Homework 1, due Tuesday 10/3

1. Hashing the cube. You have a collection of nonzero distinct binary vectors, \( x_1, \ldots, x_m \in \{0, 1\}^n \). To facilitate later lookup, you decide to hash them down to vectors of length \( p < n \) by means of a linear mapping

\[ x_i \mapsto Ax_i, \]

where \( A \) is a \( p \times n \) matrix with \( 0-1 \) entries, and all computations are performed modulo 2. Suppose the entries of this matrix are picked uniformly at random (each an independent coin toss).

(a) Pick any \( 1 \leq i \leq m \), and any \( b \in \{0, 1\}^p \). Show that the probability (over the choice of \( A \)) that \( x_i \) hashes to \( b \) is exactly \( 1/2^p \). Hint: focus on a coordinate \( 1 \leq j \leq n \) for which \( x_{ij} = 1 \).

(b) Pick any \( 1 \leq i < j \leq m \). What is the probability that \( x_i \) and \( x_j \) hash to the same vector? This is called a collision.

(c) Show that if \( p \geq 2 \log_2 m \), then with probability at least 1/2, there are no collisions among the \( x_i \). Thus: to avoid collisions, it is enough to linearly hash into \( O(\log m) \) dimensions.

2. Almost-orthogonal points on the unit sphere. Fix any \( \epsilon > 0 \). We want to pick \( M \) points on the surface of the unit sphere \( S^{d-1} \) such that every pair of points \( x_i, x_j (i \neq j) \) is almost orthogonal: \( |x_i \cdot x_j| \leq \epsilon \). Show that it is possible to make \( M \) exponentially large in \( d \) (hint: pick the points randomly and use a bound from class). (Note: if we wanted the points to be perfectly orthogonal, then of course \( M \leq d \))

3. Norms. A norm on \( \mathbb{R}^d \) is a function \( \| \cdot \| : \mathbb{R}^d \to \mathbb{R} \) which satisfies the following properties:
   
   - Positivity: for any \( x \in \mathbb{R}^d \), \( \|x\| \geq 0 \), with equality iff \( x = 0 \).
   - Homogeneity: for any \( x \in \mathbb{R}^d \) and \( t \in \mathbb{R} \), \( \|tx\| = |t| \cdot \|x\| \).
   - Triangle inequality: for any \( x, y \in \mathbb{R}^d \), \( \|x + y\| \leq \|x\| + \|y\| \).

A useful family of norms are the \( l_p \) norms, defined as follows for \( p \geq 1 \):

\[ \|x\|_p = \left( \sum_{i=1}^d |x_i|^p \right)^{1/p}. \]

These include the familiar \( l_1 \), \( l_2 \), and \( l_\infty \) norm (the latter is \( \max_i |x_i| \)). You may assume all of these satisfy the definition of norm.

(a) Show that for any \( x \in \mathbb{R}^d \),

\[ \|x\|_2 \leq \|x\|_1 \leq \|x\|_2 \cdot \sqrt{d} \]

and give examples where each of these inequalities is tight. (You will need Cauchy-Schwarz.)

(b) Show that for any \( x \in \mathbb{R}^d \), and any \( p \geq 1 \), \( \|x\|_1 \geq \|x\|_p \).

(c) Show that for any \( x \in \mathbb{R}^d \), and any \( 1 \leq p \leq q \), \( \|x\|_p \geq \|x\|_q \). That is, the \( l_p \) norm of a vector is always larger than its \( l_q \) norm if \( p \leq q \).

(d) Show that for any \( x \in \mathbb{R}^d \) and \( 1 \leq p \leq q \),

\[ \|x\|_p \leq \|x\|_q \cdot d^{(1/p) - (1/q)}. \]

You will need Holder’s inequality, which says that \( |x \cdot y| \leq \|x\|_a \|y\|_b \) for any vectors \( x, y \) and any \( a, b \geq 1 \) with \((1/a) + (1/b) = 1 \).

4. Suppose \( \| \cdot \| \) is some norm on \( \mathbb{R}^d \). Show that for any vectors \( v_1, \ldots, v_n \in \mathbb{R}^d \),

\[ \sum_{\sigma \in \{+1, -1\}^n} \left\| \sum_{i=1}^n \sigma_i v_i \right\| \geq 2^n \max_i \|v_i\|. \]