. אזי,

$$P = P_{\text{SAT}}(\varphi)$$

$$P' = \frac{1}{S}\sum_{i=1}^{S}\chi_i \implies \mathbb{E}[P'] = P = P_{\text{SAT}}(\varphi)$$

נשתמש במשפט צ'רנוף עם $\gamma = \varepsilon$, ונקבל שאם

$$S = \frac{3}{P_{\text{SAT}}(\varphi)\cdot\varepsilon^2}\cdot\log\left(\frac{2}{\delta}\right)$$

נקבל הצלחה בהסתברות לפחות $1-\delta$.

הערה 8. שתי בעיות:

1. $P_{\text{SAT}}(\varphi)$ אינו ידוע.
2. $P_{\text{SAT}}(\varphi)$ יכול להיות אקספוננציאלי קטן.

ניעזר באלגוריתם 14 לבעיית גודל האיחוד:

- בהינתן DNF $\varphi = \bigvee_{j=1}^{m} T_j$, נגדיר $\mathcal{U} = \{0,1\}^n$ ו-$S_j \subseteq \mathcal{U}$ אוסף ההשמות המספקות בהן T_j מסתפקת.
- מספר ההשמות המספקות של φ הוא בדיוק $\left|\bigcup_{j=1}^{m} S_j\right|$.
- נוודא שההנחות מתקיימות:
1. אם T_j מכיל k ליטרלים, אזי $|S_j| = 2^{n-k}$: כל השאר חופשיים.
2. קל לבחור השמה באופן אחיד שמספקת את T_j: קביעת ערכי המשתנים שנמצאים ב-T_j, והשמה מקרית לשאר.
3. לכל השמה, ניתן לבדוק ביעילות אם היא מספקת את T_j (פשוט לבדוק שכל הליטרלים שמופיעים ב-T_j מסתפקים).

2.8.2 בעיית גודל האיחוד

קלט: עולם $\mathcal{U}$ ואוסף תתי-קבוצות $\{S_i\}_{i=1}^{m}$.

פלט: קירוב של $\left|\bigcup_{i=1}^{m} S_i\right|$.

- ניתן לחשב במדויק באמצעות הכלה והדחה, אך זה אקספוננציאלי ב-m.
- האלגוריתם שנראה משתמש בהנחות הבאות לכל $j \in [m]$:
1. ניתן לחשב את $|S_j|$ ביעילות.
2. ניתן לבחור איבר מקרי אחיד $u \in S_j$ ביעילות.

Algorithm 14: $\texttt{UnionSize}\left(\mathcal{U},\{S_i\}\right)$

עבור $i=1,\ldots,S$:
דגום קבוצה S_t בהסתברות $\left|S_t\right| / \sum_{j=1}^{n}\left|S_j\right|$.
דגום באופן אחיד $u\in S_t$.
מצא את ה-$j\in[m]$ המינימלי כך ש-$u\in S_j$.
אם $j=t$ הגדר $\chi_i\leftarrow 1$, ואחרת 0.
$\chi\leftarrow\sum_i\chi_i$.
החזר $F\leftarrow\frac{\chi}{S}\cdot\sum_{j=1}^{m}\left|S_j\right|$.

3. בהינתן $u\in\mathcal{U}$, ניתן לבדוק האם $u\in S_j$ ביעילות.

נתבונן באלגוריתם 14. בדיקת שפיות - נפריד למקרים:

1. כל הקבוצות S_j זרות. במקרה זה $\left|\bigcup_{j=1}^{m}S_j\right|=\sum_{j=1}^{m}\left|S_j\right|$, ואז לכל $i\in[S]$ מתקיים $\chi_i=1$:

$$\chi=|S|\implies F=\frac{S}{S}\sum_{j=1}^{m}\left|S_j\right|=\left|\bigcup_{j=1}^{m}S_j\right|.$$

2. כל הקבוצות S_j זהות. במקרה זה $\left|\bigcup_{j=1}^{m}S_j\right|=\frac{1}{m}\sum_{j=1}^{m}\left|S_j\right|$, ולכן $\chi_i=1$ רק אם $t=1$ - שקורה בהסתברות $1/m$. לכן, נצפה ש-

$$\frac{\chi}{S}\approx\frac{1}{m}.$$

למה 10. נגדיר $\eta=\frac{\left|\bigcup_{j=1}^{m}S_j\right|}{\sum_{j=1}^{m}\left|S_j\right|}$. אז לכל $i\in[S]$,

$$\Pr\left[\chi_i=1\right]=\eta.$$

משפט 15. אם

$$S\geq\frac{3m}{\varepsilon^2}\ln\left(\frac{2}{\delta}\right),$$

אז בהסתברות לפחות $1-\delta$ מתקיים

$$(1-\varepsilon)\left|\bigcup_{j=1}^{m}S_j\right|\leq F\leq(1+\varepsilon)\left|\bigcup_{j=1}^{m}S_j\right|$$

הוכחה. (משפט 15) האבחנה החשובה היא ש-

$$\eta=\frac{\left|\bigcup_{j=1}^{m}S_j\right|}{\sum_{j=1}^{m}\left|S_j\right|}\geq\frac{\max_j\left|S_j\right|}{m\cdot\max_j\left|S_j\right|}\geq\frac{1}{m}.$$

לכן $S \geq \frac{3}{\eta\varepsilon^2}\ln\left(\frac{2}{\delta}\right)$. נשתמש בצ'רנוף עם $P=\eta,\gamma=\varepsilon$ ונקבל ש-

$$\Pr\left[\frac{\chi}{S} \geq (1+\varepsilon)\,\eta\right] \leq e^{-\frac{\varepsilon^2\eta S}{3}} \leq \frac{\delta}{2}$$

$$\Pr\left[\frac{\chi}{S} < (1-\varepsilon)\,\eta\right] \leq \frac{\delta}{2}$$

נשתמש בחסם איחוד, ונקבל שבהסתברות לפחות $1-\delta$ מתקיים

$$(1-\varepsilon)\,\eta \leq \frac{\chi}{S} \leq (1+\varepsilon)\,\eta$$

$$\implies (1-\varepsilon)\left|\bigcup_{j=1}^{m} S_j\right| \leq F \leq (1+\varepsilon)\bigcup_{j=1}^{m} S_j.$$

□

כעת נותר להוכיח את למה 10.

הוכחה. (למה 10)

- נסתכל על כל הזוגות (t,u) כך ש-$t\in[m]$ ו-$u\in S_t$. מספר הזוגות הוא $\sum_{j=1}^{m}|S_j|$.
- מה ההסתברות לבחור זוג ספציפי (t,u)?

$$\begin{aligned}\Pr\left[(t,u)\ \text{לבחור}\right] &= \Pr\left[t\ \text{לבחור}\right]\cdot\Pr\left[u\ \text{לבחור}\mid t\ \text{בחרנו}\right]\\ &= \frac{|S_t|}{\sum_{j=1}^{m}|S_j|}\cdot\frac{1}{|S_t|}\\ &= \frac{1}{\sum_{j=1}^{m}|S_j|}.\end{aligned}$$

מכאן, כל זוג נבחר באותה ההסתברות.

- לכל $u\in\mathcal{U}$, נסמן ב-$j(u)$ את האינדקס המינימלי כך ש-$u\in S_j$.
- לפי הגדרת האלגוריתם, $\chi_i=1 \iff t=j(u)$.
- מכיוון ש-$j(u)$ הוא ייחודי, מספר הזוגות (t,u) כך ש-$t=j(u)$ הוא $\left|\bigcup_{j=1}^{m}S_j\right|$.
- בסך הכל נקבל:

$$\Pr\left[\chi_i=1\right] = \frac{\left|\bigcup_{j=1}^{m}S_j\right|}{\sum_{j=1}^{m}|S_j|} = \eta.$$

□

2.9 בדיקת תכונות מדגמית

2.9.1 מבוא

- עד כה דיברנו על זמן פולינומי בגודל הקלט כ-"יעיל".
- עם זאת, בעידן ה-Big Data אפילו זמן לינארי לא בא בחשבון.
- אלגוריתמים כאלה לא יכולים בכלל לקרוא את כל הקלט אבל מסוגלים "לדגום לתוכו".
  - בדרך כלל רנדומיים, שמספקים תשובה מקורבת.
- נדבר בעיקר על תת-תחום שנקרא בדיקות תכונות מדגמית (property testing).

הגדרה 14. אלגוריתם לבדיקת תכונות לתכונה P מקבל פרמטר מרחק ε יחד עם גישת שאילתא לקלט O.

1. אם O הוא בתכונה P, האלגוריתם יחזיר <u>כן</u> בהסתברות $2/3$.
2. אם O הוא ε-רחוק מ-P, האלגוריתם יחזיר <u>לא</u> בהסתברות $2/3$.

נרצה לבצע כמה שפחות שאילתות לקלט.

הערה 9. 1. ניתן בקלות לקבל הסתברות $1-\delta$ (ע"י חזרות).

2. אם ההסתברות במקרה הראשון היא 1, אז האלגוריתם בעל <u>שגיאה חד-צדדית</u>.
3. במידה והאובייקט ε-קרוב, אין שום הבטחה על הפלט.

דוגמה 20. בדיקת אפסיות של מחרוזות.

האובייקט: מחרוזות מעל $\{0,1\}^n$.

התכונה: המחרוזת היא "הכל אפסים".

- אלגוריתם מדויק לבעיה זו עובד בזמן $\Theta(n)$.

המרחק: המינג מנורמל:

$$d_H(x,y)=\frac{1}{n}\left|\{i\in[n]\mid x_i\neq y_i\}\right|.$$

שאילתא: בהינתן $i\in[n]$, החזר את $w_i\in\{0,1\}$.

- הצעה לאלגוריתם:
  - דגום באופן אחיד וב"ת S אינדקסים $i_1,\ldots,i_S$, ודגום את w בהם.
  - אם כולם 0 קבל, ואחר דחה.

ננתח את האלגוריתם:

- אם w היא "הכל אפסים" האלגוריתם מקבל בהסתברות 1 (כלומר השגיאה חד-צדדית).

- אם w היא ε-רחוקה מ-"הכל אפסים", אזי המרחק שלה הוא לפחות ε, ולכן קיימים לפחות εn אינדקסים i כך ש-$w_i = 1$.
- נחשב את ההסתברות שהאלגוריתם מקבל:

$$\Pr\left[\text{קבלה}\right] \leq (1-\varepsilon)^S \leq e^{-\varepsilon S} < 1/3,$$

עבור $S > 2/\varepsilon$.

דוגמה 21. בדומה לדוגמא 20, האובייקט הוא מחרוזת, המרחק הוא $d_H(\cdot,\cdot)$ והשאילתות מהצורה $i \to w_i$. התכונה היא "מחרוזת חזקה" (מכילה לפחות $n/2$ 1-ים).

- הצעה לאלגוריתם:
  - דגום באופן אחיד וב"ת S אינדקסים $i_1, \ldots, i_S$, ודגום את w בהם.
  - אם לפחות $S\left(1/2 - \varepsilon/2\right)$ ביטים הם 1 קבל, ואחרת דחה.

ננתח את האלגוריתם - נשתמש בצ'רנוף חיבורי:

- נסמן ב-$\alpha(w)$ את החלק היחסי של ה-1 במחרוזת w.
- אם w היא מחרוזת חזקה, אזי $\alpha(w) \geq 1/2$.
- אם w היא $\varepsilon/2$-רחוקה מהתכונה, אז $\alpha(w) < 1/2 - \varepsilon/2$.
- נסמן ב-$i_1, \ldots, i_S$ את האינדקסים שנבחרו.
- נגדיר מ"מ $\chi_1, \ldots, \chi_S$ כך ש-$\chi_j = 1$ אם $w_{i_j} = 1$. אזי, לכל j מתקיים

$$\Pr\left[\chi_j = 1\right] = \alpha(w).$$

  - אם w מחרוזת חזקה, אזי עבור $S \geq 4/\varepsilon^2$ נקבל

$$\begin{aligned}\Pr\left[\text{נדחית } w\right] &= \Pr\left[\frac{1}{S}\sum_{j=1}^{S}\chi_j < \frac{1}{2} - \frac{\varepsilon}{2}\right] \\ &\leq \Pr\left[\frac{1}{S}\sum_{j=1}^{S}\chi_j < \alpha(w) - \frac{\varepsilon}{2}\right] \\ &\leq e^{-2\left(\frac{\varepsilon}{2}\right)^2 S} \\ &\leq \frac{1}{3}.\end{aligned}$$

  - אם w ε-רחוקה מחזקה:

$$\begin{aligned}\Pr\left[\text{מתקבלת } w\right] &= \Pr\left[\frac{1}{S}\sum_{j=1}^{S}\chi_j \geq \frac{1}{2} - \frac{\varepsilon}{2}\right] \\ &\leq \Pr\left[\frac{1}{s}\sum_{j=1}^{S}\chi_j \geq \alpha(w) - \frac{\varepsilon}{2}\right] \\ &\leq e^{-2S\frac{\varepsilon^2}{4}} \\ &< 1/3.\end{aligned}$$

תהייה: למי אכפת?

1. להשתמש באלגוריתם כעיבוד מקדים (preprocessing).
2. הדאטה עצום ואי אפשר להרשות זמן לינארי.
3. הבעיה קשה (כמו צביעה) אך ניתן להסתפק בקירוב כלשהו.

סוגי בעיות שנחקרו:

1. תכונות של גרפים (למשל דו-צדדיות, צביעה).
2. תכונות אלגבריות (של קודים).
3. תכונות של מחרוזות.
4. תכונות של פונקציות בוליאניות.
5. תכונות של התפלגויות.

2.9.2 בדיקת מונוטוניות

אובייקט: פונקציה $f:[n]\to R$ עבור אוסף R עם סדר מלא.

תכונה: הפונקציה f מונוטונית: $\forall i<j: f(i)\leq f(j)$.

דגימה: בהינתן $i\in[n]$, פלוט $f(i)$.

מרחק: המינג מנורמל.

- קיימים אלגוריתמים שרצים בזמן $\mathcal{O}(\log n/\varepsilon)$.
- נתחיל באלגוריתם נאיבי:
  - דגום באופן אחיד אינדקסים ב-$[n]$.
  - נדחה אמ"מ קיימת הפרה ($i<j$ כך ש-$f(i)>f(j)$).
- קיים חסם תחתון של $\Omega(\sqrt{n})$ לאלגוריתם זה (עבור ε קבוע).
- נסתכל על הפונקציה הבאה:

$$f(i)=\begin{cases} i+1 & i\equiv 1 \mod 2\\ i-1 & i\equiv 0 \mod 2\end{cases}$$

כלומר $f(1)=2, f(2)=1, f(3)=4, f(4)=3,\ldots$.

טענה 10. הפונקציה f היא $1/2$-רחוקה ממונוטונית.

טענה 11. ההסתברות שדגימה אחידה ב"ת של $S\leq\sqrt{n}/2$ אינדקסים מכילה זוג משודך $(i,i+1)$ היא לכל היותר $2/3$.

Algorithm 15: MonotoneTest (f)

1. דגום $s=2/\varepsilon$ אינדקסים $i_1,\ldots,i_s$ מ-$[n]$ באופן אחיד ב"ת.
2. לכל אינדקס i_r נסמן $x_r=f\left(i_r\right)$, ובצע חיפוש בינארי של x_r.
3. אם קיים x_r שעבורו החיפוש נכשל דחה, ואחרת קבל.

הוכחה. (טענה 10) תהי f' פונקציה מונוטונית.

- האלגוריתם הנאיבי דוחה רק אם הוא מוצא זוג משודך.
- הפונקציה f' חייבת לא להסכים עם f בלפחות איבר אחד מכל זוג משודך (אחרת עבור $(i,i+1)$ כלשהם נקבל $f'\left(i\right)>f'\left(i+1\right)$ - סתירה).
- מכאן, f ו-f' לא מסכימות על לפחות $n/2$ אלמנטים, והמרחק הוא לפחות $1/2$.

□

הוכחה. (טענה 11) נסתכל על האינדקסים שנבחרו ב-$i_1,\ldots,i_S$.

- לכל $1\leq j<k\leq S$, נסמן ב-$E_{j,k}$ את המאורע ש-i_j ו-i_k הם זוג משודך.
  - בפרט, זה אומר ש-i_j אי-זוגי ו-$i_k=i_j+1$, או ש-i_j זוגי ו-$i_k=i_j-1$.

$$\Pr\left[E_{j,k}\right]=\frac{1}{n}\implies\Pr\left[\bigcup_{j<k}E_{j,k}\right]\leq\sum_{j<k}\Pr\left[E_{j,k}\right]\leq\frac{S^2}{2}\cdot\frac{1}{n}\leq\frac{1}{4}.$$

□

- כעת נתבונן באלגוריתם 15 שרץ בזמן $\Theta\left(\log n/\varepsilon\right)$.
- נניח בה"כ כי f חח"ע (אחרת היינו עובדים עם פונקציה שפולטת $\left(f\left(i\right),i\right)$- סדר משני לפי האינדקסים):
$$\forall i\neq j:f\left(i\right)\neq f\left(j\right).$$
- האלגוריתם מבצע חיפוש בינארי על הערכים של f (נחשוב על f בתור מערך).
- ברור כי זמן הריצה של האלגוריתם הוא $\Theta\left(\log n/\varepsilon\right)$.

דוגמה 22. נסתכל על המערך (שמייצג פונקציה) הבא:

index	1	2	3	4	5	6	7	8	9
value	6	4	2	5	8	0	12	14	10

- נניח כי דגמנו את $i=9$, ונחפש את $f\left(9\right)=10$.
- תחילה נחפש באינדקס 5, נזוז ימינה ל-$i=7$, שמאלה ל-$i=6$, והחיפוש ייכשל.

אבחנה: אם f היא מונוטונית, האלגוריתם מקבל בהסתברות 1 (כי המערך ממויין והחיפוש הבינארי יעבוד תמיד).

נותר להראות שאם f היא ε-רחוקה, אזי האלגוריתם דוחה בהסתברות לפחות $2/3$.

הגדרה 15. אינדקס $j \in [n]$ הוא **עד** לאי-מונוטוניות של f אם חיפוש בינארי על $x = f(j)$ לא מחזיר את j (נכשל).

למה 11. אם f היא ε-רחוקה ממונוטוניות, אז קיימים יותר מ-εn עדים.

הוכחה. (למה 11) נניח בשלילה שיש לכל היותר εn עדים עבור f.

שלב 1: נסתכל על כל האינדקסים שהם לא עדים, ונראה שהם יוצרים תת-מערך ממויין.

- לכל זוג של לא-עדים j, j' כאשר $j < j'$, נסתכל השלבים בחיפוש הבינארי של $x_j = f(j)$ ו-$x_{j'} = f(j')$.
- נסמן ב-u את האינדקס הראשון בו החיפוש הבינארי מתפצל.
  - אזי $j \leq u \leq j'$ ואחד מאי-השוויונים לפחות חזק.
  - מתקיים ש-$x_j \leq x_u \leq x_{j'}$ כאשר אחד מאי-השוויונים לפחות חזק.
  - לכן $f(j) < f(j')$ כנדרש.

שלב 2: נשנה את f בכל אינדקס j שהוא עד, שינוי של לכל היותר εn ערכים.

ניתן להפוך את המערך לממוין ב-$\leq \varepsilon n$ שינויים, בסתירה לכך ש-f היא ε-רחוקה.

□

מסקנה 3. אם f היא ε-רחוקה ממונוטונית, אלגוריתם 15 דוחה בהסתברות לפחות $2/3$.

הוכחה. (מסקנה 3) אם f היא ε-רחוקה, מלמה 11 יש יותר מ-εn עדים. מכאן, ההסתברות שהאלגוריתם לא יתפוס עד בכל s האינדקסים שהוגרלו היא לכל היותר

$$(1-\varepsilon)^s \leq e^{-\varepsilon s} = e^{-2} < 1/3.$$

□

2.9.3 בדיקת תכונות בגרפים

2.9.3.1 מודלים

ישנם שלושה מודלים מרכזיים לבדיקת תכונות בגרפים:

1. המודל הצפוף:
   - א) הקלט הוא גרף (פשוט) שמיוצג ע"י מטריצה $n \times n$.
   - ב) השאילתות הן: האם יש קשת בין (i, j).
   - ג) המרחק הוא המינג מנורמל (אם קלט הוא ε-רחוק מהתכונה אז צריך לשנות לפחות εn^2 ערכים במטריצה).

2. גרפים חסומי דרגה: נתון חסום עליון על הדרגה המקסימלית d.
   א) הקלט הוא מטריצה $n \times d$.
   ב) השאילתות הן:
      i. בהינתן $v \in V$, החזר $\deg\left(v\right)$.
      ii. בהינתן $v \in V$ ואינדקס j, החזר את השכן ה-j של v (או שאין כזה).
   ג) המרחק הוא המינג מנורמנל (אם הגרף ε-רחוק, יש לשנות לפחות εnd).
3. המודל הכללי: נתון חסם עליון על מספר הקשתות בגרף.
   א) הקלט הוא רשימת שכנויות.
   ב) שאילתות ומרחק בדומה לגרפים חסומי דרגה.
   ג) הגרף הוא ε-רחוק אם צריך לשנות יותר מ-εm ערכים ברשימה.

2.9.3.2 בדיקת קשירות

הגדרה 16. גרף הוא קשיר אם קיים מסלול בין כל זוג קודקודים.

- אם גרף לא קשיר, אז הוא מתפרק לרכיבי קשירות בגודל מקסימלי.
- נניח שבגרף יש $C\left(G\right)$ רכיבי קשירות.

אבחנה: ניתן להפוך את הגרף לקשיר ע"י הוספת $C\left(G\right)-1$ קשתות.

למה 12. אם גרף הוא ε-רחוק מקשיר, אז קיימים $\varepsilon m+1$ רכיבי קשירות.

הוכחה. (למה 12) נניח בשלילה שיש פחות מ-$\varepsilon m+1$ רכיבי קשירות. אזי ניתן להפוך את הגרף ע"י εm שינויים - סתירה. □

הערה 10. ניתן להכליל את למה 12 לגרפים חסומי דרגה, אך בזהירות (לא ניתן להוסיף קשתות לקודקודים מדרגה מקסימלית).

שים לב: אם $\varepsilon m \geq n$, אז כל גרף הוא ε-קרוב: תמיד ניתן להוסיף $n-1 \leq$ קשתות כדי להפוך את הגרף לקשיר (אז ניתן לקבל תמיד). נתמקד במקרה בו $\varepsilon < n/m$.

הערה 11. עבור גרפים צפופים נקבל $\varepsilon < 1/n$. לכן, אם האלגוריתם פולינומיאלי (או אפילו לינארי) ב-$1/\varepsilon$, הוא לא יהיה תת-לינארי. עם זאת, רוב הגרפים הצפופים הם קשירים.

למה 13. אם הגרף הוא ε-רחוק מקשיר, אז הוא מכיל לפחות $\varepsilon\bar{d}n/2$ רכיבי קשירות בגודל קטן מ-$\frac{2}{\varepsilon\bar{d}}$, כאשר

$$\bar{d}=\frac{m}{n}.$$

הוכחה. מלמה 12, אם הגרף הוא ε-רחוק, אז הוא מכיל לפחות $\varepsilon\bar{d}n$ רכיבי קשירות.

- נגיד שרכיב קשירות הוא קטן אם הוא מכיל לכל היותר $\frac{2}{\varepsilon\bar{d}}$ קודקודים, ואחרת גדול.
- נניח בשלילה שיש פחות מ-$\varepsilon n\bar{d}/2$ רכיבי קשירות בגודל קטן מ-$\frac{2}{\varepsilon\bar{d}}$.

Algorithm 16: ConnectivityTest (f)

1 דגום $s = 4/\left(\varepsilon\bar{d}\right)$ קודקודים באופן אחיד ב"ת.
2 לכל קודקוד s בצע BFS המתחיל מ-s, עד שהתגלו $2/\left(\varepsilon\bar{d}\right)$ קודקודים או שלא ניתן לקבל עוד קודקוד (רכיב קשירות קטן).
3 אם אחת מהרצות ה-BFS מצאה רכיב קשירות קטן, דחה, ואחרת קבל.

- מכיוון שכל רכיב קשירות גדול מכיל לפחות $\frac{2}{\varepsilon\bar{d}}$ קודקודים, ורכיבי הקשירות זרים, יש לכל היותר
$$\frac{n}{\frac{2}{\varepsilon\bar{d}}} = \frac{\varepsilon\bar{d}n}{2}$$
רכיבי קשירות גדולים.
- מכאן, נקבל שיש פחות מ-
$$\underbrace{\frac{\varepsilon\bar{d}n}{2}}_{\text{קטנים}} + \underbrace{\frac{\varepsilon\bar{d}n}{2}}_{\text{גדולים}} = \varepsilon\bar{d}n$$
רכיבי קשירות - סתירה.

□

נתבונן באלגוריתם 16 הנובע מהלמה לעיל.

- אם הגרף קשיר, כל הרצות ה-BFS לא ימצאו רכיב קשירות קטן, ולכן האלגוריתם יקבל בהסתברות 1.
- מאידך, אם הגרף ε-רחוק:
  - מלמה 13 קיימים לפחות $\varepsilon n\bar{d}/2$ קודקודים ששייכים לרכיבי קשירות קטנים.
  - אם האלגוריתם תופס קודקוד כזה, אז הוא דוחה. ההסתברות שלא נצליח לעשות זאת היא
$$\left(1 - \frac{\varepsilon\bar{d}}{2}\right)^{4/\varepsilon\bar{d}} \leq e^{-2} < 1/3.$$

ננתח את זמן הריצה של האלגוריתם (או סיבוכיות שאילתות):

- אם הדרגה המקסימלית היא $\mathcal{O}\left(\bar{d}\right)$, נקבל
$$\mathcal{O}\left(\underbrace{\frac{1}{\varepsilon\bar{d}}}_{\text{דגימה}} \cdot \underbrace{\frac{1}{\varepsilon\bar{d}} \cdot \bar{d}}_{\text{BFS}}\right) = \mathcal{O}\left(\frac{1}{\varepsilon^2\bar{d}}\right).$$

- אחרת, נקבל

$$\mathcal{O}\left(\underbrace{\frac{1}{\varepsilon\bar{d}}}_{\text{דגימה}}\cdot\underbrace{\left(\frac{1}{\varepsilon\bar{d}}\right)^2}_{\text{BFS}}\right)=\mathcal{O}\left(\frac{1}{\left(\varepsilon\bar{d}\right)^3}\right).$$

שיפור: בניתוח התייחסנו לשני המקרים הבאים, למרות שהם לא יכולים לקרות יחד:

- כל הקודקודים ברכיבי קשירות קטנים. במקרה זה, בהסתברות 1 נופלים על רכיב קטן ומשלמים $\frac{2}{\varepsilon\bar{d}}$.
- כל רכיב קשירות קטן הוא בגודל 1, אבל אז לא צריך BFS.

בתרגיל: אלגוריתם משופר עם זמן ריצה טוב יותר - חלוקה לדליים.

2.10 למידה חישובית: Probably Approximately Correct

יש הרבה מודלים עבור למידה חישובית. מודל פורמלי מפורסם הוא מודל PAC:

Probably Approximately Correct.

2.10.1 הגדרות

הרעיון הכללי: להכליל מדוגמאות מתוייגות.

מוטיבציה: ניתן לחשוב על כלל סיווג עבור אנשים חולים.

- נניח שקיבלנו מדגם של אנשים שלחלק יש מחלה (בה"כ קורונה) ולחלק אין.
- כל פציינט מיוצג ע"י וקטור של תכונות (למשל גיל, משקל, לחץ דם וכו'), יחד עם האם הוא חולה או לא.
- בהינתן מדגם הפציינטים, היינו רוצים ללמוד כלל שיאפשר לתייג חולים פוטנציאליים חדשים.

באופן פורמלי:

- ניתן לחשוב על הדוגמאות כאילו הן מגיעות מאיזשהו תחום X (למשל $X=\{0,1\}^n$ או $X=\mathbb{R}^n$).
- בנוסף, קיימת התפלגות כלשהי D (לא ידועה לנו) מעל X.
- האלגוריתם מקבל דגימות לפי D (כלומר כל דגימה $x\sim D$), ומגיעה יחד עם תיוג $f\left(x\right)$ (פונקציית מטרה).
- אנו מניחים ש-f מגיעה ממשפחה מסוימת של פונקציות $\mathcal{F}$.
  - למען הפשטות, נניח ש-f מגיעה ממחלקה $\mathcal{F}$ של פונקציות בוליאניות:

$$\forall f\in\mathcal{F}:f:\{0,1\}^n\to\{0,1\}$$

Algorithm 17: $\texttt{SingletonLearning}\,(n,\varepsilon,\delta)$

נבחר מדגם (מתויג) $\left(\left(x^1,f\left(x^1\right),,\ldots,\left(x^s,f\left(x^s\right)\right)\right)\right)$ עבור s כמוגדר בהמשך.
נפלוט $h\in\mathcal{F}$ שעקבית עם המדגם, כלומר $\forall i\in[s]:h\left(x^i\right)=f\left(x^i\right)$.

- ניתן להכליל את התוצאות בהמשך למחלקות אחרות.

סימונים:

- תחום הדגימות הוא X.
- התפלגות D של הדגימות (קבועה אך לא ידועה).
- $\mathcal{F}$ היא משפחה של פונקציות בוליאניות מעל X.
- $f\in\mathcal{F}$ נקראת פונקציית המטרה.

הגדרה 17. אלגוריתם PAC מקבל גישה לדוגמאות מתויגות $\left(x,f\left(x\right)\right)$ כאשר $x\sim D$, ובנוסף מקבל שני פרמטרים $\varepsilon,\delta\in(0,1)$. האלגוריתם פולט היפותזה $h:X\to\{0,1\}$ כך שהבאים מתקיימים:

1. בהסתברות $1-\delta$ לפחות (מעל הדוגמאות, ואולי מעל רנדומיות פנימית של האלג'):

$$err_{f,D}\left(u\right)\leq\varepsilon;\quad err_{f,D}\left(h\right):=\Pr_{x\sim D}\left[h\left(x\right)\neq f\left(x\right)\right].$$

2. זמן הריצה וסיבוכיות המדגם יהיו פולינומיים ב-$1/\varepsilon$, $\log\left(1/\delta\right)$, מימד הדגימות n ובמידת הסיבוכיות של המחלקה $\mathcal{F}$.

הערה 12. ניתן להסתכל גם על מקרה שבו h נדרשת להיות ב-$\mathcal{F}$ (Proper Learning). אחרת, נניח ש-$f\in\mathcal{H}$ עבור $\mathcal{F}\supseteq\mathcal{H}$ כלשהי. בנוסף, יש הרבה וריאנטים של המודל, נראה כאן אחד מהם.

2.10.2 למידת סינגלטונים

דוגמא פשוטה מאוד:

$$X=\{0,1\}^n$$

$$\mathcal{F}=\left\{g_i:\{0,1\}^n\to\{0,1\}\mid g_i\left(x\right)=x_i\right\}$$

למשל, עבור $n=5$,

$$g_2\left(10111\right)=0,\quad g_3\left(10111\right)=1.$$

נתבונן באלגוריתם 17 לפתרון הבעיה.

למה 14. אם $s\geq\ln\left(n/\delta\right)/\varepsilon$, אז בהסתברות לפחות $1-\delta$ נקבל $err_{f,D}\left(h\right)\leq\varepsilon$.

Algorithm 18: $\texttt{OccamsRazor}\,(\varepsilon,\delta)$

1 דגום s דגימות המתפלגות לפי D ומתויגיות לפי f.
2 פלוט היפותזה $h\in\mathcal{H}$ שעקבית עם f על המדגם.

הוכחה. (למה 14) נסמן

$$B_{\varepsilon,D}\left(f\right):=\left\{g_i\in\mathcal{F}\mid err_{f,D}\left(g_i\right)>\varepsilon\right\},$$

ונראה שבהסתברות לפחות $1-\delta$ אין $g_i\in B_{\varepsilon,D}\left(f\right)$ שמסכימה עם f על המדגם. הנ"ל גורר שבהסתברות לפחות $1-\delta$ מתקיים $h\notin B_{\varepsilon,D}\left(f\right)$, ולכן $err_{f,D}\left(h\right)\leq\varepsilon$ כנדרש. לכל $g_i\in B_{\varepsilon,D}\left(f\right)$, נסמן ב-$E_i$ את המאורע ש-g_i קונסיסטנטית עם f על המדגם. אזי,

$$\forall g_i\in B_{\varepsilon,D}\left(f\right):\Pr\left[E_i\right]<\left(1-\varepsilon\right)^s\leq e^{-\varepsilon s}\leq\delta/n.$$

נפעיל חסם איחוד ונקבל:

$$\begin{aligned}\Pr\left[\exists g_i\in B_{\varepsilon,D}\left(f\right):f\text{ עקבית עם }\right]&=\Pr\left[\bigcup_{g_i\in B_{\varepsilon,D}(f)}E_i\right]\\&\leq\sum_{g_i\in B_{\varepsilon,D}(f)}\Pr\left[E_i\right]\\&\leq n\cdot\frac{\delta}{n}=\delta.\end{aligned}$$

□

סיבוכיות המדגם: היא $s=\mathcal{O}\left(\frac{\ln(n/\delta)}{\varepsilon}\right)$.

זמן הריצה: הוא $ns=\mathcal{O}\left(\ln\left(n/\delta\right)\frac{n}{\varepsilon}\right)$.

2.10.3 התער של אוקאם (Occam's Razor)

האלגוריתם של לממידת סינגלטונים הוא מקרה פרטי של פרדיגמת למידה כללית שעובדת לקבוצות סופיות של פונקציות.

נניח שרוצים ללמוד פונקציה ממחלקה סופית $\mathcal{F}$, ונניח שההיפותזה h נבחרת תמיד מתוך מחלקה (סופית) כאשר $\mathcal{H}\supseteq\mathcal{F}$ (מקרה פרטי הוא $\mathcal{H}=\mathcal{F}$).

משפט 16. אם $s\geq\ln\left(\left|\mathcal{H}\right|/\delta\right)/\varepsilon$, אז לכל $f\in\mathcal{F}$ והתפלגות D, בהסתברות לפחות $1-\delta$ אלגוריתם 18 פולט היפותזה h כך ש-$err_{f,D}\left(h\right)\leq\varepsilon$ (קיבלנו אלגוריתם PAC).

הוכחה. (משפט 16) באופן דומה להוכחת למה 14, נסמן

$$B_{\varepsilon,D,\mathcal{H}}\left(f\right):=\left\{g_i\in\mathcal{F}\mid err_{f,D}\left(g_i\right)>\varepsilon\right\},$$

ונראה שבהסתברות לפחות $1-\delta$ אין $g\in B_{\varepsilon,D,\mathcal{H}}\left(f\right)$ שמסכימה עם f על המדגם. לכל $g\in B_{\varepsilon,D,\mathcal{H}}\left(f\right)$, נסמן ב-$E_g$ את המאורע ש-E_g קונסיסטנטית עם f על המדגם. אזי,

$$\forall g\in B_{\varepsilon,D,\mathcal{H}}\left(f\right):\Pr\left[E_g\right]<\left(1-\varepsilon\right)^s\leq e^{-\varepsilon s}\leq\delta/\left|\mathcal{H}\right|.$$

נפעיל חסם איחוד ונקבל:

$$\begin{aligned}\Pr\left[\exists g\in B_{\varepsilon,D,\mathcal{H}}\left(f\right):f\text{ עקבית עם }g\right]&=\Pr\left[\bigcup_{g\in B_{\varepsilon,D,\mathcal{H}}(f)}E_g\right]\\&\leq\sum_{g\in B_{\varepsilon,D,\mathcal{H}}(f)}\Pr\left[E_g\right]\\&\leq\left|\mathcal{H}\right|\cdot\frac{\delta}{\left|\mathcal{H}\right|}=\delta.\end{aligned}$$

$\square$

מסקנה 4. ממשפט 18, כל עוד $\left|\mathcal{H}\right|\leq\exp\left(\mathrm{poly}\left(n\right)\right)$, מספר הדוגמאות שמספיק ללמידה הוא פולינומי ב-$1/\varepsilon$, $\log\left(1/\delta\right)$ וב-n כנדרש. בפרט, המשפט מאפשר להרחיב את התוצאה של למידת סינגלטונים למחלקות גדולות יותר (כמו מונומים - בתרגיל).

1. מכאן, הלמידה בעצם מסתכמת למציאת היפותזה שעקבית עם המדגם.
2. אם המחלקה לא כל-כך גדולה, בדי להכליל מספיק לקחת מדגם גדול דיה ולמצוא פונקציה שטובה עבור המדגם הספציפי.

קשיים ודרכי התמודדות

1. מציאת היפותזה עקבית יכולה להיות בעיה קשה (למשל k-DNF).
2. מחלקת ההיפותזות $\mathcal{H}$ יכולה להיות בגודל אינסופי (נראה בהמשך - VC dimension).
3. המידע יכול להיות רועש (ניתן להכליל עם התער של אוקאם).
4. ייתכן מצב שבו אנו לא יודעים דבר על שייכות של f למחלקה כלשהי (למידה אגנוסטית).

2.10.4 מחלקות אינסופיות של פונקציות

- נסתכל על דוגמא בא הנקודות מתפלגות על המישור $\mathbb{R}^2$, ו-f מוגדרת ע"י מלבן (מקביל לצירים), כך ש-$f\left(x,y\right)=1$ אמ"מ $\left(x,y\right)$ שייכת למלבן.
- פונקציה כזאת יכולה לתאר למשל אדם עם מבנה גוף בינוני (תכונות של משקל וגובה - נקודה במלבן אם המשקל והגובה לא נמוכים או גבוהים מדי).
- אלגוריתם ספציפי לבעיה זו הוא אלגוריתם שמוצא את המלבן הקטן ביותר שמכיל את כל הדגימות החיוביות ($f\left(x,y\right)=1$).
- ניתן לנתח אלגוריתם כזה ולהראות שמדגם בגודל $\mathcal{O}\left(\log\left(1/\delta\right)/\varepsilon\right)$ מספיק.
  - לחילופין, ניתן להשתמש בואריאציה כלשהי של אוקאם.

- כדי להכליל את התער של אוקאם למחלקות אינסופיות של פונקציות, ניתן להגדיר מידת סיבוכיות שנקראת VC-dimension (Vapnik-Chervonenkis).

הגדרה 18. (VC-dimension) עבור מחלקת פונקציות $\mathcal{F}$ (לאו דווקא סופית), אומרים שאוסף נקודות $x^1,\ldots,x^m$ מנותץ ע"י $\mathcal{F}$ אם עבור כל תיוג של $x^1,\ldots,x^m$ קיימת $f\in\mathcal{F}$ שמסכימה עם התיוג. מימד ה-VC של $\mathcal{F}$, $VC\left(\mathcal{F}\right)$, הוא הגודל המקסימלי שקיים עבורו סדרת נקודות $x^1,\ldots,x^m$ שמנותצת ע"י $\mathcal{F}$.

דוגמה 23. מחלקת פונקציות האינטרוולים $\mathcal{I}$ מעל $[0,1]$. כלומר, לכל $\tau_1\leq\tau_2$ נגדיר

$$f_{\tau_1,\tau_2}:[0,1]\rightarrow\{0,1\}$$

ב-$\mathcal{I}$ כך ש-$f_{\tau_1,\tau_2}\left(x\right)=1$ אמ"מ $x\in[\tau_1,\tau_2]$.

- מימד ה-VC של המחלקה הזו הוא 2:
  - כל שתי נקודות מנותצות ע"י $\mathcal{F}$.
  - עם זאת, כל שלוש נקודות לא מנותצות ע"י $\mathcal{F}$: עבור תיוג שנותן 0 לנקודה האמצעית ו-1 לשאר.
- ניתן להראות שלפונקציית מלבנים מימד ה-VC יהיה 4.

ניתן להכליל את משפט אוקאם למקרה בו $\mathcal{H}$ לא סופית עם סיבוכיות מדגם

$$\mathcal{O}\left(\left(VC\left(\mathcal{H}\right)+\frac{1}{\varepsilon}\right)\log\left(\frac{1}{\delta}\right)\right).$$

2.10.5 למידה אגנוסטית (Agnostic Learning)

- וריאנט של PAC בו שום דבר לא ידוע על פונקציית המטרה $f:X\rightarrow\{0,1\}$.
- בדומה למודל PAC, יש התפלגות D (קבועה אך לא ידועה) מעל X שהדגימות מתפלגות לפיה ומתויגות ע"י f.
- האלגוריתם בוחר היפותזה $h\in\mathcal{H}$.
- נסמן:

$$\varepsilon_{f,D}^{OPT}\left(\mathcal{H}\right):=\min_{h\in\mathcal{H}}\left\{ \underbrace{err_{f,D}\left(h\right)}_{\Pr_{x\sim D}[f(x)\neq h(x)]}\right\}$$

נשים לב שאם $f\in\mathcal{H}$ אז $\varepsilon_{f,D}^{OPT}\left(\mathcal{H}\right)=0$.

הגדרה 19. אלגוריתם למידה אגנוסטי מקבל גישה למדגם מתויג ופרמטרים $\varepsilon,\delta\in(0,1)$, וצריך לפלוט היפותזה $h:X\rightarrow\{0,1\}$ שעבורה מתקיים בהסתברות לפחות $1-\delta$

$$err_{f,D}\left(h\right)\leq\varepsilon_{f,D}^{OPT}\left(\mathcal{H}\right)+\varepsilon.$$

משפט 17. יהי A אלגוריתם שבהינתן מדגם מתויג פולט היפותזה $h \in \mathcal{H}$ שמביאה למינימום את השגיאה האמפירית על המדגם. כלומר, בהינתן מדגם $S = \{(x^i, f(x^i))\}_{i=1}^m$, האלגוריתם מוצא $h \in \mathcal{H}$ כך ש-

$$\mu_S(h) = \frac{1}{m}\left|\left\{j \mid h\left(x^j\right) \neq f\left(x^j\right)\right\}\right|$$

מקבל מינימום מעל כל הפונקציות ב-$\mathcal{H}$. לכל פונקציית מטרה f והתפלגות D, אם $m \geq \frac{2}{\varepsilon^2}\ln\left(\frac{|\mathcal{H}|}{\delta}\right)$, והמדגם מתפלג לפי D ומתויג לפי f, אז בהסתברות לפחות $1-\delta$ ההיפותזה h מקיימת

$$err_{f,D}(h) \leq \varepsilon_{f,D}^{OPT}(\mathcal{H}) + \varepsilon.$$

הוכחה. (משפט 17) נסמן ב-$B_{\varepsilon,D,\mathcal{H}}(f)$ את תת-קבוצת הפונקציות $g \in \mathcal{H}$ עבורן מתקיים

$$err_{f,D}(g) > \varepsilon_{f,D}^{OPT}(\mathcal{H}) + \varepsilon.$$

לכל $g \in \mathcal{H}$ ו-$j \in [m]$, נסמן ב-χ_g^j מ"מ אינדיקטור למאורע בו $g\left(x^j\right) \neq f\left(x^j\right)$. אזי,

$$\mu_S(g) = \frac{1}{m}\sum_{j=1}^m \chi_g^j, \quad \Pr\left[\chi_g^j = 1\right] = err_{f,D}(g).$$

נסמן ב-h^* את הפונקציה ב-$\mathcal{H}$ שמקיימת $\varepsilon_{f,D}^{OPT}(\mathcal{H}) = err_{f,D}(h^*)$ (הטובה ביותר שניתן למצוא ב-$\mathcal{H}$). נראה שבהסתברות לפחות $1-\delta$ מתקיים:

1. עבור h^*:

$$\mu_S(h^*) \leq \varepsilon_{f,D}^{OPT}(\mathcal{H}) + \frac{\varepsilon}{2}.$$

2. לכל $g \in B_{\varepsilon,D,\mathcal{H}}(f)$ מתקיים

$$\mu_S(g) > \varepsilon_{f,D}^{OPT}(\mathcal{H}) + \frac{\varepsilon}{2}.$$

נכונות המשפט נובעת ישירות משתי הטענות, כי בהסתברות $1-\delta$ מתקיים $\mu_S(g) > \mu_S(h^*)$ לכל $g \in B_{\varepsilon,D,\mathcal{H}}(f)$. בהינתן שזה קורה, כל פונקציה ב-$\mathcal{H}$ שמביאה למינימום את השגיאה האמפירית לא שייכת ל-$B_{\varepsilon,D,\mathcal{H}}(f)$. מכאן,

$$err_{f,D}(h) \leq \varepsilon_{f,D}^{OPT}(\mathcal{H}) + \varepsilon.$$

כעת נוכיח את הנותר: מהאבחנה השנייה, $\Pr\left[\chi_{h^*}^j = 1\right] = err_{f,D}(h^*) = \varepsilon_{f,D}^{OPT}(\mathcal{H})$. נשתמש בצ'רנוף חיבורי ונקבל

$$\Pr\left[\mu_S(h^*) > \varepsilon_{f,D}^{OPT}(\mathcal{H}) + \varepsilon/2\right] \leq \frac{\delta}{|\mathcal{H}|}.$$

מאידך, לכל $g \in B_{\varepsilon,D,\mathcal{H}}(f)$ מתקיים

$$\Pr\left[\chi_g^j = 1\right] = err_{f,D}(g) > \varepsilon_{f,D}^{OPT}(\mathcal{H}) + \varepsilon.$$

נפעיל צ'רנוף חיבורי ונקבל

$$\Pr\left[\mu_S(g) < \varepsilon_{f,D}^{OPT}(\mathcal{H}) + \varepsilon/2\right] \leq \frac{\delta}{2|\mathcal{H}|}.$$

מכאן, ניקח חסם איחוד על כל המאורעות ה-"רעים", ונסיים את ההוכחה. □