Math 608D, Assignment 2

1. Let $\mathcal{X} \subset \boldsymbol{R}^n$ be a finite point cloud, i.e., $|\mathcal{X}| = N$. We wish we wish to find a nearly isometric embedding of $\mathcal{X}$ into ℓ_1^m. To be precise, fix $0 < \epsilon < 1$. We wish to show there is a mapping $A : \boldsymbol{R}^n \to \boldsymbol{R}^m$ satisfying

$$(1-\epsilon)\left\|x-y\right\|_2 \leq \left\|A(x-y)\right\|_1 \leq (1+\epsilon)\left\|x-y\right\|_2 \qquad \text{for all } x, y \in \mathcal{X}.$$

What mapping makes this embedding (with high probability)? How small can you choose m as a function of n, N, ϵ?

2. Consider the p-norm, $p \geq 1$,

$$\left\|x\right\|_p := \left(\sum_i |x_i|^p\right)^{1/p}.$$

Let $B_p^n := \{x \in \boldsymbol{R}^n : \left\|x\right\|_p \leq 1\}$ be the unit ball for this norm. Let $N(K, \left\|\cdot\right\|_p, \epsilon)$ be the covering number of the a set $K \subset \boldsymbol{R}^n$ using the p-norm, that is, the minimal cardinality of a subset $\mathcal{X} \subset K$ satisfying: for any $x \in K$ there is $y \in \mathcal{X}$ such that $\left\|x-y\right\|_p \leq \epsilon$. Bound the following:

   (a) $N(B_1^n, \left\|\cdot\right\|_1, \epsilon)$.

   (b) $N(B_p^n, \left\|\cdot\right\|_p, \epsilon)$.

   (c) $N(K, \left\|\cdot\right\|_\infty, \epsilon)$, where $K \subset B_\infty^n$ is a polytope with N vertices. (Can you beat the bound in the previous problem when N is not too large?)

   (d) **Tensors.** Consider $\boldsymbol{R}^{n\times n\times n}$, i.e., arrays with 3 *modes*. (Picture a rubix cube.) We call elements of this set tensors. A rank-1 tensor takes the form

$$T = u \times v \times w, \qquad \text{where } u, v, w \in \boldsymbol{R}^n.$$

   In other words,

$$T_{i,j,k} = u_i \cdot v_j \cdot w_k.$$

   We say that a tensor has rank at most r if it may be written as the sum of r rank-1 tensors. The Frobenius norm of a tensor is $\left\|T\right\|_F = \sqrt{\sum_{i,j,k} T_{i,j,k}^2}$. Let K be the set of rank-r tensors with Frobenius norm at most 1. Bound $N(K, \left\|\cdot\right\|_F, \epsilon)$. (I believe such a bound was made in the literature in the past few years, but I haven't read the details.)

   (e) The packing number $P(K, \left\|\cdot\right\|, \epsilon)$ of a set K is the maximal cadinality of a subset $\mathcal{X} \in K$ satisfying $\left\|x-y\right\| > \epsilon$ for all $x, y \in \mathcal{X}$ with $x \neq y$. Lower bound $P(K, \left\|\cdot\right\|_F, \epsilon)$ with K as in the last problem. (I'm not sure whether this has been done.)

3. **Single index model/1-layer neural net.** A 1-layer neural net is a function $f : \boldsymbol{R}^n \to \boldsymbol{R}^m$ of the following form. Let $W \in \boldsymbol{R}^{m\times n}$ be a *weight matrix* and let $\sigma : \boldsymbol{R} \to \boldsymbol{R}$ be some *activation function.* Then

$$f(x) = \sigma(Wx) \qquad \text{where } \sigma \text{ acts elementwise.}$$

In other words, the ith entry of $f(x)$ is

$$(f(x))_i = \sigma(\langle W_i, x\rangle) \qquad \text{where } W_i \text{ is the } i\text{th row of } W.$$

Note: The above model is also called the *single-index model* of statistics.

In this problem we explore the behaviour of f under the assumption that the weights are Gaussian, i.e., assume $W_{i,j} \sim N(0,1)$ and all entries are independent. We will check whether f preserves Euclidean norms.

(a) Suppose that σ is a rectified linear unit i.e.,

$$\sigma(x) = \max(x, 0).$$

i. Fix $x \in \boldsymbol{R}^n$. Let $T_x = \|f(x)\|_2$. What is $\boldsymbol{E}\, T_x^2$?

ii. Bound

$$\|T_x - \boldsymbol{E}\, T_x\|_{\Psi_2}.$$

iii. Let $K \subset S^{n-1}$ be a finite set, i.e., $|K| \leq \infty$. Bound

$$\boldsymbol{E} \sup_{x \in K} |T_x - \boldsymbol{E}\, T_x|.$$

iv. We now pass to the infinite case. Bound

$$\boldsymbol{E} \sup_{x \in \boldsymbol{S}^{n-1}} |T_x - \boldsymbol{E}\, T_x|.$$

v. Now the more challenging infinite case. Let $K \subset \boldsymbol{S}^{n-1}$ be an arbitrary infinite set. Bound

$$\boldsymbol{E} \sup_{x \in K} |T_x - \boldsymbol{E}\, T_x|.$$

(Your answer should depend on how "large" K is.)

(b) Let $\sigma : \boldsymbol{R} \to \boldsymbol{R}$ be a general function. Complete Items i-v above in this setting. Your answers should depend on properties of σ. (You may assume what you need.)