1. Consider a classification problem with $\mathcal{X} = [0, 1]^d$ and $\mathcal{Y} = \{0, 1\}$. Let $\mathcal{F}$ denote the collection of all histogram classifiers $f : [0, 1]^d \rightarrow \{0, 1\}$ with $M$ equal volume bins. Do not assume that $\min_{f \in \mathcal{F}} R(f) = 0$. For a certain $\epsilon > 0$ and $\delta > 0$, how many samples $n$ are needed for an $(\epsilon, \delta)$-PAC bound? Compare with results from HW 3.

2. Consider a classification problem with $\mathcal{X} = [0, 1]^2$ and $\mathcal{Y} = \{0, 1\}$. Let $\{v_j\}_{j=1}^K$ be a collection of $K$ points uniformly spaced around the perimeter of the unit square. Let $\mathcal{F}$ denote the set of linear classifiers obtained by connecting any two points in $\{v_j\}$ with a line. Do not assume that $\min_{f \in \mathcal{F}} R(f) = 0$. Give a bound for the estimation error in terms of $K$ and the number of training data $n$. Compare with results from HW 3.

3. Decision trees can be defined by recursively partitioning the input space $\mathcal{X}$. Let $\mathcal{X} = [0, 1]^2$ and partition it into 4 subsquares of equal area by splitting the horizontal and vertical coordinate axes at $1/2$. Repeat this process to each subsquare (this results in 16 squares). If you recursively apply the partitioning process $m$ times then you obtain a uniform partition of $\mathcal{X}$ into $4^m$ subsquares of equal area. The recursive partitioning can be represented with a tree with $m$ levels, where the leaves denote the final squares, the root denotes the single initial square $[0, 1]^2$, and internal vertices denote branching/splitting steps in the process.

Now we can use training data $\{X_i, Y_i\}_{i=1}^n$ to assign a binary decision label to each square (assume the $Y_i$ are binary). If we use the full tree, then this is equivalent to a histogram classifier with $4^m$ cells. Alternatively, we could consider “pruning” the tree (which amounts to “undoing” some of the splitting of the recursive partitioning process and re-merging some of the cells). In general, the pruning process results in a non-uniform partition with $k < 4^m$ cells of varying sizes.

Devise a prefix coding scheme for representing decision trees. Let $T$ denote a specific decision tree and let $\mathcal{T}$ denote all possible decision trees obtained by recursive dyadic partitioning. Use the codes $c(T)$ to construct a bound of the following form: for any $\delta > 0$ with probability at least $1 - \delta$

$$R(T) \leq \hat{R}(T) + \sqrt{\frac{c(T) + \log(1/\delta)}{2n}} \quad \forall T \in \mathcal{T}$$