CS265, Winter 2022

Class 11: Agenda, Questions, and Links

1 Announcements

- HW5 was due today, HW6 due next Wednesday!

2 Recap/Questions?

Any questions from the minilectures and/or the quiz (second moment method and LLL)?

3 Practice with the LLL

Recall the *k-SAT* problem. There are n variables $x_1, \ldots, x_n$. We consider clauses that looks like $(x_{i_1} \vee x_{i_2} \vee \overline{x_{i_3}} \vee \cdots \vee x_{i_k})$; that is, a clause is the OR of k literals. **For today, assume that each clause has k distinct variables that appear in it.** We have a formula φ that is the AND of m clauses. We would like to know: is φ satisfiable? That is, is there a way to assign values to the variables $x_1, x_2, \ldots$ so that φ evaluates to TRUE?

Group Work

Suppose that each variable x_i is in at most t clauses, for some parameter t that will depend on k and that you'll work out in this problem. Apply the LLL to get a statement like the following:

> Suppose that each variable is in at most t clauses of φ. Then φ is satisfiable.

(You should try to get t to be as large as possible. It's not hard to see that the statement above is true if, say, $t = 1$, but you should get a value of t that grows with k.)

***Hint**: Recall that to apply the LLL, you need to define a probability distribution and a set of "bad" events. We set up this example in the minilecture video, we just didn't work out the conclusion. In the set-up of the video, we considered the probability distribution to correspond to assigning TRUE/FALSE to each variable $x_1, \ldots, x_n$ independently with probability $1/2$ each, and we defined the bad event A_i to be the event that clause i is not satisfied.*

Group Work: Solutions

We claim that if each variable is in at most $t \leq 2^{k-2}k$ clauses, then the formula is satisfiable. To see this, consider a uniformly random assignment to $x_1, \ldots x_n$ (ie setting each

x_i to be TRUE/FALSE independently with probability 1/2). Define events $A_1, \ldots, A_m$ where A_i is the indicator random variable of the ith clause NOT being satisfied. For a clause with k variables to not be satisfied, all k variables must take the "bad" assignment, and hence:

$$\Pr[A_i] = 1/2^k.$$

To apply the LLL, we now need to reason about the dependencies. To that end, we claim that A_i is mutually independent of the set of clauses whose variable are disjoint from the variables in clause i, namely the set

$$S_i = \{A_j : vbl(clause_i) \cap vbl(clause_j) = \emptyset\}.$$

Indeed, no matter the assignment to variables that occur in the clauses in S_i, since none appear in the ith clause, they can't alter the probability of A_i. Now we simply count up the number of events not in the set S_i: namely

$$[\#j \text{ such that } vbl(clause_i) \cap vbl(clause_j) \neq \emptyset] \leq kt,$$

since there are k variables in clause i, and each of them is in at most t other clauses.

To conclude, each A_i is mutually independent of all but $d = kt$ other events, and hence by the LLL with $d = kt$ and $p = 2^{-k}$, we have that $\Pr[\cap_i not(A_i] \geq (1-2p)^m > 0$ provided $dp \leq 1/4$, hence we want $kt \cdot 2^{-k} \leq 1/4$, which implies that want $t \leq 2^{k-2}/k$.

To put some concrete numbers in here, if $k = 10$, then as long as each variable appears in at most $2^{10-8}/8 = 25.6$ clauses, then the formula is always satisfiable, no matter the number of variables of clauses!!! Of course, now the big question on your mind should be "*Its great that we know such formulas are satisfiable, but how do we FIND a satisfying assignment efficiently?*" We'll get to this in the next set of minilectures, on the "Constructive LLL"!!

3.1 More Practice with LLL and Mutual Independence

Consider a set of equations over variables $x_1, \ldots, x_n$, where each equation has the form $a_1 x_{i_1} + a_2 x_{i_2} + \ldots + a_r x_{i_r} \equiv a_{r+1} \mod 17$, for some r (that might vary from equation to equation) and set of coefficients $a_1, \ldots, a_r \in \{1, 2, \ldots, 16\}$, and $a_{r+1} \in \{0, \ldots, 16\}$. Additionally, suppose that each variable, x_i, occurs in at most 4 equations.

Group Work

Prove that there exists an assignment to the variables such that *none* of the equations are satisfied.

Hint: *Recall that because 17 is prime, for any* $a \in \{1, \ldots, 16\}$ *and any* $b \in \{0, \ldots, 16\}$, *the equation* $ax \equiv b$ *has a unique solution for* $x \in \{0, \ldots, 16\}$.

Hint: *It might be helpful to go back to the definition of mutual independence when* *arguing about the value of d when applying the LLL.*

Group Work: Solutions

Consider the distribution given by assigning each variable x_i a uniformly random value in the set $\{0, \ldots, 16\}$. Define the event A_i to be the indicator random variable representing whether the ith equation is satisfied. First, we claim that $\Pr[A_i] = 1/17$. To see this, consider an equation involving r variables, and note that whatever the assignment to the first $r-1$ variables, over the randomness of assigning the rth variable, the probability the equation holds is 1/17. This is true because after the first $r-1$ variables are set, the equation becomes $a_r x_r \equiv a' \mod 17$, for some $a_r \neq 0$, and some value of a', and this equation has a unique solution modulo 17 (because 17 is prime...).

We now argue that each A_i is mutually independent of all but 4 other equations. This might seem counterintuitive, because there could be a large number of other equations sharing variables with each A_i. The argument will critically leverage the definition of mutual independence. Consider one of the equations: $a_1 x_{i_1} + a_2 x_{i_2} + \ldots + a_r x_{i_r} \equiv a_{r+1} \mod 17$, and consider the set of equations S that do not include variable x_{i_1}. [By assumption, there are at most 4 equations not in set S.] We claim that the event that the equation in question is satisfied is mutually independent from all the events corresponding to equations in set S. To see why this holds, recall that for any fixed assignment to all variables except x_{i_1}, with respect to randomly assigning x_{i_1}, the probability the equation in question is satisfied is exactly 1/17. Because the equations in set S do not involve x_{i_1}, conditioning on any outcomes for those equations does not change the distribution of x_{i_1}, and hence the probability the equation is satisfied is still exactly 1/17 even when conditioning on any subset of outcomes of equations in S. This is precisely the definition of mutual independence.

To apply the LLL, we are hoping that $d \Pr[A_i] < 1/4$, which holds because $d = 4$, and $\Pr[A_i] = 1/17$.

4 Practice with the second moment method

In a graph $G = (V, E)$, say that a vertex v is **isolated** if it has no neighboring vertices.

Group Work

Let $G \sim G_{n,p}$ be a random graph where each edge is present independently with probability p, where $p = \frac{c \ln n}{n}$ for some constant $0 < c < 1$.

1. Use the Second Moment Method to show that, with probability at least $1 - o(1)$, there is some isolated vertex in G.

   For this exercise, feel free to use the approximation $e^{-x} \approx 1 - x$ when x is small without worrying about it.

Hint: *Consider the random variable X that is the number of isolated vertices in G, and recall that the second moment method says that* $\Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{(\mathbb{E}X)^2}$.

Hint: *When computing the variance of X, you may want to consider the following question: given two distinct vertices u, v of G, what is the probability that* both *u and v are isolated?*

2. If you finish the previous part, what statement can you make about the case that $c > 1$?

Group Work: Solutions

1. First, lets make sure that the expected number of isolated vertices is reasonably large for $p = \frac{c \log n}{n}$ with $c < 1$. We do this by linearity of expectation:

$$\mathbb{E}\left[\# \text{ isolated vertices }\right] = n \Pr[v \text{ is isolated}] = n(1-p)^{n-1} \approx ne^{-p(n-1)} = n \cdot n^{-c(1-1/n)} \approx n^{1-c}.$$

When $c < 1$, $n^{1-c} \gg 1$, things are looking good in expectation.

To apply the second moment method, we need a bound on the variance of the number of isolated vertices. Letting X denote the number of isolated vertices, we have the following:

$$\begin{aligned} Var[X] &= \mathbb{E}\left[X^2\right] - \left(\mathbb{E}\left[X\right]\right)^2 \\ &= \mathbb{E}\left[\sum_{u,v} \Pr[u \text{ and } v \text{ isolated}\right] - \left(\mathbb{E}\left[X\right]\right)^2 \\ &= \mathbb{E}\left[\sum_{u} \Pr[u \text{ and } u \text{ isolated}\right] + \mathbb{E}\left[\sum_{u \neq v} \Pr[u \text{ and } v \text{ isolated}\right] \left(\mathbb{E}\left[X\right]\right)^2 \\ &\approx n^{1-c} + n(n-1)(1-p)^{2n-3} - \left(\mathbb{E}\left[X\right]\right)^2. \end{aligned}$$

In the above, we used the fact that for each of the $n(n-1)$ choices of $u \neq v$, for both to be isolated, all of the $2n - 3$ potential edges connected to either of them must be absent. Now we just simplify: $(1-p)^{2n-3} \approx e^{-(2n-3)p} = n^{-c(2-3/n)} \approx n^{-2c}$.

$$n^{1-c} + n(n-1)(1-p)^{2n-3} - \left(\mathbb{E}\left[X\right]\right)^2 \approx n^{1-c} + n^{2(1-c)} - \left(n^{1-c}\right)^2 = n^{1-c}.$$

So, to conclude, we have

$$\Pr[X = 0] \leq \frac{Var[X]}{(\mathbb{E}\left[X\right])^2} \approx \frac{n^{1-c}}{n^{2(1-c)}} = n^{c-1}.$$

Hence for $c < 1$, this probability is $o(1)$.

2. In the case that $c > 1$, the expected number of isolated vertices is roughly $n^{1-c} \ll 1$, and hence by Markov's inequality, the probability that there are any isolated vertices (ie the probability that the number of isolated vertices is at least 1), is bounded by $\frac{\mathbb{E}[X]}{1} \approx n^{1-c} \ll 1$.