Approximation Algorithm

Prof. Palash Dey

Department of Computer Science and Engineering

Indian Institute of Technology, Kharagpur

Week – 01

Lecture 07

Lecture 07 : Dual Fitting Analysis of Greedy Set Cover

Welcome. So, in the last class we have seen the greedy algorithm for weighted set cover problem and we have seen that the approximation ratio is big O of log of n. There we have used one fact that if there is an optimal solution which covers every elements with total weight opt or total cost opt. then there exist that means, that the optimal solution can cover all elements with average cost opt by n, then there exist a set which also covers its elements with average cost at most opt by n. So, let us prove that formally although this may be intuitively clear, but that was only the missing part which we assumed in the last classes analysis. So, for that let us prove it for arbitrary iteration. So, we prove the claim that we that in the k-th iteration $min_{j \in [m] : \hat{S}_j \neq \varnothing} \frac{w_j}{|\hat{S}_j|}$ this is less than equal to $\frac{opt}{n_k}$, where n_k is the number of uncovered elements in the beginning of the k-th iteration. So, for that so, we need to prove that and then this will finish the proof that the approximation ratio of greedy set cover is $\log n$ in particular H_n where H_n is the n-th harmonic number. So, let O be the indices of the sets in an optimal set cover ok. Then what is $\frac{\sum w_j}{\sum \hat{S}_j}$?

The numerator is opt because we have assumed that O is the indices of an of the sets in an optimal set cover and this remains same. And, this is less than equal to $\frac{opt}{n_k}$. So, we are considering a particular iteration k. So, here we are doing this analysis for the k-th iteration. ok and this is the last inequality this one follows since $\hat{S}_j$ covers $\cup_{j \in O} \hat{S}_j$ covers all the n_k uncovered elements ok.

And this is and this is like this is greater than equal to $min_{j \in O} \frac{w_j}{|\hat{S}_j|}$ which is greater than equal to minimum over all sets such that $\hat{S}_j$ is not an empty set. $\frac{w_j}{|\hat{S}_j|}$ ok. So, recall this follows from this fact that this follows from the fact that $\frac{\sum a_i}{\sum b_i} \geq min_{i=1}^{k} \frac{a_i}{b_i}$.

And now let j the index of a set which minimizes this ratio. So, $\frac{w_j}{|\widehat{S_j}|} \leq \frac{opt}{n_k}$. If we add S_j to our solution, if we add S_j to our solution then there will be $|\widehat{S_j}|$ few are uncovered elements. $n_{k+1} = n_k - |\widehat{S_j}|$.

$w_j \leq |\widehat{S_j}| \frac{opt}{n_k}$ but $|\widehat{S_j}| = n_k - n_{k+1}$ Hence, $w_j \leq (n_k - n_{k+1}) \frac{opt}{n_k}$. So, this concludes the proof the final piece that we have assumed. So, this concludes the proof that the approximation ratio of the greedy set cover is h n, where n where h n is the nth harmonic number. Next we make this make this claim stronger. So, let g be the maximum cardinality of any input set. that is $g = max_{j=1}^{m} |S_j|$.

So, then we claim that the approximation factor of the greedy set cover is actually H_g. So, this is less than equal to n. So, we claim theorem approximation the approximation factor of greedy set cover is at most h g proof and we will prove this using a technique called dual fitting. So, we will the technique called dual fitting. So, what is it? We will assign values to dual variables in such a way that the sum of the weights of the sets picked by the algorithm which is ALG is same as dual objective.

So, the idea is we will construct an infeasible dual solution, why such that which is sum of the weights of the sets in i which are picked by the algorithm w_j is same as $\sum y_i$. And then we will show that $y_i' = \frac{1}{H_g} y_i, i \in [n]$ is a dual feasible solution and then the result will follow by weak duality. Then by weak duality we know that, $\sum w_j = \sum y_i = H_g \sum y_i'$ and this is the these are dual feasible solution y_i'. So, this is less than equal to $H_g LP - opt$ which is less than equal to $H_g opt$ ok.

So, with duality we are using here. and this technique is called dual fitting. So, to construct an the infeasible dual solution dual solution suppose we choose the set S_j in the kth iteration Then for each element that are covered in the k-th iteration which were previously uncovered. That means, for each $e_i \in \widehat{S_j}$ we set $y_i = \frac{w_j}{|\widehat{S_j}|}$ ok. So, since each e_i was uncovered before k th iteration and from $k+1$ th iteration it will be covered.

So, we observe that each dual variable is set exactly once ok. and clearly if I am picking the jth set in the kth iteration S_j in kth iteration and it is cost is w_j that cost or weight is distributed among the dual variables that are set in that iteration. So, clearly alg equal to $\sum w_j$ which is $\sum y_i$ ok. Next we scale this by H_g we divide each dual variable with by H_g and we will show that is a dual feasible solution. So, define $y_i^{'}=\frac{1}{H_g} y_i, i \in [n]$

We claim that $y_i^{'}$ is dual feasible solution that is all dual constraints are satisfied. So, that is for each set S_j we must show that for all $i \in [n]$ such that $\sum y_i \leq w_j$ ok. So, let j be any arbitrary set. S_j be any arbitrary set ok. Let a_k be the number of elements of S_j that are uncovered in the beginning of the k-th iteration. So, in particular we have $a_1=|S_j|$ and $a_{l+1}=0$. Recall we assume that the greedy algorithm iterates over l many times the while loop iterates over l many iterations ok. So, let A_k be the uncovered be the set of uncovered elements of S_j that are covered in the kth iteration that k-th iteration.

That means, $|A_k|=a_k-a_{k+1}$. So, let the greedy algorithm picks the set S_p in the k-th iteration ok. Then $e_i \in A_k$ all the variables in the jth set is in S_j which are set in the kth iteration

$$y_i^{'}=\frac{w_p}{H_g|\hat{S}_p|}\leq\frac{w_j}{H_g|\hat{S}_j|}=\frac{w_p}{H_g a_k} \quad .$$

So,

$$\sum_{i\in[n]:e_i\in S_j} y_i^{'}=\sum_{k=1}^{l}\sum_{e_i\in A_k} y_i^{'}\leq\sum_{k=1}^{l}\frac{a_k-a_{k+1}}{H_g a_k}\cdot w_j$$

$$=\frac{w_j}{H_g}\sum_{k=1}^{l}\frac{a_k-a_{k+1}}{a_k}\leq\frac{w_j}{H_g}\sum_{k=1}^{l}\left(\frac{1}{a_k}+\frac{1}{a_{k-1}}+\ldots+\frac{1}{a_{k+1}+1}\right)=\frac{w_j}{H_g}\sum_{i=1}^{|S_j|}\frac{1}{i}=\frac{w_j}{H_g}H_{|S_j|}\leq w_j$$

So, this finishes the proof. So, the approximation guarantee is far superior than H_n if all the sets are of small cardinality in the input.

So, let us stop here. Thank you.