1. Analyze the construction below of a stretch 3 distance oracle for a graph $G$, and show that its storage requirement almost matches that shown in class. (It is not really a distance oracle because its query time is not fast enough.) Analyze also the accuracy (stretch bound). Explain whether your bounds (storage and accuracy) hold in the worst-case, in expectation, or with high probability.

Preprocess($G$): Choose $L \subseteq V$ as a random set of $l = \sqrt{n \log n}$ “landmark” vertices (with or without repetitions). For every vertex $v \in V$, store its distances (i) to the $n$ vertices closest to it (denoted $B_v \subseteq V$); and (ii) to all the landmark vertices.

Query($u,v$): If $u \in B_v$, i.e., $u$ is among the $n$ closest to $v$, report the distance. Otherwise, report $\min_{w \in L} [d(u,w) + d(w,v)]$.

Hint: in the “otherwise” case, show that $L \setminus B_v \neq \emptyset$.

2. Let $B$ be a randomized algorithm that approximates some function $f(x)$ as follows:

$$\forall x, \quad \Pr \left[ B(x) \leq (1 + \varepsilon) f(x) \right] \geq 2/3.$$ 

Let algorithm $C$ output the median of $O(\log \frac{1}{\varepsilon})$ independent executions of algorithm $B$ on the same input. Prove that

$$\forall x, \quad \Pr \left[ C(x) \leq (1 + \varepsilon) f(x) \right] \geq 1 - \delta.$$ 

3. Design a streaming algorithm for the $\ell_1$-point query problem, i.e., producing an estimate $\tilde{x}_i \in x_i \pm \varepsilon \|x\|_1$.

For simplicity, ignore the issue of storing the random bits.

Hint: Show a linear sketch by extending the count-min sketch seen in class (so as to remove the restriction $x_i \geq 0$).

**Extra credit:**

4. Let $\varphi$ be an arbitrary 2-SAT formula on $n$ variables $x_1, \ldots, x_n$. Assume that every clause $c$ has weight $w_c \geq 0$, and the total weight is $\sum_c w_c = 1$ (by normalization). A 2-SAT formula $\varphi'$ will be called a *sparsifier* of $\varphi$ if it contains a subset of the clauses of $\varphi$, with arbitrary new weights $w'_c$. 
Show that \( \varphi \) admits a sparsifier \( \varphi' \) with \( O(n/\varepsilon^2) \) clauses, such that for every truth assignment \( A \) to the \( n \) variables, the value of \( \varphi \) (i.e., total weight of clauses satisfied by \( A \)) differs from that of \( \varphi' \) by at most \( \varepsilon \) (additively).

Hint: Sample exactly \( t = O(n/\varepsilon^2) \) clauses from \( \varphi \) with repetitions, and give each of them weight \( 1/t \), and analyze any fixed truth assignment using Hoeffding’s inequality (not Chernoff).