CS861: Theoretical Foundations of Machine Learning Lecture 16 - 10/11/2023
University of Wi[illegible] 2023

Lecture 16: [illegible] prediction problems, Stochastic Bandits

Lecturer: Kirthe[illegible] *Scribed by: Ransheng Guan, Haoran Xiong*

Disclaimer: *These n[illegible]ected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the instructor.*

In this lecture, we will continue our discussion on proving minimax lower bounds for prediction problems, and use it to prove a lower bound for classification in a VC class. Finally, we will briefly introduce `Stochastic Bandits`.

1 Excess risk in classification/regression (Cont'd)

Let $\mathcal{Z}$ be a data space, $\mathcal{P}$ be a family of distribution, and $\mathcal{H}$ be a hypothesis space. Let $f : \mathcal{H} \times \mathcal{Z} \to \mathbb{R}$ be the instance loss, where $f(h, Z)$ is the loss of hypothesis h on instance Z. Let $F(h, P) = \mathbb{E}_{Z\sim P}[f(h, Z)]$ be the population loss of hypothesis h on distribution P, and let $L(h, P) = F(h, P) - \inf_{h\in\mathcal{H}} F(h, P)$ denote the excess population loss.

Then a dataset S drawn from some $P \in \mathcal{P}$; an estimator $\hat{h}$ mapping the data to a hypothesis in $\mathcal{H}$. Thus, the risk would be

$$R(\widehat{h}, P) = \mathbb{E}[L(\widehat{h}, P)] = \mathbb{E}[F(\widehat{h}, P)] - \inf_{h\in\mathcal{H}} F(h, P),$$

and the minimax risk is

$$R^\star = \inf_{\widehat{h}} \sup_{P\in\mathcal{P}} R(\widehat{h}, P).$$

Example 1 (Estimation error in a hypothesis class). $\mathcal{H} \subseteq \{h : \mathcal{X} \to \mathcal{Y}\}$

Our estimator $\hat{h}$ will choose some hypothesis in $\mathcal{H}$ using data. We can now view $L(h, P)$ as the estimation error. Recall, that letting h^* be the Bayes' optimal classifier, we can write

$$F(h, P) - F(h^*, P) = \underbrace{F(h, P) - \inf_{h'\in\mathcal{H}} F(h', P)}_{\texttt{estimation error}=L(h,P)} + \underbrace{\inf_{h'\in\mathcal{H}} F(h', P) - F(h^*, P)}_{\texttt{approximation error}}.$$

In Homework 1, we saw that for ERM, when $\mathcal{H}$ has VC dimension $d_{\mathcal{H}}$, we have

$$R(\hat{h}_{ERM}, P) = \underbrace{\mathbb{E}_S[F(\hat{h}_{ERM}(S), P)]}_{\mathbb{E}_S[\mathbb{E}_{X,Y\sim P}[(\mathbb{1}(\hat{h}_{ERM}(S)(X)\neq Y)]]} - \inf_{h\in\mathcal{H}} F(h, P) \in \tilde{O}\left(\sqrt{\frac{d_{\mathcal{H}}}{n}}\right)$$

We will use this framework to show a corresponding lower bound

$$\inf_{\hat{h}} \sup_{P\in\mathcal{P}} \left(\mathbb{E}_{S\sim P}[F(\hat{h}(S), P)] - \inf_{h'\in\mathcal{H}} F(h', P)\right) \in \Omega\left(\sqrt{\frac{d_{\mathcal{H}}}{n}}\right)$$

To proceed, we will first define the separation of two distributions, with respect to a given hypothesis class and loss L.

Definition 1 (Separation)**.** *For two distributions P, Q, define the separation $\Delta(P, Q)$ as*

$$[\text{illegible}]\; L(h,P) \leq \delta \Rightarrow L(h,Q) \geq \delta, \forall h \in \mathcal{H}$$
$$L(h,Q) \leq \delta \Rightarrow L(h,P) \geq \delta, \forall h \in \mathcal{H}\}$$

- *P, Q are δ-separ*[illegible]*at does well on P (i.e. $L(h,P) \leq \delta$), does poorly on Q (i.e. $L(h,Q) \geq \delta$)*
- *We say a collecti*[illegible]*$\cdots, P_N\}$ are δ-separated if $\Delta(P_i, P_k) \geq \delta$, $\forall j \neq k$.*

The following theor[illegible]g a similar technique to our previous thoerem on reducing estimation to testing. You will do this in your homework.

Theorem 2 (Reduction to testing)**.** *Let $\{P_1, \cdots, P_N\}$ be a δ-separated subset of $\mathcal{P}$. Let ψ be any test which maps the dataset to $[N]$. Then*

$$R^* \geq \delta \inf_{\psi} \max_{j \in [N]} P_j(\psi \neq j)$$

We can then establish the following statements from the above result when S consists of n i.i.d data points.

Theorem 3 (Le Cam & Fano Method)**.** 1. ***Le Cam:*** *If $\{P_0, P_1\}$ are δ-separated,*

$$R^* \geq \frac{\delta}{2} \|P_0 \wedge P_1\| \geq \frac{\delta}{4} e^{-\mathrm{KL}(P_0,P_1)}$$

Hence, for i.i.d. data $S \sim P^n$, if $\mathrm{KL}(P_0, P_1) \leq \frac{\log(2)}{n}$, then $R^* \geq \frac{\delta}{8}$

2. ***Local Fano Method:*** *If $\{P_1, \cdots, P_N\}$ are δ-separated, then*

$$R^* \geq \delta \left(1 - \frac{\frac{1}{N^2}\sum_{j,k} \mathrm{KL}(P_j, P_k) + \log(2)}{\log(N)}\right)$$

Hence, for i.i.d. data $S \sim P^n$, if $\mathrm{KL}(P_j, P_k) \leq \frac{\log(N)}{4n}$, and $N \geq 16$, then $R^* \geq \frac{\delta}{8}$

Remark While our focus is on prediction problems, this framework and theorems apply to any problem for which

$$\inf_{h \in \mathcal{H}} L(h, P) = 0 \quad \forall P \in \mathcal{P}.$$

2 Application: Classification in a VC class

We will now use the above results to prove a lower bound for classification in a VC class.

Theorem 4. *Let $\mathcal{P}$ be the set of all distributions supported on $\mathcal{X} \times \{0,1\}$. Let $\mathcal{H} \subseteq \{h : \mathcal{X} \to \mathcal{Y}\}$ be a hypothesis class with VC dimension $d \geqslant 8$. Let $S = \{(X_1, Y_1), \ldots, (X_n, Y_n)\} \sim_{iid} P$, where $P \in \mathcal{P}$. Then, for any estimator $\hat{h}$ which maps the data set S to a hypothesis in $\mathcal{H}$,*

$$R^* = \inf_{\hat{h}} \sup_{P \in \mathcal{P}} \left(\mathbb{E}[F(\hat{h}, P)] - \inf_{h' \in \mathcal{H}} F(h', P)\right) \geqslant C_1 \sqrt{\frac{d}{n}}$$

for some global constant C_1.

Proof Our proof will follow the usual four step recipe when applying Fano/Le Cam methods.

Step 1: Construct alternatives.

Let $\mathcal{X}_d = \{x_1, \ldots, x$ [illegible] attered by $\mathcal{H}$. Let $\gamma \leqslant 1/4$ be a value which will be specified later. Define

$$\mathcal{P}' = \{P_\omega : I \text{ [illegible] },\ P_\omega(Y = 1|X = x_i) = \frac{1}{2} + (2\omega_i - 1)\gamma,\ \omega \in \Omega_d\},$$

where Ω_d is the VG-pru [illegible] $\}^d$.

Remark To illustrat [illegible] on, consider the class of two-sided threshold classifiers with $d = 2$, i.e. $\mathcal{X}_2 = \{x_1, x_2$ [illegible] istribution for $\omega = (0, 1)$ with $P_\omega(X = x_1) = P_\omega(X = x_2) = 1/2$. Then the conditional distribution of Y should be

$$P_\omega(Y = 1|X = x_1) = \frac{1}{2} - \gamma, \qquad P_\omega(Y = 1|X = x_2) = \frac{1}{2} + \gamma$$

Step 2: Lower bound the separation $\min_{\omega,\omega'} \Delta(P_\omega, P_{\omega'})$.

We have the following claim: For any $P_\omega, P_{\omega'} \in \mathcal{P}'$, the separation satisfies

$$\Delta(P_\omega, P_{\omega'}) \geqslant \frac{\gamma}{d} H(\omega, \omega').$$

We will prove this claim in homework. Then by the Varshamov-Gilbert lemma, we have

$$\min_{\omega,\omega'} \Delta(P_\omega, P_{\omega'}) \geqslant \frac{\gamma}{d}\frac{d}{8} = \frac{\gamma}{8} \triangleq \delta.$$

Step 3: Upper bound the KL divergence $\max_{\omega,\omega'} KL(P_\omega, P_{\omega'})$. We have,

$$\begin{aligned}
KL(P_\omega, P_{\omega'}) &= \mathbb{E}_{X,Y}\left[\log \frac{P_\omega(X,Y)}{P_{\omega'}(X,Y)}\right] \\
&= \sum_{i=1}^{d} P_\omega(x_i) \sum_{y\in\{0,1\}} P_\omega(y|x_i) \log \frac{P_\omega(y|x_i)}{P_{\omega'}(y|x_i)} \qquad (\text{as } P_\omega(x) = P_{\omega'}(x)) \\
&= \sum_{i=1}^{d} \frac{1}{d} \mathbb{I}_{\{\omega \neq \omega'\}} \underbrace{\left[(1/2 + \gamma)\log\frac{1/2+\gamma}{1/2-\gamma} + (1/2 - \gamma)\log\frac{1/2-\gamma}{1/2+\gamma}\right]}_{=O(r^2)} \\
&\leqslant C_2 \frac{\gamma^2}{d} H(\omega, \omega').
\end{aligned}$$

Therefore, with $H(\omega, \omega') \leqslant d$, we have

$$\max_{\omega,\omega'} KL(P_\omega, P_{\omega'}) \leqslant C_2 \gamma^2.$$

Step 4: To conclude the proof, we will choose $\gamma = C_3\sqrt{d/n}$. Then we have

$$\max_{\omega,\omega'} KL(P_\omega, P_{\omega'}) \leqslant C_4 \frac{d}{n} \leqslant \frac{\log(2^{d/8})}{4n} \leqslant \frac{\log(|\mathcal{P}'|)}{4n},$$

where the last inequality is by the Varshamov-Gilbert lemma. Then, by the local Fano method, we have

$$R^* \geqslant \frac{\delta}{2} \geqslant C_5\sqrt{\frac{d}{n}}.$$

$\square$

3 Stochastic Bandits

In the next series of lec[illegible]sing sequential/adaptive decision making problems in which there exists a sequence [illegible] a learner and an environment. Specifically, on round t, the learner chooses an actio[illegible] a set of possible actions. Then the environment reveals an observation O_t. In retu[illegible] a reward $X_t = X_t(O_t, A_t)$. The learner's goal is to maximize the sum of rewards $\sum$[illegible]ersarial bandits and online learning are typical examples of sequential/adaptive dec[illegible] We will first focus on stochastic bandits.

A stochastic bandit [illegible]ollowing components:

- Let $\nu = \{\nu_a, a \in$ [illegible]ibutions indexed by actions in $\mathcal{A}$. ν is called a <u>bandit model</u> and is a subset of some family $\mathcal{P}$.
- On round t, the learner chooses $A_t \in \mathcal{A}$ and observes a reward X_t sampled from ν_{A_t}.

- The learner is characterized by a policy $\Pi = (\Pi_t)_{t \in N}$, where Π_t maps the history $\{(A_s, X_s)\}_{s=1}^{t-1}$ to an action in $\mathcal{A}$.
- If Π is a randomized policy, Π_t maps the history to a probability distribution on $\mathcal{A}$, and then an action is sampled from this distribution. It can also be a deterministic policy.

- $\mu_a = \mathbb{E}_{X \sim \nu_a}[X]$ is defined to be the expected reward of the action a. Let $a^* \in \arg\max_{a \in \mathcal{A}} \mu_a$ be an optimal action, and let $\mu_* = \mu_{a^*}$ be the corresponding optimal value of the expected reward.

- Finally, we define the regret after T rounds of interaction as

$$R_T = R_T(\pi, \nu) = T\mu^* - \mathbb{E}[\sum_{t=1}^{T} X_t]$$

where $\mathbb{E}$ is with respect to the distribution of the action-reward sequence $A_1, X_1, A_2, X_2, ..., A_T, X_T$ induced by the interaction between the policy π and bandit model ν. Here, μ_a, a^*, and μ_* should be viewed as functions of the the bandit model ν and can be written as $\mu_a(\nu)$, $a^*(\nu)$, and $\mu_*(\nu)$.

When designing an algorithm for bandits, at the bare minimum, we will require $R_T \in \mathcal{O}(T)$, i.e. $\lim\limits_{T \to \infty} \frac{R_T}{T} = 0$. This implies that over time, a learner is able to eventually learn the optimal arm.