15-496/859X: Computer Science Theory for the Information Age — Spring 2012
Carnegie Mellon University — V. Guruswami & R. Kannan

Problem Set 1

Due in class, Thursday, February 2

Instructions

- You should think about *each* problem by yourself for *at least 30 minutes* before choosing to collaborate with others.
- You are allowed to collaborate with fellow students taking the class in solving the problems (in groups of at most 3 people for each problem). In fact this is encouraged so that you interact with and learn from each other. However, *you must write up your solutions on your own.* If you collaborate in solving problems, you should clearly acknowledge your collaborators for each problem.
- Reference to any external material besides the course text and material covered in lecture is not allowed. In particular, you are not allowed to search for answers or hints on the web. You are encouraged to contact the instructors or the TA for a possible hint if you feel stuck on a problem and require some assistance.
- Solutions typeset in LaTeX are preferred.
- Feel free to email the instructors or the TA if you have any questions or would like any clarifications about the problems.
- You are urged to start work on the problem set early. There are seven problems. Only students signed up for the graduate version need to turn in the last problem (others may do so as extra credit if they have time).

---

1. Consider the unit radius ball in the ℓ_4-norm:

$$K = \{\mathbf{x} = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d \mid x_1^4 + x_2^4 + \cdots + x_d^4 \le 1\} .$$

   (a) Consider the subset $S = \{\mathbf{x} = (x_1, x_2, \ldots, x_d) \mid x_1^4 + x_2^4 + \cdots + x_d^4 \le 1/2\}$. What fraction of the volume of K does S occupy?

   (b) For any $c > 0$, prove that the fraction of volume of K outside the slab $|x_1| \le c/d^{1/4}$ is at most $\frac{1}{c^3} e^{-c^4/4}$.

2. Exercise 1.8 of the book Chapter 1. (Overlap of spheres)

3. Let u, v be two fixed unit vectors in d-dimensions. Ler $r \in \mathbb{R}^d$ be sampled by picking its coordinates independently according to the standard Gaussian. The notation $\langle x, y \rangle$ stands for the inner product between vectors $x, y \in \mathbb{R}^d$.

   (a) What is the expected value (over the choice of r) of $\langle u, r \rangle$?

   (b) What is the expected value of $|\langle u, r \rangle|$? (Hint: Use the rotational symmetry of the d-dimensional Gaussian.)

   (c) What is the expected value of $\langle u, r \rangle \cdot \langle v, r \rangle$?

   (d) What is the probability that $\langle u, r \rangle$ and $\langle v, r \rangle$ have the same sign?

4. Prove that for every fixed dimension reduction matrix $A \in \mathbb{R}^{k \times d}$ with $k < d$, there is a pair of vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ such that the distances between their images $A\mathbf{x}, A\mathbf{y}$ is hugely distorted (compared to the distance between $\mathbf{x}$ and $\mathbf{y}$).

5. Fix an even integer $d \ge 2$ and $\epsilon > 0$. Let us say that $\mathsf{COVER} \subseteq \mathbb{R}^d$ is a (d, ϵ)-covering set if for every point x in the d-dimensional unit cube $[0, 1]^d$, there is some point $c \in \mathsf{COVER}$ such that the Euclidean distance $\|x - c\|$ between x and c is at most ϵ.

(a) Prove that the size of any (d, ϵ)-covering set COVER must satisfy the lower bound

$$|\mathsf{COVER}| \geq \frac{(d/2)!}{(\pi\epsilon^2)^{d/2}} .$$

(b) Prove that there exists a (d, ϵ)-covering set COVER with

$$|\mathsf{COVER}| \leq \left(\frac{2d}{\epsilon^2}\right)^{d/2} .$$

6. NASA's Jet Propulsion Laboratory (JPL) needs to periodically send m-bit *messages* to their robots out in space exploring our solar system. But they've found that when you transmit bits all the way into space, about 5% of them get corrupted (0's changed to 1's or 1's changed to 0's). So the JPL employs a *codebook* for communication with their robots in space. For each m-bit string x, they have a designated n-bit *codeword*, $C(x)$, for $n = 10m$.

   Whenever JPL wants to send the message x, they send $C(x)$ instead. They would like the codewords $C(x)$ have the following very useful property:

   Useful Property: For any two distinct m-bit messages x and y, the codewords $C(x)$ and $C(y)$ differ on more than $n/10$ of the n bit-positions.

   (In your spare time, you can think about why this is a good idea for their goal.)

   JPL tries to make such a codebook probabilistically as follows:

   `for each` m`-bit string` x

   $\quad C(x) \leftarrow$ `uniformly random` n`-bit string`

   You will now argue that the code will have the above useful property with high probability.

   (a) Let x and y be two distinct m-bit strings. Show that

   $$\Pr[C(x) \text{ and } C(y) \text{ differ on more than } n/10 \text{ of the } n \text{ bit-positions}] \geq 1 - 2^{-3m} .$$

   Clearly explain what tail bounds you use and how.

   (b) Show that except with exponentially small probability, JPL's codebook has the above-stated *Useful Property.*

7. **(For graduate version only; Undergraduates may attempt as an extra credit problem)** Suppose we are given a simple undirected graph $G = (V, E)$. The goal is to place each vertex u at a point x_u on the surface of a sphere of radius 1 in d dimensions so that the average of (squared) distances along the edges

   $$\frac{1}{|E|} \sum_{e=(u,v)\in E} \|x_u - x_v\|^2 \qquad (1)$$

   is maximized. (Here $\|x - y\|$ denotes the Euclidean distance between vectors x and y.)

   (a) Argue that it suffices to take $d = |V|$ to achieve this maximum.

   (b) Let the maximum value of (1) be Θ (achieved with some $d \leq |V|$). Prove that for all $\epsilon > 0$, there is a set of unit vectors $\{x_u\}_{u\in V}$ in $O(\frac{1}{\epsilon^2}\log(1/\epsilon))$ dimensions for which the quantity (1) is at least $\Theta - \epsilon$. (Note that this number of dimensions is independent of the number of vertices!)