Assignment 4

CS-UH-2218: Algorithmic Foundations of Data Science

Assignments are to be submitted in groups of two or three. *Upload the solutions on NYU classes as one PDF file for theoretical assignments and separate source code files for each programming assignment. Submit only one copy per group. Clearly mention the participant names on* ***each file*** *you submit.*

Problem 1 (10 points)**.**
Suppose that we wish use Bloom Filtering to store a set of a million items. What is the minimum amount of memory (in terms of bits) do we need in order to have a 10^{-6} probability of false positives? You can use the estimate for false positives obtained in class.

Problem 2 (10 points)**.**
Suppose that we run the Misra-Gries algorithm on a stream of length m with $k-1$ counters. Let $\hat{f}_x$ be the count returned by the algorithm for a key x. If there is no counter associated with x, we take $\hat{f}_x$ to be 0. Let $\hat{m}$ be the sum of all counters at the end of the algorithm. Prove that for any element x, $\hat{f}_x$ provides a crude estimate of the frequency f_x of x in the following sense:

$$f_x - \frac{(m-\hat{m})}{k} \leq \hat{f}_x \leq f_x$$

Hint: Recall that every time we decrement the counters in the algorithm, we "destroy" the records of k items in the stream. How is $(m-\hat{m})$ related to the number of records that got destroyed? What can you say about the number of decrement steps?

Problem 3 (10 points)**.**
Suppose that given a stream of length m, we want to output all elements with frequency more than m/k but we do not want to output any element with frequency less than $(1-\epsilon)\, m/k$ for some given $\epsilon \in (0,1]$. How would you use the Misra-Gries algorithm to do this in one pass? How many counters (in terms of k and ϵ) do you need ?

Problem 4 (**Optional.** 10 bonus points)**.**
Let A be an $m \times u$ matrix with entries in $\{0,1\}$ s.t. for each $i, j > 1$, $A_{i,j} = A_{i-1,j-1}$ and let b be a vector in $\{0,1\}^m$. For any choice of such A and b, define the function $h_{A,b}(x) = Ax + b \pmod 2$ and let $\mathcal{H}$ be the set of all such functions. Prove that $\mathcal{H}$ is 2-universal.