IT for Statistics and Learning
2024

Assignment 9
Assigned: Fr, Jan 19, 2024
Due: before the lecture on Fr, Jan 26, 2024 T. Oechtering

Problem 9.1: *Complete the proofs.* Show that the following properties hold for the KL divergence:

(i) Given non-negative weights λ_i with $\sum_{i=1}^{n} \lambda_i = 1$, then we have

$$D(\sum_{i=1}^{n} \lambda_i P_i \| Q) \leq \sum_{i=1}^{n} \lambda_i D(P_i \| Q) \qquad D(Q \| \sum_{i=1}^{n} \lambda_i P_i) \leq \sum_{i=1}^{n} \lambda_i D(Q \| P_i)$$

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

$$D(P^{1:n} \| Q^{1:n}) = \sum_{i=1}^{n} D(P_i \| Q_i)$$

(iii) For mixture distribution $\bar{Q} = \frac{1}{n} \sum_{i=1}^{n} P_i$ and any other distribution Q we have

$$\frac{1}{n} \sum_{j=1}^{n} D(P_j \| \bar{Q}) \leq \frac{1}{n} \sum_{j=1}^{n} D(P_j \| Q)$$

Problem 9.2: *Complete the proofs.* Show that we have the following relations between packing and covering

$$M(2\delta, \Theta, \rho) \leq N(\delta, \Theta, \rho) \leq M(\delta, \Theta, \rho)$$

Problem 9.3: *Covering number of unit cubes.* Show that we have $N(\delta, [-1, 1]^d, \| \cdot \|_\infty) \leq (\frac{1}{\delta} + 1)^d$ for unit cubes in $\mathbb{R}^d$.

Problem 9.4: *Unit sphere.* Let $S^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ denote the unit sphere in $\mathbb{R}^n$.

(i) Show that $M(\delta, S^{n-1}, \| \cdot \|_2) \leq \left(1 + \frac{2}{\delta}\right)^n$

(ii) Show that the bound $M(\delta, S^{n-1}, \| \cdot \|_2) \leq \left(1 + \frac{2}{\delta}\right)^n$ is not tight by deriving the exact covering number for the case for $n = 1$.

Problem 9.5: *Local Fano.* Consider the mean estimation of a vector-valued Gaussian distribution with $\mathcal{P} = \{\mathcal{N}(\mu, I_d) : \mu \in \mathbb{R}^d\}$ and loss $\| \cdot \|_2$. Given n observations $x_1, \ldots x_n$ independently drawn according to $\mathcal{N}(\mu, I_d)$.

(i) Show that for any $\mu \in \mathbb{R}^d$ the empirical mean estimator $\hat{\theta}(x_1, \ldots, x_n) = \frac{1}{n} \sum_{k=1}^{n} x_k$ results in

$$E\big[\|\hat{\theta}(X_1, \ldots, X_n) - \mu\|_2^2\big] = d/n$$

(ii) Show that we have $D(\mathcal{N}(\mu_1, I_d) \| \mathcal{N}(\mu_2, I_d) = \frac{1}{2\log(2)} \|\mu_1 - \mu_2\|_2^2$.

(iii) Let $B_d(c, r) = \{x \in \mathbb{R} : \|x - c\|_2 \leq r\}$ denote the ball in $\mathbb{R}^d$ with center c and radius r. Show that we have '

$$M(\delta, B_d(0, 1), \| \cdot \|_2) \geq N(\delta, B_d(0, 1), \| \cdot \|_2) \geq \frac{1}{\delta^d}$$

(iv) Use the local Fano method using a 0.5-packing of $B_d(0,1)$ to show

$$\inf_{\hat{\theta}} \sup_{\mu \in \mathbb{R}^d} E\big[\|\hat{\theta} - \mu\|_2^2\big] \geq \delta^2 \left(1 - \frac{n(32\delta^2/\ln 2) - 1}{d}\right)$$

(v) Can you choose δ such that it follows that the empirical mean estimator is asymptotically optimal?