

week5_lec1

Main Ideas

- Set Cover
- Approximations

Set Cover

Let's consider the following:

- U = set of elements
- $C = \{S_1, S_2, \dots, S_n\} \subseteq U, \forall S_i \subseteq U.$
- Set cover: A sub-collection $C' \subseteq C$ whose union is $U, \bigcup_{x \in C'} x = U$

Set Cover Problems

Input : A set of numbers B such that $S_1, S_2, \dots, S_m \subseteq B$

Let: $B = \{1235, 90, 319, 457, 523\}$

In order to cover this set B , we have $\{0, 1, 2, 3, 4, 5, 7, 9\}$

Greedy Approach

Pick the set S_i with the largest number of uncovered elements and repeat this until all elements of B are covered.

The steps would be:

- First pick: 1235
covered elements: $\{1, 2, 3, 5\}$
uncovered elements: $\{0, 4, 7, 9\}$
- Second pick: 457
covered elements: $\{1, 2, 3, 4, 5, 7\}$
uncovered elements: $\{0, 9\}$
- Third pick: 90

covered elements: $\{0,1,2,3,4,5,7,9\}$

All covered by this point

Here Greedy Method is optimal.

But it is not so in all cases

For example, consider the following problem:

Let: $B = \{1, 2, 3, 4, 5, 6\}$

Let the set family would $\{1, 2, 3, 4\}, \{1, 3, 5\}, \{2, 4, 6\}$.

Here, the Greedy approach would not work as it would pick all three subsets.

However, the most optimum solution would be to pick $\{1, 3, 5\}$ and $\{2, 4, 6\}$.

Time Complexity : $O(\log n)$

If B consists of n elements with optimal cover containing k of them.

The algorithm complexity is at most $k * \ln(n)$

Proof of Approximations

Assume n_t to be the number of elements that are not covered after t iterations.

$$n_0 = n$$

We know that the optimal solution uses k sets, therefore there must be some set A , such that $|A| \geq \frac{1}{k}$ of the points.

As the algorithm decides the set which has the most points, it will at least cover that many.

After the first set of iteration, there are at the maximum $n * (1 - \frac{1}{k})$ uncovered elements left.

Continuing, since we know solution uses k sets, there will be a set covering at least a $\frac{1}{k}$ fraction of the remainder.

Since the algorithm chooses the set that covers the maximum number of uncovered elements, there are at most $n * (1 - \frac{1}{k})^2$ elements left.

Thus, after i rounds, there will be at most $n * (1 - \frac{1}{k})^i$ elements left. After $i = k * \ln(n)$ rounds

$$n * (1 - \frac{1}{k})^{k * \ln(n)} < n(1/e)^{\ln(n)} = 1$$