COMPSCI 614: Problem Set 3

Due: 3/26 by 11:59pm in Gradescope.

Instructions:

- You are allowed to, and highly encouraged to, work on this problem set in a group of up to three members.
- Each group should **submit a single solution set**: one member should upload a pdf to Gradescope, marking the other members as part of their group in Gradescope.
- You may talk to members of other groups at a high level about the problems but **not work through the solutions in detail together**.
- You must show your work/derive any answers as part of the solutions to receive full credit.

Hint: The following two inequalities may be helpful throughout the course: for any $x > 0$, $(1+x)^{1/x} \leq e$ and $(1-x)^{1/x} \leq 1/e$.

1. Error Bounds for Count Sketch (14 points)

In class we showed that Count Sketch implemented with sketch size $O(\log(1/\delta)/\epsilon^2)$ returns an estimate for any entry of a vector $x \in \mathbb{R}^n$ satisfying $|\tilde{x}(i) - x(i)| \leq \epsilon \|x\|_2$ with probability $\geq 1 - \delta$.

1. (6 points) For any $x \in \mathbb{R}^n$ and integer $k \leq n$, let x_k be its best k-sparse approximation – i.e., x_k matches x on just its k largest magnitude entries and is 0 everywhere else. Prove that Count Sketch implemented with sketch size $O(\log(1/\delta)/\epsilon^2)$ returns an estimate for any entry of a vector $x \in \mathbb{R}^n$ satisfying $|\tilde{x}(i) - x(i)| \leq \epsilon \|x - x_{1/\epsilon^2}\|_2$ with probability $\geq 1 - \delta$. **Hint:** Consider a single repetition. Argue that with good probability, any index i is not hashed to the same position as any of the $1/\epsilon^2$ largest magnitude entries. Then compute an error bound conditioned on this event.

2. (2 points) Describe a scenario in which you think that the error bound above will be much better than the error bound shown in class.

3. (4 points) Prove that Count Sketch with sketch size $O(\log(1/\delta)/\epsilon)$ (note the $1/\epsilon$ rather than $1/\epsilon^2$) dependence) returns an estimate for any entry of a vector $x \in \mathbb{R}^n$ satisfying $|\tilde{x}(i) - x(i)| \leq \epsilon \|x\|_1$ with probability $\geq 1 - \delta$, where $\|\cdot\|_1$ is the ℓ_1 norm. **Hint:** Replace Chebyshev's inequality by Markov's inequality in the Count Sketch analysis.

4. (2 points) How does the bound in part (3) compare to the bound shown in class? That is, when roughly do you expect one to be better than the other?

2. One-Sided Importance Sampling for Matrix Multiplication (5 points)

1. (3 points) Suppose we run the approximate matrix multiplication algorithm described in class with $p_i = \frac{\|A_{:,i}\|_2^2}{\|A\|_F^2}$ for all $i \in \{1, \ldots, n\}$. I.e., our sampling probabilities only take into account the row norms of A, and not the column norms of B. Prove that if we take $t = O\left(\frac{1}{\delta \cdot \epsilon^2}\right)$ samples then with probability at least $1 - \delta$, $\|AB - \overline{\mathbf{C}}\|_F \leq \epsilon \|A\|_F \cdot \|B\|_F$. I.e., up to constant factors on the worst-case error bound, this approach matches the optimal approach analyzed in class.

2. (2 points) Consider the following application: I have an orthonormal basis $V \in \mathbb{R}^{n \times k}$ and I would like to repeatedly apply the projection operation $V^T \in \mathbb{R}^{k \times n}$ to various input vectors $x \in \mathbb{R}^n$. What is the runtime required to compute $V^T x$ for a single $x \in \mathbb{R}^n$ exactly? What is the runtime required to approximate $V^T x$ to error $\epsilon \|x\|_2$ with probability $\geq 1 - \delta$ using the sampling approach of part (1)? **Note:** You may assume that you precompute a sample of V's rows that is repeatedly used for different x, without factoring this precomputation step into your runtime. You may also use that the approximate matrix multiplication result generalizes to rectangular matrix multiplication, via an identical proof.

3. Approximate Matrix Multiplication with Random Sketching (9 points)

1. (6 points) Let $\mathbf{\Pi} \in \mathbb{R}^{n \times m}$ be a random matrix with each entry set independently to $1/\sqrt{m}$ with probability 1/2 and $-1/\sqrt{m}$ with probability 1/2. Show that for any $A, B \in \mathbb{R}^{n \times n}$, $\mathbb{E}[\|A\mathbf{\Pi}\mathbf{\Pi}^T B - AB\|_F^2] \leq \frac{2\|A\|_F^2 \|B\|_F^2}{m}$.

   Hint: Write $A\mathbf{\Pi}\mathbf{\Pi}^T B = \frac{1}{m}\sum_{i=1}^m A\boldsymbol{\pi}_i \boldsymbol{\pi}_i^T B$, where $\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_m \in \mathbb{R}^n$ are independent random vectors with each entry set independently to 1 with probability 1/2 and -1 with probability 1/2. Analyze the expectation and variance of the entries of $A\boldsymbol{\pi}_i \boldsymbol{\pi}_i^T B$.

2. (1 point) Show that if $m = O(\frac{1}{\delta \epsilon^2})$, then with probability at least $1 - \delta$, $\|A\mathbf{\Pi}\mathbf{\Pi}^T B - AB\|_F \leq \epsilon \|A\|_F \|B\|_F$.

3. (2 points) How does this approximate matrix multiplication algorithm compare to the sampling based algorithm presented in class in terms of runtime? How about in terms of suitability to implementation in a data stream where the entries of A and B are updated over time and you want to maintain a small compression of these matrices?

4. Importance Sampling with Machine Learned Advice (8 points)

Consider the following scenario: a group of scientists have set up a camera at a fish ladder that automatically takes images when fish swim by. They would like to count the number of images that contain a certain fish species – let's say salmon. Assume there are n images and m of them contain a salmon. The goal is to compute m. Since it is time consuming to hand-check every image for the presence of a salmon, they would like to use random sampling to get a good estimate of m. Throughout, we will think of m as being much smaller than n.

1. (2 points) Consider the following estimator: sample t images uniformly at random with replacement from the set of images. Count how many contain a salmon and multiply this count by n/t. Let $\tilde{\mathbf{m}}$ be the returned estimator. Prove that $\mathbb{E}[\tilde{\mathbf{m}}] = m$ and that $\text{Var}[\tilde{\mathbf{m}}] \leq \frac{mn}{t}$.

2. (6 points) Assume the scientists develop a machine learning model that is able to quickly process all their images and detect the presence of salmon. The model is not exact, but is

fairly accurate: it correctly detects 90% of images that contain salmon, while missing 10% of these images. Overall, it predicts the presence of salmon in 5% of all images. Describe an algorithm that uses this model to give an estimator $\tilde{\mathbf{m}}$ for m that is correct in expectation, requires just t hand image checks, and has variance at most $\frac{.28 \cdot mn}{t}$. That is, the estimator reduces the variance by a factor of roughly three as compared to the naive uniform sampling estimator. **Hint:** Apply importance sampling where the probability of sampling an image depends on whether or not the ML model predicts that image contains a salmon or not. You may assume that the accuracy statistics of the model are known and can be used in setting these sampling probabilities.