

ECE 901 Homework 4

1. Histogram classifiers vs. Tree classifiers. Consider the following set-up. Let $\mathcal{X} = [0, 1]^2$, $\mathcal{Y} = \{0, 1\}$, and suppose we are given n iid training examples $\{X_i, Y_i\}$. Furthermore, assume that the Bayes classifier is $1_{\eta(x) > 1/2}$, where $\eta(x) = P(Y = 1|X = x)$. Assume that the boundary of the Bayes decision region is a non-fractal, 1-d curve. Specifically, assume that the boundary has box-counting dimension 1. This means that if we divide the unit square into m^2 boxes, each of side-length $1/m$, then the number of boxes that the boundary intersects is bounded by Cm , for some constant $C > 0$ and for every positive integer m .
 - a. Using the fact that the Bayes decision boundary has box-counting dimension 1, prove that the histogram classifier selected according to the complexity regularization procedure described in class converges to the Bayes classifier at rate of at least $n^{-1/4}$.
Hint: Consider a histogram classifier with m^2 boxes, each of sidelength $1/m$, apply the risk bound we derived in class, and minimize with respect to m .
 - b. Prove that the dyadic tree classifier selected according to the complexity regularization procedure described in class converges to the Bayes classifier at rate of at least $n^{-1/3}$.
Hint: Begin by considering the partition associated with the histogram classifier with m^2 boxes, where m is a power of 2. Prove that there exists a pruned dyadic partition with at most $8Cm$ regions such that the Bayes decision boundary is completely contained in small boxes of sidelength $1/m$. Finish by applying the risk bound from class in this case and optimizing over choice of m .
2. Complexity regularization in regression. Consider learning under squared error loss. Suppose we have n iid training examples $\{X_i, Y_i\}_{i=1}^n$ and a collection \mathcal{F} of candidate functions mapping $\mathcal{X} = \mathbb{R}^d$ to $\mathcal{Y} = \mathbb{R}$. Assume that the support of the Y_i and the range of the candidate functions $f \in \mathcal{F}$ is in a known interval of length b . Our empirical and true risks are given by

$$\hat{R}_n(f) = \frac{1}{n} \sum_{i=1}^n (f(X_i) - Y_i)^2$$

$$R(f) = E[(f(X) - Y)^2] = \frac{1}{n} \sum_{i=1}^n E[(f(X_i) - Y_i)^2]$$

Notice that we can write $R(f) - \hat{R}_n(f) = \frac{1}{n} \sum_{i=1}^n (U_i - E[U_i])$, where $U_i = (f(X_i) - Y_i)^2$. Assume that we have the following concentration inequalities for $\sum_{i=1}^n (U_i - E[U_i])$:

$$P\left(\sum_{i=1}^n (U_i - E[U_i]) > \epsilon\right) \leq e^{-\epsilon} \quad \text{and} \quad P\left(\sum_{i=1}^n (E[U_i] - U_i) > \epsilon\right) \leq e^{-\epsilon}$$

Select a model from \mathcal{F} according to

$$\hat{f}_n = \arg \max_{f \in \mathcal{F}} \left\{ \hat{R}_n(f) + \frac{c(f) \log 2}{n} \right\}$$

where $\{c(f)\}$ are positive numbers satisfying $\sum_{f \in \mathcal{F}} 2^{-c(f)} \leq 1$. Using the concentration inequalities above and mimicking the derivation of the risk bound for complexity regularized classifier selection in Lecture 10, prove that

$$E[R(\hat{f}_n)] \leq \min_{f \in \mathcal{F}} \left\{ R(f) + \frac{c(f) \log 2}{n} \right\} + \frac{\log n + b^2 + 1}{n}$$

Compare this result to the bound one obtains using Hoeffding's inequality

$$P\left(\sum_{i=1}^n (U_i - E[U_i]) > \epsilon\right) \leq e^{-2\epsilon^2/n} \quad \text{and} \quad P\left(\sum_{i=1}^n (E[U_i] - U_i) > \epsilon\right) \leq e^{-2\epsilon^2/n}$$