IT for Statistics and Learning
2023

Assignment 5
Assigned: Fri Dec 8 2023
Due: Fri Dec 15 2023 M. Skoglund

Problem 5.1: Assume that $X_1, \ldots, X_n$ are iid $\sim \mathcal{N}(\theta, 1)$. Let $\hat{\theta}_1(X_1, \ldots, X_n)$ be the median of the samples, and let

$$T = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Find a rule $P_{\hat{\theta}|T=t}$ based on T that has the same risk as $\hat{\theta}_1$ no matter what the loss function is.

Problem 5.2: For $\theta \to X \to \hat{\theta}$ on $(\Omega, \mathcal{A}, P)$, where all variables are discrete and θ is uniformly distributed, let $E = 1$ on $\{\omega : \hat{\theta} \neq \theta\}$ and $E = 0$ o.w. Derive Fano's inequality using the simple fact that

$$H(E, \theta|\hat{\theta}) = H(\theta|\hat{\theta}) + H(E|\theta, \hat{\theta}) = H(E|\hat{\theta}) + H(\theta|E, \hat{\theta})$$

(thus arriving at an alternative proof to the one shown in class).

Problem 5.3: Consider the Gaussian location model $X_i \sim \mathcal{N}(\theta, I_p)$, $i = 1, \ldots, n$. Assume $\ell(\theta, \hat{\theta}) = \|\theta - \hat{\theta}\|^2$ where $\|\cdot\|$ is the Euclidean norm in p dimensions. Using the Shannon lower bound, verify the (tight) lower bound

$$R_n^* \geq \frac{p}{n}$$

on the minimax risk.

Problem 5.4: Assume that X_i, $i = 1, \ldots, n$, are iid uniformly distributed over the interval $[\theta, \theta + 1]$. Assuming $\ell(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$ prove that the minimax risk decreases as n^{-2} with the number of samples n.

Problem 5.5: Assume that S is drawn uniformly at random from $\{-1, 0, 1\}^d$ subject to $\|S\|^2 = k < d$. That is, the d-dimensional vector S is drawn uniformly subject to having k non-zero components (-1 or 1). Assume we observe

$$Y = X(\theta S) + W$$

where $W \sim \mathcal{N}(0, I_n)$ and where $\theta > 0$ is a fixed and known scalar. The $n \times d$ matrix X, with $d < n$, is also fixed and known. This is a model for *sparse linear regression.* Assume now we wish to estimate S from Y as $\hat{S} \in \{-1, 0, 1\}^d$ s.t. $\|\hat{S}\|^2 = k$. Show that there is a constant $C > 0$ such that unless

$$n \geq C \frac{\frac{d}{k} \log \binom{d}{k}}{\|n^{-1/2} X\|^2 \, \theta^2}$$

(where $\|\cdot\|^2$ is the Frobenius norm) we have $\Pr(\hat{S} \neq S) \geq 1/2$ for any $\hat{S}$. You can assume that $k \geq 4$ and $\log \binom{d}{k} \geq 10$.