ECE 901
Homework 2

1. Consider the histogram classification rule discussed in Lecture 5. In lecture, we showed that the histogram rule is consistent if the number of bins $M \to \infty$ and $n/M \to \infty$ as the number of training data $n \to \infty$. The rate of convergence can be arbitrarily slow, depending on the complexity of the unknown conditional probability function $\eta(x) = P(Y = 1|X = x)$. However, if we make some assumptions about $\eta(x)$, then it is possible to determine a good choice for M and a rate of convergence. Specifically, let's assume that $\eta(x)$ is a Lipschitz function on $[0, 1]^d$. Derive a good choice for M in terms of the sample size n, and determine the resulting rate of convergence.

Hint: Show that the approximation error of $\hat{\eta}_n$ will be $O(M^{-1/d})$ in this case. To bound the estimation error, we need to bound $E[N^{-1/2}|N > 0]$ where N is the number of samples falling in a particular bin. Assuming that the density of X is bounded below and above by a constants C_1 and C_2, respectively, it follows that the probability p of falling into a given bin $C_1/M < p < C_2/M$. If we have n training samples, then $N \sim \text{Binomial}(n, p)$, for some $p \in [C_1/M, C_2/M]$. Use a first order Taylors series with remainder to approximate the function $f(k) = 1/\sqrt{k}$ about the point $k = np$ and then compute the expectation (the first order term is zero in expectation and the second order remainder term tends to zero as $n \to \infty$). Thus, $E[N^{-1/2}|N > 0] = O(1/\sqrt{np}) = O(\sqrt{M/n})$.

2. 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 $f : [0, 1]^d \to \{0, 1\}$ with M equal volume bins. 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 (ϵ, δ)-PAC bound?

3. 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. 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.