Homework 7
MATH 123 - Spring 2023
Tufts University, Department of Mathematics
Due: March 27, 2023

Question 1

Consider the cube in $C_r^D = [-r/2, r/2]^D \subset \mathbb{R}^D$ in D-dimensions. Let $\text{vol}_D(A)$ denote the volume of a set A in $\mathbb{R}^D$, namely $\text{vol}_D(A) = \int_A dx_1 \dots dx_D$.

(a) Prove using integration that $\text{vol}_D(C_r^D)$ is r^D.

(b) For $\epsilon > 0$, let $A_{\epsilon,r}^D = \{x \in C_r^D \mid x \notin C_{r-\epsilon}^D\}$. Calculate $\frac{\text{vol}_D(A_{\epsilon,r}^D)}{\text{vol}_D(C_r^D)}$.

(c) Use (b) to argue that "most" of the volume of a high dimensional cube is near the boundary. Can you make this precise?

Question 2

Let $w \in \mathbb{R}^{D\times 1}$.

(a) Show that $\{x \in \mathbb{R}^{D\times 1} \mid w^T x = 0\}$ is a $(D-1)$-dimensional linear subspace of $\mathbb{R}^D$ if $w \neq 0$.

(b) Let $b \in \mathbb{R}$. Is it necessarily the case that $\{x \in \mathbb{R}^{D\times 1} \mid w^T x = b\}$ is a $(D-1)$-dimensional linear subspace of $\mathbb{R}^D$? Prove or given a counterexample.

Question 3

Download the dataset "kNN_ClassifierSyntheticData.mat". Randomly select 100 different testing points in the dataset, and run a kNN-classifier for $k\text{NN} = \{1, 10, 50, 100, 500, 900\}$ using the remaining points as training points. How does performance change with the change in kNN?

Question 4

Consider the Salinas A dataset, which may be found at `http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes`. Randomly select 100 different testing points in the dataset, and run a kNN-classifier for $k\text{NN} = \{1, 10, 50, 100, 500, 900\}$ using the remaining points as training points. How does performance change with kNN?