MCS 549 – Mathematical Foundations of Data Science

Fall 2021

Problem Set 1

Lev Reyzin

Due: 10/4/21 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 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. A Markov chain is said to be symmetric if for all i and j, $p_{ij} = p_{ji}$. What is the stationary distribution of a connected symmetric Markov chain?

6. Given a Markov chain on an undirected graph, we modify the chain as follows: at the current state, we stay there with probability 1/2; with the other probability 1/2, we move as in the old chain. Show that the new chain has the same stationary distribution.