Lecture 5

Monday, December 5, 2022 16:37

For two strings s_1, s_2 of the same length L, we define the HAMMING DISTANCE between s_1 & s_2 to be $d(s_1, s_2) := |\{i : s_1[i] \neq s_2[i]\}|$

so $d(\text{horse}, \text{force}) = 2$

For a string c and set S of strings of length L

$$d(S, c) := \max_{s \in S} d(s, c)$$

Closest String problem ← (over alphabet Σ)

In: n strings of length L

Task: find center string c (of length L)

s.t $d(S, c)$ is minimized

GOAL: $(1+\varepsilon)$-Approximation Algorithm

ALGORITHM 1:

ILP For closest string:

For every $i \leq L$, $\underbrace{\alpha \in \Sigma}_{\text{letter in alphabet}}$ $\Longrightarrow$ $x_{i,\alpha} \in \{0, 1\}$

"1 means $c[i] = \alpha$"

Constraints: "at every position we pick exactly one letter"

$$\forall i : \sum_{\alpha \in \Sigma} x_{i,\alpha} = 1$$

How to encode distance ??

Min d

s.t. $\forall s \in S : \sum_{i=1}^{L} 1 - x_{i,s[i]} \le d$

↓

Solve LP and get optimal values for $\{x_{i,\alpha}\}$ and d.

Note: $d = OPT_{LP} \le OPT_{ILP} = OPT$

How to round?

For every i we have $$\underbrace{\sum_{\alpha \in \Sigma} x_{i,\alpha} = 1}_{\text{THIS IS A PROBABILITY DISTRIBUTION!}}$$

Set $c[i] = \alpha$ with probability $x_{i,\alpha}$

Fix some $s \in S$.

what is $E[d(c,s)]$?

Method of indicator random vars!

(1 if $c[i] \neq s[i]$, 0 if $c[i] = s[i]$)

$$\Pr[c[i] \neq s[i]] = 1 - x_{i,s[i]}$$

So $E[d(c,s)] = \sum_{i=1}^{L} 1 - x_{i,s[i]} \leq d = OPT_{LP}$!

Furthermore, $d(c,s)$ is sum of independent 0-1 variables, so CHERNOFF applies

We have that $E[d(c,s)] \leq d = OPT_{LP} \leq OPT_{ILP}$

$$\Pr[d(c,s) - OPT > \varepsilon \cdot OPT] \leq e^{-\varepsilon^2 OPT/3}$$

Now, by union bound over all n choices of $s \in S$

We get

$$\Pr[d(c,S) \geq OPT(1+\varepsilon)] \leq \frac{n}{e^{\varepsilon^2 \cdot OPT/3}}$$

So, if $\dfrac{n}{e^{\varepsilon^2 \cdot OPT/3}} \leq \dfrac{1}{2}$ we win!

$$\Updownarrow$$

$$n \leq e^{\varepsilon^2 OPT/3}$$

$$\Updownarrow$$

$$\ln n \le \varepsilon^2 \mathrm{OPT}/3$$

$$\Updownarrow$$

$$\frac{3 \ln n}{\varepsilon^2} \le \mathrm{OPT}$$

So if OPT is at least $\sim \ln n/\varepsilon^2$ we win!

what to do when OPT is really small?!

IDEA: CAN AFFORD RUNNING TIME

$$(1/\varepsilon)^{O(\mathrm{OPT})} \cdot \mathrm{poly}(n)$$

$$\|$$

$$e^{\ln 1/\varepsilon \cdot \ln n/\varepsilon^2} = n^{\ln(1/\varepsilon)/\varepsilon^2}$$

NEXT: $(1+\varepsilon)$-APPROXIMATION WITH RUNNING TIME poly(n)

& success probability $1/e^{\ln 1/\varepsilon \cdot \ln n/\varepsilon^2}$

(recall time $\approx 1/$success prob)

Suppose we know OPT

(what if we dont?)

ALG:

set $c = s$ for arbitrary $s \in S$

```
for i = 1 to L
        if $d(c,S) \le (1+\varepsilon)\,OPT$
                return S
        let s be the string in S furthest from c.
        let $Z = \{j : s[j] \neq c[j]\}$
        Select $p \in Z$ u.a.r
        Set $c[p] = s[p]$
RETURN FAIL
```

- ALG always runs in poly time.
- IF it doesn't fail, it returns a c s.t $d(c,S) \le (1+\varepsilon)\,OPT$

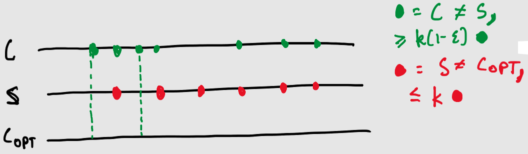
What is Pr [doesn't fail] ?

Let C_{OPT} be some unknown optimal solution

- We say ALG succeeds in step i if, for the selected position p it holds that : $s[p] = C_{OPT}[p]$
- If ALG succeeds in step i then $d(c, C_{OPT})$ decreases by 1 [WHY ?]

- $\Pr[\text{ALG succeeds in step 1}] = \dfrac{\{p \in Z : s[p] = C_{OPT}[p]\}}{|Z|}$

But $\{p \in Z : S[p] = C_{OPT}[p]\} \geq \varepsilon\, OPT$



$$= \frac{\{p \in Z : S[p] = C_{OPT}[p]\}}{2 \cdot OPT} \geq \frac{\varepsilon \cdot OPT}{2 \cdot OPT} \geq \frac{\varepsilon}{2}$$

- $\Pr[\text{ALG succeeds in step } 2] = \dfrac{\{p \in Z : S[p] = C_{OPT}[p]\}}{|Z|}$

$$\geq \frac{\varepsilon}{2}$$

- $\Pr[\text{ALG succeeds in step } 3] \geq \dfrac{\varepsilon}{2}$

. . .

$$\Pr[\text{ALG succeeds OPT steps in a row}] \geq \left(\frac{\varepsilon}{2}\right)^{OPT}$$

Note that if ALG succeeds OPT steps in a row
then $d(c, C_{OPT}) = 0$ and so alg outputs c!

Hence $\Pr[\text{Alg outputs } c] \geq \left(\frac{\varepsilon}{2}\right)^{OPT}$

↳ For success prob ≥ 1/2, repeat
$\left(\frac{2}{\varepsilon}\right)^{OPT}$ times.

COMBINED ALG:

If $\frac{3\ln n}{\varepsilon^2} \le OPT$ Solve LP & Round

otherwise, do local search ALG in time $\left(\frac{2}{\varepsilon}\right)^{3\ln n/\varepsilon^2}$

$\frac{3\ln n}{\varepsilon^2} \ge OPT$

$\le n^{O(\log 1/\varepsilon / \varepsilon^2)}$

HOEFFDING BOUND (Kind of like CHERNOFF)

If $X = X_1 + X_2 \dots + X_n$ and

- $a_i \le X_i \le b_i$
- X_i's are independent

$$\Pr\left[\,|X - E[X]| > t\,\right] \le 2e^{-2t^2/\sum_i (b_i - a_i)^2}$$

MAX CUT (AGAIN!)

In: G, min degree in G is $\ge c\cdot n$

Task: Find $S \subseteq V(G)$, maximizing $\partial(X)$

Will see a $(1+\varepsilon)$-appx in time $n^{O(1/\varepsilon^2\cdot c)}$??

OBS 1: $m \ge \frac{cn^2}{2}$ AND $\exists$ cut of size $\ge \frac{cn^2}{4}$

⇒ Will go for an algorithm that produces in time $n^{O(1/\varepsilon^2)}$ a cut X with $\partial(X) \geq OPT - \varepsilon n^2$

Why is this enough for a $(1+\varepsilon')$-approximation??

OBS 2: wlog the optimal soln X_{OPT} SATISFIES

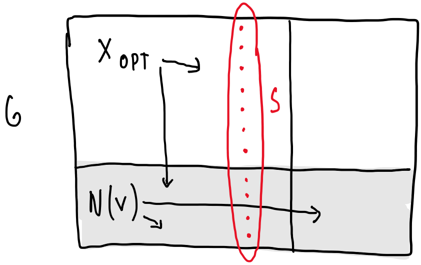
$$|X_{OPT}| \geq n/2$$

OBS 3: Let $x = |X_{OPT}|$. We can assume we know x

[why?]

Algorithm: Sample a set S on $\frac{\ln n}{\varepsilon^2}$ vertices from $V(G)$

Define $\hat{d}(v) = \frac{|S \cap N(v)|}{|S|} \cdot x$



EXPECT $\sim |S| \cdot \frac{N(v) \cap X_{OPT}}{X_{OPT}}$ vertices in $N(v)$

Good EVT 1: $S \subseteq X_{OPT}$

Good EVT 2: $\forall v \in V(G)$

$$\Big|\,|N(v)\cap X_{OPT}| - \hat{d}(v)\Big| \leq \varepsilon \cdot n$$

$$\Pr[S \subseteq X_{OPT}] \geq 1/2^{|S|} \geq 1/n^{1/\varepsilon^2}$$

$$\Pr[\text{Good Evt } 2 \mid S \subseteq X_{OPT}] = ??$$

$$E\big[|S\cap N(v)| \;\big|\; S \subseteq X_{OPT}\big] = ?$$

METHOD OF INDICATOR R.VARS!

1 if i'th vertex of S is in N(v)

$$E[\text{indicator} \mid S \subseteq X_{OPT}] = \Pr[\text{i'th vertex in } S \text{ is in } N(v) \mid S \subseteq X_{OPT}]$$

$$= \frac{|N(v)\cap X_{OPT}|}{|X_{OPT}|}$$

$$E\big[|S\cap N(v)| \;\big|\; S \subseteq X_{OPT}\big] = \frac{|S|\cdot|N(v)\cap X_{OPT}|}{X}$$

By HOEFFDING

$$\Pr\Big[\Big|\,|S\cap N(v)| - E\big[|S\cap N(v)|\,\big|\,S \subseteq X_{OPT}\big]\Big| > \varepsilon\cdot|S| \;\Big|\; S \subseteq X_{OPT}\Big]$$

$$\leq 2e^{-2\varepsilon^2|S|^2/|S|}$$

$$\leq 2e^{-2\varepsilon^2|S|}$$

$$\leq 2e^{-2\varepsilon^2 \ln n/\varepsilon^2}$$

$$= 2/n^2$$

Thus, conditioned on $S \subseteq X_{OPT}$ by UNION BOUND with probability $\geq 1 - 2/n$ it holds that

$$\forall v : \Big| |S \cap N(v)| - E\big[|S \cap N(v)| \,\big|\, S \subseteq X_{OPT}\big] \Big| \leq \varepsilon \cdot |S|$$

BUT WE SHOWED

$$E\big[|S \cap N(v)| \,\big|\, S \subseteq X_{OPT}\big] = \frac{|S| \cdot |N(v) \cap X_{OPT}|}{X}$$

Thus

$$\forall v : \left| |S \cap N(v)| - \frac{|S| \cdot |N(v) \cap X_{OPT}|}{X} \right| \leq \varepsilon \cdot |S|$$

But $\hat{d}(v) = |S \cap N(v)| \cdot \frac{X}{|S|}$

So $\forall v : \big| \hat{d}(v) - |N(v) \cap X_{OPT}| \big| \leq \varepsilon \cdot X \leq \varepsilon n$

We now have that:

- $\partial(X_{OPT}) = \sum_{u \notin X_{OPT}} |N(u) \cap X_{OPT}|$
- So $\left| \partial(X_{OPT}) - \sum_{u \notin X_{OPT}} \hat{d}(u) \right| \leq \varepsilon n^2$

Thus $\sum_{u \notin X_{OPT}} \hat{d}(u)$ can act as a good "proxy" for the actual objective function!

Consider the following ILP:

variable x_v for every vertex $v \in V(G)$

$x_v = 1$ means $v \in X$

Maximize $\sum_{v \in V(G)} (1 - x_v) \cdot \hat{d}(v)$

s.t. $\sum_{v \in V(G)} x_v = x$ $(= |X_{OPT}|)$ delete?

$\forall u$: $\hat{d}(u) - \varepsilon n \leq \sum_{v \in N(u)} x_v \leq \hat{d}(u) + \varepsilon n$

X_{OPT} (or rather its x_v-variables) is a feasible solution to this ILP,

AND ITS OBJ. VALUE IS $\geq \partial(X_{OPT}) - \varepsilon n^2$

Thus, we can solve this LP and get a LP solution $\{x_v\}$ with value

$$\geq \partial(X_{OPT}) - \varepsilon n^2$$

We have an LP solution, need a set X.

Do the randomized rounding trick:

add v to X with probability x_v

By linearity of expectation

$$E\left[\sum_{v\in V(G)} \left[\begin{matrix}1 \text{ if } v\notin X\\ 0 \text{ otherwise}\end{matrix}\right]\cdot \hat{d}(v)\right] = E\left[\sum_{v\notin X}\hat{d}(v)\right]$$

$$= OPT_{LP} \geq \partial(X_{OPT}) - \varepsilon n^2$$

Furthermore, by HOEFFDING

$$\sum_{v\notin X}\hat{d}(v) \geq \partial(X_{OPT}) - 2\varepsilon n^2$$

with probability $\geq 1-2e^{-2\varepsilon^2 n}$ [GOOD EVENT 3]

For every vertex u it holds that

$$\hat{d}(u) - \varepsilon n \leq E\left[|N(u)\cap X|\right] \leq \hat{d}(u) + \varepsilon n$$

Since $|N(u)\cap X|$ is a sum of independent 0-1 vars
we get by HOEFFDING + UNION BOUND [TUTORIAL Q!]
with prob $\geq 1-2e^{-2\varepsilon^2 n}$

[GOOD EVENT 4]

that $\forall u$:

$$\hat{d}(u) - 2\varepsilon n \leq E\left[|N(u)\cap X|\right] \leq \hat{d}(u) + 2\varepsilon n$$

If both good event 3 and 4 happen
(and they do with probability close to 1)

Then

$$\begin{aligned} \partial(X) &= \sum_{u\notin X} |N(u)\cap X| \\ \text{(4)} \qquad &\geq \sum_{u\notin X} \hat{d}(u) - 2\varepsilon n \end{aligned}$$

$$\geq -2\varepsilon n^2 + \sum_{u \notin X} \hat{d}(u)$$

$$(3) \quad \geq -2\varepsilon n^2 + \partial(X_{OPT}) - 2\varepsilon^2 n$$

$$= \partial(X_{OPT}) - 4\varepsilon^2 n$$

Overall algorithm:

```
repeat $n^{1/\varepsilon^2}$ times
    Set $s = \frac{\ln n}{\varepsilon^2}$
    Sample $S \subseteq V(G)$ of size $s$ u.a.r.
    Compute $\hat{d}(u)$ for every $u$
    Formulate and solve LP using $\hat{d}(u)$'s.
    $X \leftarrow$ randomized rounding
Keep best $X$.
```