TA Session 10: Random Algorithms and De-Randomization

בעיית *max-3-CNF-SAT*

נתונה נוסחת *3-CNF*, φ. נרצה למצוא השמה לנוסחה, כך שכמה שיותר פסוקיות יהיו מסופקות.

<u>אלגוריתם אקראי</u>: לכל משתנה ניתן 0 או 1 בהסתברות חצי.

התוחלת של הפסוקיות המסופקות: לפסוקית מסויימת, צריך שכל 3 המשתנים לא יהיו מסופקים. כלומר $\frac{1}{8} = \left(\frac{1}{2}\right)^3$.

אז התוחלת של מספר הפסוקיות המסופקות היא $m \cdot \frac{7}{8}$, כלומר קיבלנו $\left(\frac{7}{8}\right)$-קירוב.

<u>נעשה דה-רנדומיזציה</u>: נגדיר W משתנה מקרי שסופר את מספר הפסוקיות שסופקו. $E[W] = \frac{7}{8}m$.

ניזכר בנוסחת התוחלת השלמה. לכל משתנה x:

$$E[w] = E[w \mid x = T] \cdot \underbrace{P[x = T]}_{0.5} + E[w \mid x = F] \cdot \underbrace{P[x = F]}_{0.5} = \frac{1}{2}\left(\underbrace{E[w \mid x = T]}_{a} + \underbrace{E[w \mid x = F]}_{b}\right)$$

אז בוודאות, אחד מתוך a או b הם לפחות $E[w]$. נחשב את שניהם ונבחר את מה שמגדיל את התוחלת (אם הם שווים אז נבחר אחד שרירותי).

לכל $i = 1$ עד n:

נחשב את $E(w \mid x_1 = b_1, x_2 = b_2, \dots, x_{i-1} = b_{i-1}, x_i = 0)$ ואת $E(w \mid x_1 = b_1, x_2 = b_2, \dots, x_{i-1} = b_{i-1}, x_i = 1)$.

ונקבע את b_i לפי מ שמגדיל את התוחלת. בגלל שבכל שלב, התוחלת לא קטנה, אז האלגוריתם נותן לפחות $\left(\frac{7}{8}\right)$-קירוב.

נשים לב שזה עובד לכל k, לא רק 3. כלומר לכל נוסחת *k-CNF*, נקבל $\left(1 - \left(\frac{1}{2}\right)^k\right)$-קירוב.

<u>ניסיון לשיפור האלגוריתם</u>: במקום השמה בהסתברות חצי, נבחר 1 בהסתברות p ו-0 בהסתברות $1 - p$

עבור פסוקית עם a ליטרלים חיוביים (x) ו-b ליטרלים שליליים $(\bar{x})$, ההסתברות שהיא לא תסופק היא $p^b(1-p)^a$.

נניח בה"כ ש- $p > 1/2$. ההסתברות שפסוקית C תסופק:

$$P[C\ is\ SAT] = 1 - p^b \underbrace{(1-p)}_{<0.5}{}^a > 1 - p^b p^a = 1 - p^k$$

אז $E[w] > m \cdot \left(1 - p^k\right)$. וכיוון ש- $p > 0.5$, קיבלנו תוצאה פחות טובה.

ניסיון שיפור באמצעות *LP*

<u>נגדיר את בעיית ה-*IP*</u>: לכל פסוקית C, נגדיר $P(C)$ את הליטרלים החיוביים שלה, ו- $N(C)$ את הליטרלים השליליים.

ונגדיר $Z_C \in \{0,1\}$ אינדיקטור לכך שהפסוקית מסופקת. לכל משתנה x_i נגדיר $y_i \in \{0,1\}$.

$$\max \sum_C Z_C\,, \qquad \sum_{i \in P(C)} y_i + \sum_{i \in N(C)} 1 - y_i \geq Z_C$$

אם פסוקית לא מסופקת, זה אומר שכל המשתנים $i \in P(C)$ קיבלו 0, וכל המשתנים $i \in N(C)$ קיבלו 1. אז הסכום יהיה 0, וזה יגביל את Z_C.

<u>נעשה *relaxation*</u>:

הכל אותו דבר אבל: $Z_C \in [0,1],\ \ y_i \in [0,1]$.

נקבל (y^*, Z^*) אופטימליים עבור ה-*LP*. לכל משתנה, ניתן השמה 1 בהסתברות y_i^*.

<u>נחשב את תוחלת יחס הקירוב</u>: לכל פסוקית, ההסתברות שהיא לא מסופקת:

$$p_C := \prod_{i \in P(C)} (1 - y_i^*) \cdot \prod_{i \in N(C)} y_i^*$$

TA Session 10: Random Algorithms and De-Randomization

נגדיר:

$$a_i \coloneqq \begin{cases} y_i^*, & i \in N(C) \\ 1 - y_i^*, & i \in P(C) \end{cases}$$

לפי אי-שוויון $AM\text{-}GM$ (קבלו את זה כנתון, לא צריך להוכיח), מתקיים:

$$p_C = \prod_{i=1}^{k} a_i \le \left(\frac{1}{k}\sum_{i=1}^{k} a_i\right)^k = \left(\frac{1}{k}\left[\sum_{i\in P(C)} (1-y_i^*) \cdot \sum_{i\in N(C)} (y_i^*)\right]\right)^k = \left(\frac{1}{k}\left[\sum_{i\in P(C)} 1 - \sum_{i\in P(C)} y_i^* - \sum_{i\in N(C)} (1-y_i^*) + \sum_{i\in N(C)} 1\right]\right)^k$$

$$= \left(\frac{|P(C)| + |N(C)|}{k} - \frac{1}{k}\left[\sum_{i\in P(C)} y_i^* + \sum_{i\in N(C)} (1-y_i^*)\right]\right)^k = \left(1 - \frac{1}{k}\left[\sum_{i\in P(C)} y_i^* + \sum_{i\in N(C)} (1-y_i^*)\right]\right)^k$$

ולפי ההגבלות שקבענו,

$$\sum_{i\in P(C)} y_i^* + \sum_{i\in N(C)} (1-y_i^*) \ge Z_C^*$$

אז:

$$p_C \le \left(1 - \frac{1}{k} Z_C^*\right)^k$$

אז ההסתברות ש-C מסופקת היא לפחות:

$$\Pr[C\ is\ SAT] \ge 1 - \left(1 - \frac{1}{k} Z_C^*\right)^k$$

מתקיים (עוד אי שוויון נתון):

$$1 - \left(1 - \frac{x}{k}\right)^k \ge \left(1 - \left(1 - \frac{1}{k}\right)^k\right)x, \qquad \forall k \ge 0, x \in [0,1]$$

אז נוכל לרשום:

$$\Pr[C\ is\ SAT] \ge \left(1 - \left(1 - \frac{1}{k}\right)^k\right) z_C^*$$

אז התוחלת של מספר הפסוקיות המסופקות:

$$E[w] = \sum_C \Pr[C\ is\ SAT] \ge \sum_C \left(1 - \left(1 - \frac{1}{k}\right)^k\right) z_C^* = \left(1 - \left(1 - \frac{1}{k}\right)^k\right) \sum_C z_C^* =$$

ונשים לב ש- $\sum_C Z_C$ זה הביטוי שאנחנו צריכים למקסם, ו-Z^* זה הפתרון האופטימלי. אז:

$$= \left(1 - \left(1 - \frac{1}{k}\right)^k\right) \mathrm{OPT}_f \ge \left(1 - \left(1 - \frac{1}{k}\right)^k\right) OPT$$

נשים לב שזה בערך $\left(1 - \frac{1}{e}\right)$-מקרב, כי:

$$\lim_{k\to\infty} \left(1 - \frac{1}{k}\right)^k = \frac{1}{e}$$

אז

$$1 - \left(1 - \frac{1}{k}\right)^k \ge 1 - \frac{1}{e}, \qquad \forall k > 1$$

זה בעצם פחות טוב מהיחס $\left(1 - \frac{1}{8}\right)$ שהיה לנו בהתחלה, עם האלגוריתם הכי פשוט.