IT for Statistics and Learning
2024

Assignment 10
Assigned: Fr, Jan 26, 2024
Due: before the lecture on Fr, Feb 2, 2024 T. Oechtering

Problem 10.1: *Complete the proofs.* Show that the following properties hold for the Hellinger distance:

(i) Le Cam 's Inequality (Hint: Cauchy-Schwarz inequality could be useful.)

$$||P_1 - P_2||_{TV} \leq H(P_1||P_2)\sqrt{1 - \frac{H^2(P_1||P_2)}{4}}$$

(ii) Decoupling property for product measures $P^{1:n} = \otimes_{i=1}^n P_i$ and $Q^{1:n} = \otimes_{i=1}^n Q_i$

$$\frac{1}{2}H^2(P^{1:n}||Q^{1:n}) = 1 - \prod_{i=1}^n (1 - \frac{1}{2}H^2(P_i||Q_i)) \overset{iid}{\leq} n\frac{1}{2}H^2(P_1||Q_1)$$

Problem 10.2: *Bounds on Gaussian distribution family.* Recall the mean estimation of Gaussian distribution family from the lecture. Use the two-point form of Le Cam's method and Pinsker inequality to derive the following shaper lower bounds

$$\inf_{\hat{\theta}} \sup_{P_\theta \in \mathcal{P}} E_{P_\theta}\left[|\hat{\theta}(X_1^n) - \theta|\right] \geq \frac{\sigma}{8\sqrt{n}} \qquad \inf_{\hat{\theta}} \sup_{P_\theta \in \mathcal{P}} E_{P_\theta}\left[|\hat{\theta}(X_1^n) - \theta|^2\right] \geq \frac{\sigma^2}{16n}$$

Problem 10.3: *Complete the proof.* For the lower bound in the logistic regression problem show that we have

$$\frac{1}{2^d d}\sum_{j=1}^d \sum_{v\in\{-1,+1\}^d} \|P_{v,+j} - P_{v,-j}\|_{TV}^2 \leq \frac{\delta^2}{d}\sum_{j=1}^d\sum_{i=1}^n X_{ij}^2 = \frac{\delta^2}{d}\|X\|_F^2 \tag{1}$$

To this end

(i) show that for two distributions Bernoulli(p_a) and Bernoulli(p_p) with $p_a = \frac{1}{1+e^a}$ and $p_b = \frac{1}{1+e^b}$ we have

$$D(p_a\|p_b) + D(p_b\|p_a) \leq (a-b)^2.$$

(ii) Use the previous inequality and two times Pinsker's inequality (note $\|P_1 - P_2\|_{TV} = \frac{1}{2}(\|P_1 - P_2\|_{TV} + \|P_2 - P_1\|_{TV})$) to show

$$\|P_v^n - P_{v'}^n\|2 \leq \frac{\delta^2}{4}\sum_{i=1}^n \left(X_i^T(v-v')\right)^2$$

Show how the inequality (1) follows from this inequality.

Problem 10.4: *Distribution estimation.* Given n independent samples $x_1, \ldots, x_n$ drawn according to distribution P defined on finite set $\{1, 2, \ldots, k\} = [k]$. In this problem we will show that the minimax rate[1] for estimating P with respect to the total variation is given as follows

$$\inf_{\hat{P}} \sup_{P\in\mathcal{P}([k])} E_P\left[\|\hat{P}(X_1,\ldots,X_n) - P\|_{TV}\right] \asymp \min\{\sqrt{\tfrac{k-1}{n}}, 1\}$$

[1]Let L_n and U_n denote lower and upper bounds of the minimax risk. If $L_n = cn^{-\alpha}$ and $U_n = Cn^{-\alpha}$ for some c, C and α, then we have established the *minimax rate* $n^{-\alpha}$ and we write $R_n \asymp n^{-\alpha}$.

(i) Show that the maximal likelihood estimator P_{MLE} coincides with the empirical distribution.

(ii) Show that MLE is rate-optimal, i.e., the minimax rate is achieved within a constant factor.

(iii) Establish the minimax lower bound via Assouard's lemma.

(iv) Establish the minimax lower bound via Fano's inequality + volume or explicit packing.

(v) (optional to trade one of the previous subquestion) Establish the minimax lower bound via mutual information method.