MCS 549 – Mathematical Foundations of Data Science
Fall 2022
Problem Set 1

Lev Reyzin

Due: 10/3/22 at the beginning of class

Instructions: Atop your problem set, please write your name and list your collaborators.

Problems

Prove all your answers.

1. Show that for any $c \geq 1$ there exist distributions for which Chebyshev's inequality is tight, i.e. for which $\mathrm{P}(|x - E(x)| \geq c) = \mathrm{Var}(x)/c^2$.

2. For what value of d is the volume of the d-dimensional unit ball maximized?

3.* Suppose we are given n unit vectors in R^n divided into two sets P, Q with the guarantee that there exists a hyperplane $a \cdot x = 0$ such that every point in P is on one side of it and every point in Q is on the other. Furthermore, assume that the ℓ_2 distance of each point to the hyperplane is at least γ (this is sometimes called the "margin"). Show that a random projection (as defined in the book) to some $c \log n/\gamma^2$ dimensions will have the property that with high probability, the two sets of points will still remain separated by a hyperplane with margin $\gamma/2$.

4. Show that if A is a symmetric matrix with distinct singular values, then the left and right singular vectors are the same and $A = VDV^T$.

5. Find the threshold for $p(n)$ for the existence of 4-cliques in $G(n, p(n))$. Prove your answer correct.

6. The example at the end of Section 8.1.1 in the book computes that if the degrees in $G(n, \frac{1}{n})$ were independent, there would be a vertex of degree

$$d = \Omega\left(\frac{\log n}{\log \log n}\right)$$

with constant positive probability. However, the degrees are not independent. Show how to overcome this difficulty and reach the same conclusion.

7. Show that in $G(n, 1/2)$ there are almost surely are no cliques of size greater than or equal to $2 \log_2 n$. Then, use the second moment method to show that in $G(n, 1/2)$, almost surely there are cliques of size $(2 - \varepsilon) \log_2 n$ (for any constant $\varepsilon > 0$).

*This problem is extra challenging.