1. Complexity Regularization for Parameter Estimation. Suppose make observations according to the following model:

\[ Y_i = \theta^* + W_i, \quad i = 1, \ldots, n \]

where \( \theta^* \in \mathbb{R}^d \) and \( \{W_i\} \) are iid vectors whose entries are iid random variables with mean zero and variance 1.

a. Using standard statistical methods, derive an estimator \( \hat{\theta}_n \) satisfying

\[ E[\|\hat{\theta}_n - \theta^*\|^2_2] = \frac{d}{n} \]

b. Derive a similar estimator using the results of Lecture 12. Formulate a squared error empirical risk function for this problem. Assume that \( \theta^* \) satisfies \( \|\theta^*\|_2^2 \leq 1 \). Then we can restrict our selection to vectors in the unit ball. Furthermore, assume that the elements of \( \{W_i\} \) are bounded in magnitude by \( B > 0 \). Then we can use the Craig-Bernstein (CB) inequality to gauge how close the empirical risk is to the true risk. In order to use the CB inequality in the context of model selection, we need a uniform bound that holds over all models under consideration. To this end, it is necessary to consider only a finite collection of models. Note that the unit ball in \( d \) dimensions contains an infinite number of vectors, but we can discretize this set by finely quantizing the elements of the vectors. For example, if we quantize a vector \( \theta \) in the unit ball with \( 1/2 \log n \) bits of precision for each coordinate, then it is easy to show that the error of the quantized vector, \( \theta_q \), satisfies \( \|\theta_q - \theta\|_2^2 \leq 4d/n \). Show that the CB inequality leads to a bound of the form

\[ E[\|\hat{\theta}_n - \theta^*\|^2_2] \leq (C \log n) \frac{d}{n} \]

where \( C > 0 \) is a constant.

2. Complexity Regularization for Denoising. Suppose we observe samples of a Lipschitz function in noise:

\[ Y_i = f^*(X_i) + W_i, \quad i = 1, \ldots, n \]

where \( f^* \) is a bounded Lipschitz function on the unit interval \([0, 1]\) (assume that the Lipschitz constant is 1 and the function is bounded \(|f^*(x)| \leq 1\), the \( X_i \) are i.i.d. random variables on \([0, 1]\), and \( W_i \) are i.i.d. zero-mean random variables (noises) satisfying \(|W_i| \leq 1\). This is essentially the denoising problem we examined in Lecture 4, except that in this case the sampling points \( \{X_i\} \) are random according to an unknown distribution, rather than deterministic. Consider histogram estimators for \( f^* \). Derive a complexity regularization scheme that balances the bias-variance tradeoff by automatically selecting the appropriate number of bins in the histogram. The design of this scheme will be analogous to the design of histogram classifiers (cf. Lecture 9), but instead of each bin being labeled 0 or 1, in this case each bin will be assigned a level that predicts the values of \( f^* \) on that bin. In order to apply the complexity regularization theory and methods of Lecture 12, you must discretize the space of all histogram models (i.e., quantize the levels in a fashion similar to the exercise above). Show that your resulting complexity regularization procedure yields an estimator \( \hat{f}_n \) whose mean squared error is \( E[\|f^* - \hat{f}_n\|^2_2] = O(n^{-2/3} \log n) \), nearly the same order that we obtained in Lecture 4 with a fixed, deterministic sampling scheme.