

CSE 422 Autumn 26 HW 1

Note: For all problems, you will be expected to explain your reasoning. Providing an expression without reasoning will not receive any partial credit.

1 Problem 1: Optimal Homework Strategies

Consider a purely hypothetical class that has n homework assignments, where each homework assignment consists of 3 problems. Suppose that homework is graded in the following non-standard manner: for each assignment, one uniformly random problem is selected, and the student must write the solution to that problem during class without external aid, and the score of that solution is the grade the student receives for that homework.

Consider the following, purely hypothetical, model for a student (call him “Jerry”), whose objective is to achieve as high expected value on these n assignments as possible, with as little work as possible. We will model Jerry as follows. For some parameters p, q and parameter $k \in \{0, \dots, 3\}$, we will assume that Jerry solves k of the problems each week, and doesn’t solve the remaining $3 - k$ problems. On the problems he solves, he will get p points, and on the problems he doesn’t solve, he will get q points, and let’s assume that $p > q$.

(a) [5 points] As a function of p, q , and k , what is the expected number of points that Jerry should expect to get on all n of the homework problems?

(b) [6 points] Luckily for Jerry, the professor is super nice, and also allows students the drop their lowest grade out of the n , so their total score is actually the sum of their highest $n - 1$ grades. Now, as a function of p, q , and k , how many points does Jerry expect to get? How does this expression change if instead we drop the 2 lowest grades? Provide as simple an expression as you can.

(c) [9 points] What is the variance of Jerry’s score under this “best-of- $(n - 1)$ ” system? How does this change if Jerry’s score is no longer deterministic, conditioned on whether or not he did the homework, but random?

2 Problem 2: Approximate matrix multiplication

Our goal for this problem is to develop a fast algorithm for (approximate) matrix multiplication. Formally, the goal of this algorithm is to, given two matrices $A, B \in \mathbb{R}^{n \times n}$, output a matrix M so that $M \approx AB$ as efficiently as possible. If we want to do this exactly, then the best known runtime of this is $O(n^\omega)$, where $\omega < 2.371177$ is the *matrix multiplication constant*.

Our goal is to beat this, if we allow some error. For the sake of normalization, we will assume all the matrices will have bounded *Frobenius* norm; this is just the ℓ_2 norm of the matrix if we think of it as a vector:

$$\|M\|_F = \sqrt{\sum_{i,j} M_{ij}^2}.$$

Concretely, let us assume that $\|A\|_F = \|B\|_F = 1$, and also let $a_1, \dots, a_n \in \mathbb{R}^n$ denote the columns of A , and $b_1^\top, \dots, b_n^\top$ denote the rows of B , so that

$$A = \begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix}, \quad B = \begin{bmatrix} - & b_1^\top & - \\ & \vdots & \\ - & b_n^\top & - \end{bmatrix}, \quad AB = \sum_{i=1}^n a_i b_i^\top.$$

Let us further assume that $\|a_i\|_2^2 \leq C/n$ for all $i = 1, \dots, n$, for some parameter C . We will consider the following algorithm: for some parameter $p \in [0, 1]$, form a set $J \subseteq [n]$ by including each element $i \in J$ with probability p independently, and approximating $AB \approx X$, where

$$X = c_p \sum_{i \in J} a_i b_i^\top,$$

for some parameter c_p .

(a) [2 points] Compute the expectation of X . As a function of p , what is the value of c_p that make this an unbiased estimator of AB ?

(b) [3 points] Give the best bound you can on the expected runtime of computing X , as a function of p and ω .

(c) [5 points] Give a bound, in terms of C , on the variance of X as an estimator of AB , where here by variance we mean

$$\mathbb{E}[\|X - AB\|_F^2].$$

(d) [5 points] Putting it all together, for parameters $\varepsilon, \delta > 0$, by choosing internal parameters carefully, derive the runtime that this algorithm obtains to achieve an estimate X of AB so that

$$\Pr[\|X - AB\|_F \geq \varepsilon] \leq \delta.$$

(e) [5 points] Suppose now you don't have the assumption that $\|a_i\|_2^2 \leq C/n$ for all $i = 1, \dots, n$. How does the previous runtime bound change? How might you change the algorithm to get around this issue?

3 Problem 3: The moment method

In this problem, we will re-derive and generalize the standard Chernoff-Hoeffding bound, a very important tool in probability theory and in the analysis of randomized algorithms. Recall that (one version of) the standard Chernoff-Hoeffding bound is the following:

Theorem 3.1. *Let $X_1, \dots, X_n$ be independent, identically distributed mean-zero random variables satisfying $|X_i| \leq 1$ for all i . There is some universal constant $C > 0$ so that*

$$\Pr \left[\left| \frac{1}{n} \sum_{i=1}^n X_i \right| \geq t \right] \leq 2 \cdot \exp(-Cnt^2) . \quad (1)$$

In fact, we seek to prove something similar under a weaker assumption: namely, rather than assuming that the X_i 's are bounded random variables, we will simply assume that there is some even integer parameter k so that $\mathbb{E}[|X_i|^\ell] \leq 1$ for all positive integers $\ell \leq k$.

Our proof strategy will be to compute an upper bound on

$$\mathbb{E} \left[\left(\frac{1}{n} \sum_{i=1}^n X_i \right)^k \right] ; \quad (2)$$

analysis of such expressions is a standard technique and is known as the *moment method*.

We will just expand this out into a bajillion terms:

$$\mathbb{E} \left[\left(\frac{1}{n} \sum_{i=1}^n X_i \right)^k \right] = \frac{1}{n^k} \mathbb{E} \left[\underbrace{\left(\sum_{i=1}^n X_i \right) \cdots \left(\sum_{i=1}^n X_i \right)}_{k \text{ times}} \right] = \frac{1}{n^k} \sum \mathbb{E} \left[\prod_{j=1}^k X_{i_j} \right] ,$$

where the sum in the last expression is over all $i_1, \dots, i_k \in \{1, \dots, n\}$.

(a) [4 points] You should notice that some (in fact, most) of the terms in this final expression will contribute 0 to the expectation. Can you characterize which terms survive this expansion?

(c) [2 points] Derive a uniform upper bound on how big the expectation of any single term can be in this expansion.

(b) [6 points] Show that the total number of terms that contribute a non-zero amount to this expectation is at most $(Ank)^{k/2}$ for some constant $A \geq 1$. In fact you can show $A = 1$, but you do not need to.

(d) [4 points] Putting (b) and (c) together, derive an upper bound on the original expression in Equation 2. Then, combining this with Markov's inequality, derive an upper bound on the left-hand side of Equation 1 as a function of k .

(e) [4 points] Finally, by optimizing the choice of k (as a function of n and t), find the smallest value of k (up to constants) so that you can recover Theorem 3.1. Note that this directly generalizes the original theorem!