Approximation Algorithms

Lecture 3

Evaluation Policy

- 3 Assignments, one each at the end of Feb, March, April: 15% each
- Attendance and surprise quizzes: 15%
- Endsem: 40%

Last Time

- Dual Rounding
- Weak Duality Theorem
- Strong Duality Theorem
- Complementary Slackness Condition
- Primal Dual Method

Today

- Dual Fitting for Weighted Set Cover
- Randomized Rounding for Weighted Set Cover
  - Basics of probability and analyzing randomized algorithms

Today

- ❑ Dual Fitting for Weighted Set Cover
- ❑ Randomized Rounding for Weighted Set Cover
  - ❑ Basics of probability and analyzing randomized algorithms

Problem Definition (Recall)

- Universe $U = \{e_1, \ldots e_n\}$, a family $F = \{F_1, \ldots F_m\}$ of subsets of U
- Set F_i has weight w_i for all $i \in [m]$
- Output a collection of F_i's of minimum total weight whose union is U

Greedy Algorithm (Part of reading exercise)

- ❑ $I \leftarrow \emptyset$
- ❑ $\widehat{F_j} \leftarrow F_j$ for all $j \in [m]$
- ❑ while I is not a set cover
  - ❑ $\ell \leftarrow \arg\min_{j:\widehat{S_j}\neq\emptyset} \frac{w_j}{|\widehat{F_j}|}$
  - ❑ $I \leftarrow I \cup \{\ell\}$
  - ❑ $\widehat{F_j} \leftarrow \widehat{F_j} \setminus F_\ell$ for all $j \in [m]$

Theorem: The greedy algorithm is an H_n approximation for weighted set cover, where $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ is the n-th Harmonic number.

Greedy Algorithm (Part of reading exercise)

- $I \leftarrow \emptyset$
- $\widehat{F_j} \leftarrow F_j$ for all $j \in [m]$
- while I is not a set cover
  - $\ell \leftarrow \arg\min_{j:\widehat{S_j} \neq \emptyset} \frac{w_j}{|\widehat{F_j}|}$
  - $I \leftarrow I \cup \{\ell\}$
  - $\widehat{F_j} \leftarrow \widehat{F_j} \setminus F_\ell$ for all $j \in [m]$

Theorem: The greedy algorithm is an H_n approximation for weighted set cover, where $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ is the n-th Harmonic number.

Next: A different analysis for the above algorithm using the dual LP

Dual LP

$$\max \sum_{e \in U} y_e$$

$$(D) \quad \sum_{e : e \in F_j} y_e \leq w_j \quad \forall j \in [m]$$

$$y_e \geq 0 \quad \forall e \in U$$

- Let g denote the maximum cardinality among the sets in F
- **Theorem**: The greedy algorithm gives an H_g-approximation to the weighted set cover.

$1 + \frac{1}{2} + \cdots + \frac{1}{g}$

  - Guarantee strictly better than H_n whenever $g < n$

- Let g denote the maximum cardinality among the sets in F
- **Theorem**: The greedy algorithm gives an H_g-approximation to the weighted set cover.

  - Guarantee strictly better than H_n whenever $g < n$

$e_i \in F_j$ when F_j was added to the cover

- **Proof of Theorem**

- We first construct a dual solution, i.e, set values for y_i for $i \in [n]$
- Assume that element e_i is covered <u>for the first time</u> when set F_j was added to the cover
- We set $y_i \leftarrow \frac{w_j}{|\widehat{F_j}|} \cdot \frac{1}{H_g}$ for each $e_i \in U$

When I add a set F_j to the cover,

– Cover weight increases by w_j

Each of the $|\hat{F}_j|$ uncovered elements pay a price of $\frac{1}{H_g} \cdot \frac{w_j}{|\hat{F}_j|}$ $\quad \frac{1}{H_g} \cdot \frac{w_\ell}{|\hat{F}_\ell|}$ $\quad e' \in F_\ell$

$|F_{j'}| = g$

Suppose in each iteration of greedy, ~~exactly~~ at most one el. of $F_{j'}$ gets covered.

$e' \in F_{j'}$

$\frac{w_{j'}}{g} \geq \frac{w_\ell}{|\hat{F}_\ell|}$ greedy prop.

Price paid by e'

$$= \frac{1}{H_g} \cdot \frac{w_\ell}{|\hat{F}_\ell|} \leq \frac{1}{H_g} \cdot \frac{w_{j^0}'}{g}$$

---

for second element

$$\leq \frac{1}{H_g} \cdot \frac{w_{j^0}'}{g-1}$$

$\vdots$

$$\left.\begin{array}{l}\text{Total price paid by} \\ \text{elements in } F_{j^0}'\end{array}\right\} \leq \frac{1}{H_g} \cdot w_{j^0}' \overbrace{\left(\frac{1}{g} + \frac{1}{g-1} + \cdots + 1\right)}^{H_g}$$

$$= w_{j^0}'$$

$$F_{j_0,1} = \{ e_{u_1}, e_{u_2}, \dots \, e_{u_g} \}$$

$$\leq \frac{w_{j_0,1}}{g} \cdot \frac{1}{H_g} \qquad \leq \frac{w_{j_0,1}}{g-1} \cdot \frac{1}{H_g} + \dots$$

- Let g denote the maximum cardinality among the sets in F
- **Theorem**: The greedy algorithm gives an H_g-approximation to the weighted set cover.
  - Guarantee strictly better than H_n whenever $g < n$
- Proof of Theorem
- We first construct a dual solution, i.e, set values for y_i for $i \in [n]$
- Assume that element e_i is covered <u>for the first time</u> when set F_j was added to the cover
- We set $y_i \leftarrow \frac{w_j}{|\widehat{F_j}|} \cdot \frac{1}{H_g}$

This is a feasible dual solution!

- Consider set F_j and its dual constraint: $y_{u_1} + y_{u_2} + \cdots + y_{u_k} \leq w_j$
- Elements $e_{u_1}, \dots e_{u_k}$ are covered for the first time in different iterations

- Consider set F_j and its dual constraint: $y_{u_1} + y_{u_2} + \cdots + y_{u_k} \leq w_j$
- Elements $e_{u_1}, \ldots e_{u_k}$ are covered for the first time in different iterations

- Let r denote the overall number of iterations of the greedy algo.
- Let $A_t \subseteq F_j$ be elements that got covered for the first time in t-th iteration
- Let a_t be the number of elements in F_j that are uncovered at the beginning of t-th iteration: $|A_t| = a_t - a_{t+1}$

$$\underbrace{y_{u_1} + y_{u_2} + \cdots + y_{u_k}}_{\text{LHS of dual constraint of } F_j} = \sum_{t=1}^{r} \sum_{i:\, e_i \in A_t} y_i$$

- Let F_p be the set added by the greedy algorithm in t-th iteration
- Then $y_i = \frac{w_p}{|\widehat{F_p}|} \cdot \frac{1}{H_g}$ for each $e_i \in A_t$

$A_t \subseteq F_j$ that got covered for first time in t-th iteration

- Now, $y_i = \frac{w_p}{|\widehat{F_p}|} \cdot \frac{1}{H_g} \leq \frac{w_j}{a_t} \cdot \frac{1}{H_g}$ for each $e_i \in A_t$ by the greedy property

by defn. of dual soln.

uncovered elements in F_j at the beginning of t^{th} iteration

$p \in [m]$

$t \in [r]$

- Let F_p be the set added by the greedy algorithm in t-th iteration
- Then $y_i = \frac{w_p}{|\widehat{F_p}|} \cdot \frac{1}{H_g}$ for each $e_i \in A_t$
- Now, $y_i = \frac{w_p}{|\widehat{F_p}|} \cdot \frac{1}{H_g} \leq \frac{w_j}{a_t} \cdot \frac{1}{H_g}$ for each $e_i \in A_t$ by the greedy property

$\leq H_{|F_j|}$

$\leq H_g$

$$\star \quad \sum_{i : e_i \in A_t} y_i \leq \frac{w_j}{H_g} \cdot \frac{|A_t|}{a_t} = \frac{w_j}{H_g} \cdot \frac{(a_t - a_{t+1})}{a_t}$$

$$\star \quad \underbrace{y_{u_1} + y_{u_2} + \cdots + y_{u_k}} = \sum_{t=1}^{\gamma} \sum_{i : e_i \in A_t} y_i \leq \frac{w_j}{H_g} \cdot \sum_{t=1}^{\gamma} \frac{a_t - a_{t+1}}{a_t}$$

- We proved that the dual solution is feasible
- Can we say anything about the quality of the solution of the greedy algorithm based on this dual solution?

- We proved that the dual solution is feasible

- Can we say anything about the quality of the solution of the greedy algorithm based on this dual solution?

$$\star \quad \sum_{i \in [n]} y_i \overset{?}{=} \frac{1}{H_g} \cdot \sum_{j \in I} w_j = \frac{1}{H_g} \cdot \overline{\left(\text{Cost of Soln. output by greedy}\right)}$$

indices of sets in set cover (greedy)

$$\star \quad \sum_{i \in [n]} y_i \leq \text{Value of optimal solution to primal LP} \leq \text{OPT}$$

Weak duality

$\left.\begin{array}{l}\text{Cost of solution} \\ \text{output by greedy}\end{array}\right\} \leq H_g \cdot OPT$

When F_j gets added to greedy set cover,

$y_i \leftarrow \frac{1}{H_g} \cdot \frac{w_j}{|\hat{F}_j|}$ for each element $e_i \in \hat{F}_j$

$$\sum_{e_i \in \hat{F}_j} y_i = \frac{w_j}{H_g}$$

$$\sum_{e_i \in U} y_i = \sum_{j \in I} \sum_{e_i \in \hat{F}_j} \frac{y_j}{H_g}$$

Randomized Rounding

- Based on a random process determined by an optimal solution of the primal LP

LP:

$$\min \sum_{i \in [m]} w_i x_i$$

$$\sum_{i : e \in S_i} x_i \geq 1 \qquad \forall e \in U \qquad (P)$$

$$x_i \geq 0$$

Randomized Rounding Algorithm

- Determine an optimal solution $\{x_j^*\}_{j\in[m]}$ to the primal LP
- $I \leftarrow \emptyset$

Randomized Rounding Algorithm

$[0,1]$

- ❑ Determine an optimal solution $\{x_j^*\}_{j\in[m]}$ to the primal LP
- ❑ $I \leftarrow \emptyset$
- ❑ For each $j \in [m]$: → for each set $F_j \in \mathcal{F}$

  c is some constant

  - ❑ Consider a coin with Heads probability x_j^*
  - ❑ Toss the coin $\underline{c \ln n}$ times independently of each other
  - ❑ If any of the tosses shows up Heads, then $I \leftarrow I \cup \{j\}$
- ❑ Output I as the set cover

Randomized Rounding Algorithm

- ❑ Determine an optimal solution $\{x_j^*\}_{j\in[m]}$ to the primal LP
- ❑ $I \leftarrow \emptyset$
- ❑ For each $j \in [m]$:
  - ❑ Consider a coin with Heads probability x_j^*
  - ❑ Toss the coin $c \ln n$ times independently of each other
  - ❑ If any of the tosses shows up Heads, then $I \leftarrow I \cup \{j\}$
- ❑ Output I as the set cover

Theorem: The algorithm is a $O(\ln n)$-approximation algorithm that produces a set cover with high probability

- What is the probability that the algorithm outputs a set cover?
- What is the probability that a specific element $e \in U$ is not covered?

Need a language to answer these questions concretely!

Brief Intro to Probability

- The probability space associated with a random process is defined by:

  countable or even finite

  - Sample space Ω = Set of outcomes of the random process
  - Set $\mathcal{F}$ of allowable events; each event is a subset of Ω → $\Omega \in \mathcal{F}$
  - A probability function $\Pr: \mathcal{F} \to [0,1]$ satisfying:
    - For any event E, we have $\Pr[E] \in [0,1]$
    - $\Pr[\Omega] = 1$

      & disjoint
    - For any countable sequence of events $E_1, E_2, \ldots$ we have $\Pr[\cup_{i \geq 1} E_i] = \sum_{i \geq 1} \Pr[E_i]$

- Two events E and F are independent iff $\Pr[E \cap F] = \Pr[E] \cdot \Pr[F]$

  if & only if

- What is the probability that the algorithm outputs a set cover?
- What is the probability that a specific element $\boldsymbol{e} \in \boldsymbol{U}$ is not covered?

- For a set F_j s.t. $e \in F_j$, prob. that F_j is not in the cover

$$= (1 - x_j^*)^{c \ln n}$$

- $$\underbrace{\Pr(e \text{ is not covered})}_{\Pr\left[\bigcap_{j: e \in F_j} F_j \text{ not picked}\right]} = \prod_{j: e \in F_j} \Pr(F_j \text{ not picked})$$

$$\Pr(e \text{ is not covered}) = \prod_{j:\, e \in F_j} (1 - x_j^*)^{c \ln n}$$

$$\leq \prod_{j:\, e \in F_j} e^{-x_j^* \cdot c \ln n}$$

$$= e^{-(c \ln n) \underbrace{\sum_{j:\, e \in F_j} x_j^*}_{\geq 1}}$$

$$\leq \frac{1}{n^c}$$

$$\boxed{1 - x \leq e^{-x} \quad \text{for } x \in [0, 1]}$$

- What is the probability that there exists some element $e \in U$ that is not covered?

- **Union Bound:** For a countable sequence of events $E_1, E_2, \ldots$

$$\Pr[\cup_{i\geq 1} E_i] \leq \sum_{i\geq 1} \Pr[E_i]$$

- What is the probability that **there exists** some element $e \in U$ that is not covered?

- **Union Bound:** For a countable sequence of events $E_1, E_2, \ldots$

$$\Pr[\cup_{i\geq 1} E_i] \leq \sum_{i\geq 1} \Pr[E_i]$$

$$\Pr(\text{output not a set cover}) = \Pr(\exists\, e \in U \text{ that is not covered})$$

$$\leq \sum_{e \in U} \frac{1}{n^c} = \frac{1}{n^{c-1}}.$$

e

- What is the weight of the set system output?
- Need the concept of random variables, which are functions $X\colon \Omega \to \mathbb{R}$

discrete random variable

- Expected value of X, denoted $E[X] = \sum_i i \cdot \Pr[X = i]$
- Linearity of expectations: $E\left[\sum_j X_j\right] = \sum_j E[X_j]$
  - For a constant c and a random variable X, we have $E[cX] = c \cdot E[X]$
- What is the expected weight of the set system output?

- $Y_j = 1$ if F_j is included in the output and $Y_j = 0$ otherwise, for $j \in [m]$

indicator r.v.'s

- Weight of set system output $Y = \sum_{j \in [m]} w_j Y_j$

(lin. of exp.)

- Expected weight = $E[Y] = \sum_{j \in [m]} w_j E[Y_j]$

defn. of expectation

- $E[Y_j] = 1 \cdot \Pr[F_j$ is included in output$] + 0 \cdot \Pr[F_j$ is not included in output$]$

- $Y_j = 1$ if F_j is included in the output and $Y_j = 0$ otherwise, for $j \in [m]$

$\leq c \ln n \cdot OPT$

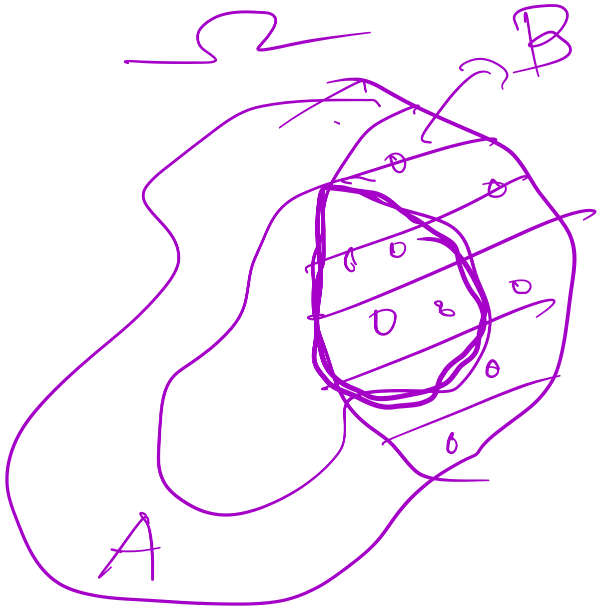
- Weight of set system output $Y = \sum_{j \in [m]} w_j Y_j$
- Expected weight = $\mathbf{E}[Y] = \sum_{j \in [m]} w_j E[Y_j]$ $\leq c \ln n \cdot \sum w_j \cdot x_j^*$
- $E[Y_j] = 1 \cdot \Pr[F_j$ is included in output$] + 0 \cdot \Pr[F_j$ is not included in output$]$

$$= 1 - (1 - x_j^*)^{c \ln n}$$

$$\leq x_j^* \cdot c \ln n$$

$$\boxed{(1-x)^a \geq 1 - ax \quad \text{for } x \in [0,1],\ a \geq 1}$$

- We want the expected weight, conditioned on the event that the set system output is a set cover...



- **Conditional Probability**: For two events A and B, we have
$$\Pr[A|B] = \frac{\Pr[A \cap B]}{P[B]}$$
- **Conditional Expectation**: $E[X|B] = \sum_i i \cdot \Pr[X = i|B]$

- We want the expected weight, conditioned on the event that the set system output is a set cover…

- **Conditional Probability**: For two events A and B, we have

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{P[B]}$$

- **Conditional Expectation**: $E[X|B] = \sum_i i \cdot \Pr[X = i|B]$

- Let H be the event that the set system output is a set cover

$\Pr(H) \geq 1 - \frac{1}{n^{c-1}} \geq \frac{1}{2}$

- $E[Y] = \underbrace{E[Y|H] \cdot \Pr[H] + \overbrace{E[Y|\bar{H}] \cdot \Pr[\bar{H}]}^{\geq 0}}_{\text{law of total expectation}} \geq \frac{1}{2} \cdot E[Y|H]$

$\Rightarrow E(Y|H) \leq 2E(Y) \leq 2c \ln n \, OPT$

Next Time:

* Christofides' approx. algo. for metric TSP

*