Problem Set 5

Electronic submission via Gradescope, due 11am Thursday 10/29.

You are strongly encouraged to submit a homework with a partner—that is, submit one homework with both of your names.

YOU MUST NOT RE-USE THE SAME PARTNER AS AN EARLIER PROBLEM SET!!!!

[You may discuss these problems with classmates. Please do not troll the internet looking for solutions to these problems. You must understand everything that is written on the homework you submit. We suggest that you write solutions using LaTeX.]

1. In class we saw a very simple probabilistic argument that in any graph \((V, E)\), the max-cut cuts at least \(|E|/2\) edges. Here we consider two improvements/variants:

   (a) Consider the greedy algorithm that, given an ordering \(v_1, \ldots, v_n\) of the vertices, assigns \(v_1\) to set \(A\), then greedily partitions the other vertices (by sequentially assigning each unassigned vertex \(v\) to either \(A\) or \(B\) according to whether \(v\) has more neighbors already assigned to \(B\) or more neighbors already assigned to \(A\).) Assume that ties are broken by assigning the point to set \(A\). Prove that the cut found by this greedy algorithm cuts at least \(|E| + |B|/2\) edges, where \(|B|\) is the size of set \(B\) at the end of the algorithm. [Hint: think about what happens when a vertex actually gets put in set \(B\).]

   (b) Prove that if the graph has an even number of nodes, \(n\), then there exists a cut with at least \(|E|/2 + n/2\) edges crossing the cut. [Hint: the fact that that \(n\) is even is just meant to suggest that it can be divided into two equal sets....]

   (c) Double Bonus: Can you turn the above probabilistic argument into an efficient deterministic algorithm? [This problem will not be graded...it is simply food for thought.]

2. An independent set in a graph is a subset of vertices with the property that no pair of them are connected via an edge. Given a graph \(G = (V, E)\) with \(|V| = n\), prove that \(G\) has an independent set of size at least

   \[
   \sum_{i=1}^{n} \frac{1}{d_i + 1},
   \]

   where \(d_i\) is the degree of vertex \(v_i\). [Hint: Given a permutation \(\pi = v_{\pi(1)}, \ldots, v_{\pi(n)}\) of the vertices, define the corresponding (independent) set \(S(\pi)\) by including vertex \(v_{\pi(i)}\) if none of \(v_{\pi(1)}, \ldots, v_{\pi(i-1)}\) is a neighbor of \(v_{\pi(i)}\). Note that \(S(\pi)\) is always an independent set. Now, define a randomized scheme for generating a permutation \(\pi\), such that \(\mathbb{E}[|S(\pi)|]\) is the claimed quantity above.]

**Theorem 1.** (Johnson and Lindenstrauss, 84’) Given any \(\epsilon \in (0, 1)\), and a set \(X \subset \mathbb{R}^k\) with \(|X| = n\), there exists a randomized linear map \(f : \mathbb{R}^k \to \mathbb{R}^d\) with \(d = O\left(\frac{\log n}{\epsilon^2}\right)\) that embeds \((X, \ell_2)\) into \((\mathbb{R}^d, \ell_2)\) so that with high probability

   \[
   (1 - \epsilon)\|x - y\|_2 \leq \|f(x) - f(y)\|_2 \leq (1 + \epsilon)\|x - y\|_2
   \]

   for any \(x, y \in X\).
3. In Tuesday’s class, we saw a proof of the above theorem by letting $f(x) = Ax$, where each entry of $A$ is independently drawn from $N(0, 1/d)$. In this problem, we will show that we can also select each entry independently from a Bernoulli-like distribution: $\Pr(A_{ij} = s) = \Pr(A_{ij} = -s) = 1/2$. Being able to select $A$ in this fashion significantly speeds up the generation of the matrix, as well as reducing the time to compute $Ax$.

(a) What should the value of $s$ be, as a function of $k$, $n$, or $d$? [Hint: the expected value of $\|f(x) - f(y)\|^2_2$ should be $\|x - y\|^2_2$.]

(b) Can we assume vector $v = x - y$ to be a basis vector of $\mathbb{R}^k$, as we did in class for the Gaussian case? Why or why not?

(c) Assume $\|v\|_2 = 1$. For $i = 1, 2, \ldots, k$, let $Z_i$ be the independent random variable such that $\Pr(Z_i = 1) = \Pr(Z_i = -1) = 1/2$, and let $C = \sum_{i=1}^k v_i Z_i$. Show that for any $0 < t < \frac{1}{2}$,

$$E[e^{tC^2}] \leq 1 + t + \frac{4t^2}{1 - 2t}$$

(d) Using the previous part, prove that with probability approaching 1 as $n$ approaches infinity,

$$\|f(x) - f(y)\|^2_2 \leq (1 + \epsilon)\|x - y\|^2_2$$

for any $x, y \in X$.

(e) BONUS. Prove that with probability approaching 1 as $n$ approaches infinity, $\|f(x) - f(y)\|^2_2 \geq (1 - \epsilon)\|x - y\|^2_2$, for any $x, y \in X$. 

\[\]