Assignment 3

CMSC 726: Machine Learning
October 9$^{\text{th}}$, 2018 **Name:**

LATE SUBMISSIONS WILL NOT BE ACCEPTED FOR THIS HOMEWORK.

1. In this question, we compute the VC dimension of a Hypothesis class. Consider the following hypothesis class $\mathcal{H} = \{h(x) : sign(\mathbf{w}^T\mathbf{x}) | x \in \mathbb{R}^n\}$, where $sign(z) = 1$ if $z \geq 0$ and $sign(z) = 0$ if $z \leq 0$.

   (a) Show that VCdim$(\mathcal{H}) \geq n$.
   Hint: Consider a set of points $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$ that correspond to the standard basis in $\mathbb{R}^n$, i.e. $\mathbf{x}_k^{(i)} = 1$ if $k = i$ and $\mathbf{x}_k^{(i)} = 0$ if $k \neq i$. Show that using $\mathbf{w} = \sum_{i=1}^d y^{(i)}\mathbf{x}^{(i)}$, we can classify all points $(\mathbf{x}^{(i)}, y^{(i)})$ correctly.

   (b) Show that VCdim $\leq n$.
   Hint 2: Suppose there exists a set of points $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n+1)}$ (more than n) such that they can be shattered by $\mathcal{H}$. Form a matrix $\mathbf{H} = \mathbf{X}\mathbf{W}$ where $\mathbf{X} = [\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n+1)}]^T$ and $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_{2^n}]$ where each $\mathbf{w}_i, 1 \leq i \leq 2^n$ corresponds to a possible labeling of points. Show that this implies that $rank(\mathbf{H}) \leq n$. Moreover, show that the fact that $\mathcal{H}$ shatters $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n+1)}$ implies that $rank(\mathbf{H}) \geq n + 1$. This leads to a contradiction.

2. Recall that the Rademacher complexity of a class of functions $\mathcal{F}$ is defined as

$$R_m(\mathcal{F}) = \frac{1}{m}\mathbb{E}_\sigma\left[\sup_{f \in \mathcal{F}} \sum_{i=1}^m \sigma_i f(z_i)\right].$$

   Suppose $\mathbf{x} \in \{0, 1\}^n$ has only k non-zero entries. Define the following class of linear functions where the parameters are bounded by $B > 0$, i.e.: $\mathcal{F} = \{\mathbf{x} \to \mathbf{w}^T\mathbf{x} : \max_i |w_i| \leq B\}$. Compute an upper bound on the Rademacher complexity as a function of B, k, n, m.

   Hint: see the proof of Lemma 26.10 in section 26.2 of the textbook.

3. (Programming Assignment) The Hoeffding inequality states that:

$$\mathbb{P}\left[\left|\frac{\theta_1 + \cdots + \theta_m}{m} - \mathbb{E}[\theta]\right| \geq \epsilon\right] \leq 2e^{\frac{-2m\epsilon^2}{(b-a)^2}}$$

   where θ_i's are i.i.d. random variables such that $a \leq \theta_i \leq b$.
   Let each θ_i be generated from $\mathbb{P}_\theta$ which is a uniform $[0, 1]$ distribution.

   (a) Generate $k = 100$ many sets where each set S_i consists of $m = 100$ i.i.d. samples from $\mathbb{P}_\theta$.

   (b) What is the fraction of k sets that satisfies the following bound: $\mathbb{P}[|\frac{\theta_1+\cdots+\theta_m}{m} - \frac{1}{2}| \leq 0.1]$?

   (c) Compare the number from part (b) with the Hoeffding bound.