CSCI B609 – Foundations of Data Science **Fall 2017**

Homework 4: December 04

Name: YOUR NAME HERE *Due: Monday, December 11, 11:59pm EST*

Problem 4.1 (Sparse recovery) *In this problem all vectors are in $\mathbb{R}^n$. Recall that the sparse recovery error is defined as:*

$$Err^k(f) = \min_{g: \|g\|_0 = k} \|f - g\|_1.$$

Give a formal argument that shows that $Err^k(f) = \sum_{i \notin S} |f_i|$ where S is the set of indices of k largest (by absolute value) entries of f.

Problem 4.2 (Dyadic intervals) *Let n be a power of two. Consider the following family of partitions of the interval $1, \dots, n$ into intervals:*

$$\begin{aligned}
&I_0 = \{\{1\}, \{2\} \dots, \{n\}\}\} \\
&I_1 = \{\{1,2\}, \{3,4\}, \{5,6\}, \dots, \{n-1, n\}\} \\
&I_2 = \{\{1,2,3,4\}, \{5,6,7,8\}, \dots, n-3, n-2, n-1, n\} \\
\dots& \\
&I_{\log n} = \{\{1, \dots, n\}\},
\end{aligned}$$

where the partition I_k consists of intervals of length 2^k. Show that any subinterval $i, \dots, j$ where $i \le j$ can be represented as a disjoint union of at most $2 \log n$ intervals from the above family.

Problem 4.3 (Generating uniform distribution) *Given a stream of numbers $a_1, \dots, a_n$ where each number is an integer between 1 and m design an algorithm that scans the stream and at every point during the scan maintains a uniformly at random chosen sample of k numbers from the stream. Your algorithm should use space $O(k \log m)$.*

Problem 4.4 (Bipartiteness via connectivity) *Consider the following reduction: given a connected undirected graph $G(V, E)$ construct a new graph $G'(V_1 \cup V_2, E')$ where V_1 and V_2 are copies of V and for each edge $(u, v) \in E$ we create two edges (u_1, v_2) and (u_2, v_1) where u_i and v_i are copies of u and v in V_i. Prove the following two statements:*

1. *If G is bipartite then the number of connected components in G' equals 2.*
2. *If G is non-bipartite then G' is connected.*