Lecture 8: Home Assignment, Due Dec 31st

Lecturer: Edo Liberty

Warning: *This note may contain typos and other inaccuracies which are usually discussed during class. Please do not cite this note as a reliable source. If you find mistakes, please inform me.*

1 Randomized meta-algorithms

setup

In this question we assume the common case where we have an input $x \in X$ and we wish to approximate a function $f : X \to \mathbb{R}^+$ (i.e. $\forall x \; f(x) \geq 0$). For that we have a black box randomized algorithm $A : X \to \mathbb{R}^+$ such that $\mathbb{E}[A(x)] = f(x)$. The questions ask you to designing meta algorithms using A as a black box.

question

1. Show that
$$\Pr[A(x) \geq 3f(x)] \leq \frac{1}{3}$$

2. Assume that for all x we have that $\mathrm{Var}[A(x)] \leq c \cdot [f(x)]^2$. Describe an algorithm B_2 such that for any two constants $\varepsilon, \delta > 0$:
$$\Pr[|B_2(x) - f(x)| \geq \varepsilon f(x)] \leq \delta$$

3. Assume that $\Pr[|A(x) - f(x)| \leq t|] \geq \frac{1}{2} + \eta$ for some fixed value $\eta > 0$. In words, the algorithm gets an additive approximation t with probability slightly better than 1/2. (Here we do not assume anything on the variance of $A(x)$). Design and algorithm B_3 such that for any prescribed $\delta > 0$

$$\Pr[|B_3(x) - f(x)| \leq t|] \geq 1 - \delta$$

That means the algorithm achieves the same additive approximation with probability arbitrary close to one.

2 Set intersections

setup

We have a universe of N items $A = \{a_1, \ldots, a_N\}$ and m subsets $S_i \subset A$, $i \in \{1, \ldots, m\}$. We assume that given a set S_i we can iterate over its elements one by one. The exercise will deal with approximating the size of different unions of these sets. Here you are tasked with designing an algorithm. Your algorithm is allowed to preprocess the sets S in and produce data structures ($preprocess(S)$) It should then be able to take as input a set of indexed $I \subset \{1, \ldots, m\}$ and produce an ε approximation to $|\cup_{i \in I} S_i|$ with probability at least $1 - \delta$ ($estimateUnionSize(I)$). The aim is to create an algorithm which runs in time $o(\sum_{i \in I} |S_i|)$ and requires $o(|\cup_{i \in I} S_i|)$ space. That means that simply iterating through the lists and keeping items in a hash lookup table is not an adequate solution.

1. Describe $preprocess(S)$ which is the preparatory stage of the algorithm and results in our choice of data structures.

2. Describe $estimateUnionSize(I)$ which return an ε approximation to $|\cup_{i \in I} S_i|$ with probability $1 - \delta$.

3. Prove your algorithm's correctness.

4. What is the space usage of your data structures?

5. What is the runtime complexity of $estimateUnionSize(I)$?

3 Weak random projections

setup

In this question we will construct a simple and weak proof of the Johnson-Lindenstrauss lemma. Given two vectors $x, y \in \mathbb{R}^d$ we will find two new vectors $x', y' \in \mathbb{R}^k$ such that from x' and y' we could approximate the value of $||x - y||$. The idea is to define k vectors $r_i \in \mathbb{R}^d$ such that each $r_i(j)$ takes a value in $\{+1, -1\}$ uniformly at random. Setting $x'(i) = r_i^T x$ and $y'(i) = r_i^T y$ the questions will lead you through arguing that $\frac{1}{k}||x' - y'||_2^2 \approx ||x - y||_2^2$.

questions

1. Let $z = x - y$, and $z' = x' - y'$. Show that $z'(\ell) = r_\ell^T z$ for any index $\ell \in [1, \ldots, k]$.

2. Show that $E[\frac{1}{k}||z'||_2^2] = E[(z'(\ell))^2] = ||z||_2^2$.

3. Show that
$$\text{Var}[(z'(\ell))^2] \leq 4||z||_2^4.$$

   Hint: for any vector w we have $||w||_4 \leq ||w||_2$.

4. From 3 (even if you did not manage to show it) claim that

$$\text{Var}[\frac{1}{k}||z'||_2^2] \leq 4||z||_2^4/k.$$

5. Use 3 and Chebyshev's inequality do obtain a value for k for which:

$$(1 - \varepsilon)||x - y||_2^2 \leq \frac{1}{k}||x' - y'||_2^2 \leq (1 + \varepsilon)||x - y||_2^2$$

   with probability at least $1 - \delta$.

4 SVD and the power method

setup

Here we will prove some basic facts about singular values, matrices, and the power method. For the reminder of the question we assume $A \in \mathbb{R}^{m \times n}$ is an arbitrary matrix. For convenience and w.l.o.g. assume $m \leq n$ and denote by $\sigma_1 \geq \ldots \sigma_m \geq 0$ the singular values of A. Define the numeric rank of a matrix $\rho(A)$ to be $\rho(A) = \|A\|_{fro}^2 / \|A\|_2^2$. $\rho(A)$ is a smoothed version of the algebraic rank $rank(A)$. It is always true that $1 \leq \rho(A) \leq Rank(A) \leq \min(m, n)$. If $\rho(A) \leq 1 + \varepsilon$ for a sufficiently small ε the matrix is "close" to being of rank 1.

question

1. Let $P \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ be unitary matrices. Show that $\|PAQ\|_{fro} = \|A\|_{fro}$. Hint, begin with the case where one of the matrices P or Q are the identity matrix.

2. Using the above show that for any matrix A we have that

$$\|A\|_{fro} = \sqrt{\sum_{i=1}^{m} \sigma_i^2}.$$

   It might help you to show that $\|A\|_{fro}^2 = tr(AA^T)$ where $tr(\cdot)$ stands for the matrix trace.

3. Give an expression to the numeric rank of A only in terms of its singular values σ_i.

4. Express the numerical rank of $(AA^T)^k A$ only in terms of σ_i.

5. Assume that the matrix A is such that $\sigma_2 / \sigma_1 \leq \eta$ for some $\eta < 1$. Use your expressions from above to find k such that $\rho((AA^T)^k A)) \leq 1 + \varepsilon$. How does this relate to the the Power Method for computing the largest singular value and vectors of A?