Section 1.6: A Greedy Algorithm

A natural greedy choice would be to „pay" as little as possible for each additional covered element:

Alg 1.2 for Set Cover: Greedy

$I \leftarrow \emptyset$

For $j \leftarrow 1$ to m

  $\hat{S}_j \leftarrow S_j$ (uncovered part of S_j)

While $\{S_j \mid j \in I\}$ is not a set cover

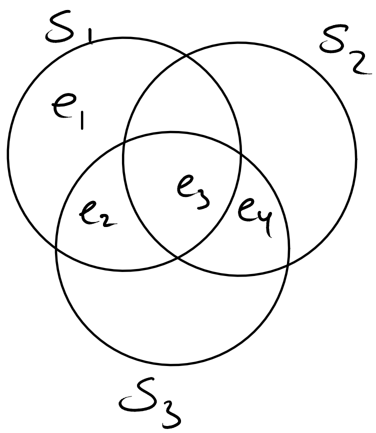
  $\ell \leftarrow \underset{j:\hat{S}_j \neq \emptyset}{\arg\min} \dfrac{w_j}{|\hat{S}_j|}$ (S_ℓ: set with smallest cost per uncovered element)

  $I \leftarrow I \cup \{\ell\}$

  For $j \leftarrow 1$ to m

    $\hat{S}_j \leftarrow \hat{S}_j - S_\ell$

Ex:



$w_1 = 12$

$w_2 = 4$

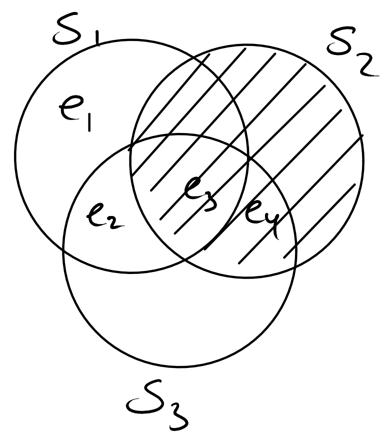
$w_3 = 9$

$$\frac{w_1}{|S_1|} = \frac{12}{3} = 4,$$

$$\frac{w_2}{|S_2|} = \frac{4}{2} = 2$$ ← price per element in first iteration

$$\frac{w_3}{|S_3|} = \frac{9}{3} = 3$$

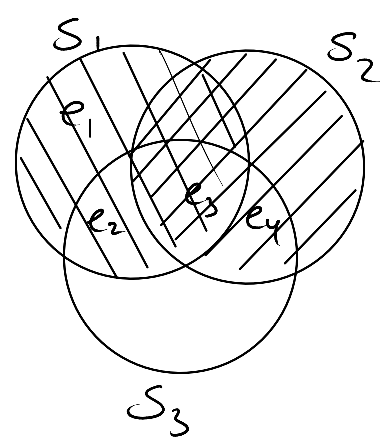
→ Pick S_2



$$\frac{w_1}{|\hat{S}_1|} = \frac{12}{2} = 6$$ ← price per element in second iteration

$$\frac{w_3}{|\hat{S}_3|} = \frac{9}{1} = 9$$

→ Pick S_1



$$\text{Total weight} = \sum_{i=1}^{4} \text{price}(e_i) = 2+2+6+6$$

$$= w_2 + w_1 = 4 + 12$$

$$= 16$$

The greedy alg. is an H_n-approx. alg

Recall: $H_n = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \approx \ln(n)$

It is „likely" that no significantly better approx. ratio can be obtained:

Thm 1.13:

Approx. factor $\frac{\ln n}{c}$, $c > 1$, for unweighted Set Cover

$\Rightarrow \underbrace{n^{O(\log \log n)}}_{\sim k^{\log n}}$-approx alg. for NPC

Thm 1.11

Alg. 1.2 is an H_n-approx. alg. for Set Cover

<u>Proof</u>:

n_k : # uncovered elements at the beginning of the k'th iteration

In the ex. above:

$n = 4$

$n_1 = 4, \quad n_2 = 2, \quad n_3 = 0$

$n_1 - n_2 = 2, \quad n_2 - n_3 = 2$

Any algorithm, including OPT, has to cover these n_k elements using only sets in $\mathcal{S} - \{S_j \mid j \in I\}$, since none of them are contained in $\{S_j \mid j \in I\}$.

Hence, there must be at least one element with a price of at most OPT/n_k. Otherwise, OPT would not be able to cover the n_k elements (and certainly not all n elements) at a cost of only OPT.

Hence, the $n_k - n_{k+1}$ elements covered in iteration k cost at most $(n_k - n_{k+1})\, OPT/n_k$ in total.

Thus, the cost of the set cover produced by the greedy alg. is

$$\sum_{j \in I} w_j \le \sum_{k=1}^{\overset{\text{\# iterations}}{r}} \frac{n_k - n_{k+1}}{n_k} OPT$$

$$= OPT \sum_{k=1}^{r} (n_k - n_{k+1}) \cdot \frac{1}{n_k}$$

$$\le OPT \sum_{k=1}^{r} \Big(\overbrace{\frac{1}{n_k} + \frac{1}{n_k - 1} + \dots + \frac{1}{n_{k+1}+1}}^{n_k - n_{k+1} \text{ terms that are each } \ge \frac{1}{n_k}} \Big)$$

$$= OPT \sum_{s=1}^{n} \frac{1}{s}$$

$$= OPT \cdot H_n \qquad \square$$

Ex from before:

$OPT = w_1 + w_2 = 12 + 4 = 16$

The cost of the greedy alg is

$$w_2 + w_1 = 4 + 12$$

$$= 2 + 2 + 6 + 6$$

$$\le \left(\frac{16}{4} + \frac{16}{4}\right) + \left(\frac{16}{2} + \frac{16}{2}\right)$$

$$\le \left(\frac{16}{4} + \frac{16}{3}\right) + \left(\frac{16}{2} + \frac{16}{1}\right)$$

$$= 16 \cdot \left(\frac{1}{4} + \frac{1}{3} + \frac{1}{2} + \frac{1}{1}\right)$$

$$= 16 \cdot H_4$$

Let $q = \max\{\, |S_j| \mid S_j \in \mathcal{S} \,\}$.

Thm 1.12

Alg. 1.2 is an H_q-approx. alg. for Set Cover

Proof: By Dual Fitting:

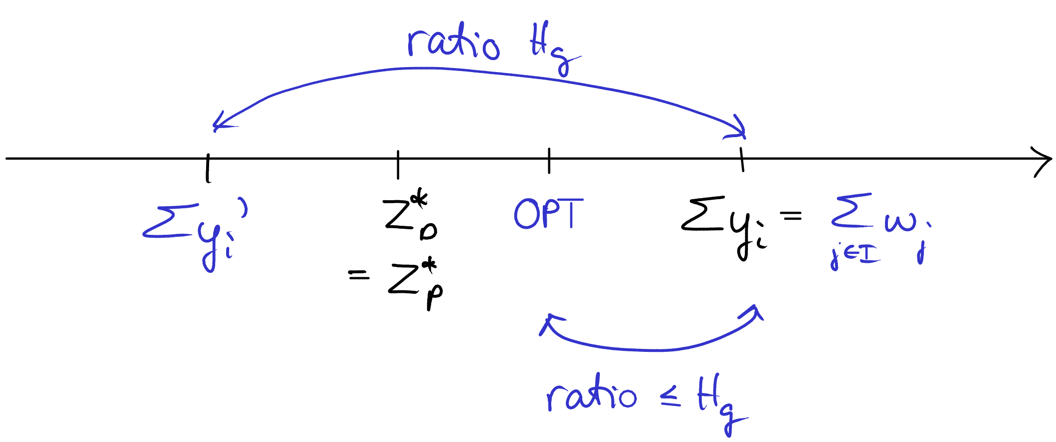
Consider the dual D of the LP for Set Cover.
We will construct an infeasible solution $\vec{y}$ and a feasible solution $\vec{y'}$ such that

- $\sum_{i=1}^{n} y_i = \sum_{j \in I} w_j$
- $y_i' = \frac{y_i}{H_q}, \quad 1 \le i \le n$

Then,

$$\sum_{j \in I} w_j = \sum_{i=1}^{n} y_i = H_q \sum_{i=1}^{n} y_i' \le H_q z_D^* \le H_q \cdot \mathrm{OPT},$$

proving the claimed approximation factor.



For $1 \le i \le n$, let $y_i = \text{price}(e_i)$. Then,

$$\sum_{1 \le i \le n} y_i = \sum_{j \in I} w_j$$

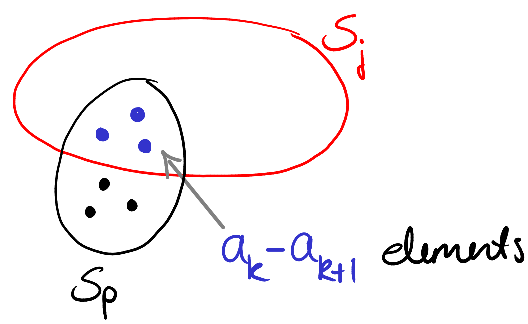
Hence, we just need to show that $\vec{y}$ is feasible:

Consider an arbitrary set S_j.

Let a_k be #uncovered elements in S_j at the beginning of the k'th iteration.

Let S_p be the set chosen by Greedy in the k'th iteration.

S_p covers $a_k - a_{k+1}$ previously uncovered elements in S_j



The price per element in S_j covered in the k'th iteration is at most

$$\frac{w_p}{|\hat{S}_p|} \le \frac{w_j}{a_k}$$

since otherwise S_j would be a <u>more</u> greedy choice. $\lightning$

Thus,

$$\sum_{e_i \in S_j} y_i \le \underbrace{\sum_{k=1}^{r} (a_k - a_{k+1}) \frac{w_j}{a_k}}_{\text{Total \#terms} = |S_j|, \text{ since } a_1 = |S_j| \text{ and } a_{r+1}=0}$$

$$\le w_j \sum_{i=1}^{|S_j|} \frac{1}{i},$$ by the same arguments as in the proof of Thm 1.12.

$$\le w_j \sum_{i=1}^{g} \frac{1}{i}$$

$$= w_j \cdot H_g$$

Hence,

$$\sum_{e_i \in S_j} y_i' = \frac{1}{H_g} \sum_{e_i \in S_j} y_i \le w_j$$

$\square$

Compare the proof of Thm 1.12 (dual fitting) to the proof of Thm 1.11:

- Simpler: Compare prices to w_j instead of OPT
- Stronger result: H_g instead of H_n
  (could also have been obtained with the technique of the proof of Thm 1.11)

Ex from before:

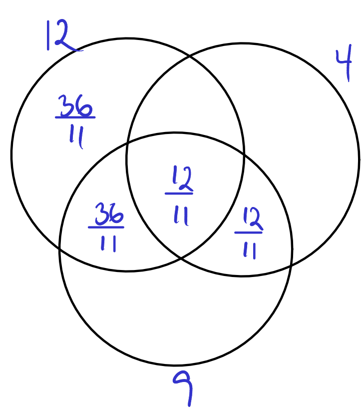
$$y_3 = y_4 = 2$$
$$y_1 = y_2 = 6$$

$$H_3 = 1 + \frac{1}{2} + \frac{1}{3} = \frac{11}{6}$$

$$y_3' = y_4' = \frac{1}{H_3} \cdot 2 = \frac{6}{11} \cdot 2 = \frac{12}{11}$$
$$y_1' = y_2' = \frac{6}{11} \cdot 6 = \frac{36}{11}$$

$\vec{y'}$ is feasible:

$$y_1' + y_2' + y_3' = 2 \cdot \frac{36}{11} + \frac{12}{11} < 8 \leq w_1$$
$$y_3' + y_4' = 2 \cdot \frac{12}{11} < 3 \leq w_2$$
$$y_2' + y_3' + y_4' = \frac{36}{11} + 2 \cdot \frac{12}{11} < 6 \leq w_3$$



Is the upper bound of H_n tight?

If it is, the matching lower bound must come from an instance with

- one set containing all elements
  (follows from the upper bound of H_g)
- only one additional element covered in each it.
  (otherwise, some of the terms in $\frac{1}{n} + \frac{1}{n-1} + \ldots + 1$ would be replaced by smaller terms.)

Ex: