MLT Homework 13

Ana Borovac
Bas Haver

December 10, 2018

Question 1

Finite Θ*.*

Subquestion 1.1

Verify that $\bar{p}$ *is a probability mass function on* $\{0,1\}^m$*.*

Solution

$$\begin{aligned}
\sum_{(z_1,\ldots,z_m)\in\{0,1\}^m} \bar{p}(z_1,\ldots,z_m) &= \sum_{(z_1,\ldots,z_m)\in\{0,1\}^m} \sum_{\theta\in\Theta} w(\theta) p_\theta(z^m) \\
&= \sum_{(z_1,\ldots,z_m)\in\{0,1\}^m} \sum_{\theta\in\Theta} \frac{1}{N} p_\theta(z^m) \\
&= \frac{1}{N} \sum_{\theta\in\Theta} \sum_{(z_1,\ldots,z_m)\in\{0,1\}^m} p_\theta(z^m) \\
&= \frac{1}{N} \sum_{\theta\in\Theta} 1 \\
&= \frac{1}{N} N \\
&= 1
\end{aligned}$$

Subquestion 1.2

Show that the worst-case case regret of $\bar{p}$ *is also bounded by* $\log N$*.*

Solution

$$\begin{aligned}
-\log \bar{p}(z^m) &= \sum_{i=1}^{m} -\log \bar{p}(z_i|z^{i-1}) \\
&= -\log \bar{p}(z_1|\epsilon) - \cdots - \log \bar{p}(z_m|z^{m-1}) \\
&= -(\log \bar{p}(z_1|\epsilon) + \cdots + \log \bar{p}(z_m|z^{m-1})) \\
&= -\left(\log \frac{\bar{p}(z^1)}{\bar{p}(\epsilon)} + \cdots + \log \frac{\bar{p}(z^m)}{\bar{p}(z^{m-1})}\right) \\
&= -\left(\log \frac{\bar{p}(z^1)}{\bar{p}(\epsilon)} \cdot \cdots \cdot \frac{\bar{p}(z^m)}{\bar{p}(z^{m-1})}\right) \\
&= -\left(\log \frac{\bar{p}(z^m)}{\bar{p}(\epsilon)}\right) \\
&= -\log \bar{p}(z^m)
\end{aligned}$$

$$\begin{aligned}
-\log \bar{p}(z^m) &= \sum_{i=1}^{m} -\log \bar{p}(z_i|z^{i-1}) \\
&\geq \max_{i \in \{1,\ldots,m\}} -\log \bar{p}(z_i|z^{i-1})
\end{aligned}$$

Subquestion 1.3

Let us fix $N = |\Theta|$ and set $\Theta = \{1/(N+1), ..., N/(N+1)\}$. For example, if we set $N = 4$ then $\Theta = \{0.2, 0.4, 0.6, 0.8\}$.

Solution

/

Subquestion 1.4

Use the law of large numbers to argue that

$$\lim_{m \to \infty} S_m = \log N.$$

Solution

Subquestion 1.5

Informally explain why, for small sample sizes m, the Shtarkov sum for $\Theta = \{0.47, 0.49, 0.51, 0.53\}$ *is significantly smaller than the Shtarkov sum for* $\Theta = \{0.2, 0.4, 0.6, 0.8\}$.

Solution

Because elements of Θ are much closer to each other.

Subquestion 1.6

Suppose that P *consists of a finite number of black-box experts, which given each history* Z_1, Z_{i-1} *provide us a distribution on* Z_i*, which may depend on the past — we don't know how they come up with their predictions, we just observe the predictions they make on the sample. Explain why we'd rather use the Bayesian than the Shtarkov predictor in such a setting.*

Solution

Question 2

(Uncountable Θ*,* $[4+1/3$ *pt* $]$*) Now consider the full Bernoulli model,* $\Theta = [0,1]$*, with the sum in* (1) *repladed by an integral, for the uniform prior probability densiwty* $w(\theta) := 1$ *for all* $\theta \in [0,1]$.

Subquestion 2.1

For each of the follwoing statements, indicate wheter it is true or false, and prove your answer.

1. $[3 + 1/3pt]$ The NML predictor p^* achieves the same regret for every sequence $x^m - x_1, \dots, x_m \in \{0,1\}^m$.

   Solution

   True.

$$p_n^*(x^n) = \frac{\sup_{\theta \in \Theta} \theta_n(x^n)}{\sum_{z^n \in Z^n} \sup_{\theta \in \Theta} \theta_n(z^n)}$$

$$
\begin{aligned}
\hat{L}(x^n) - \inf_{\theta \in \Theta} L_\theta(x^n) &= \log \frac{\sup_{\theta \in \Theta} \theta_n(x^n)}{p_n^*(x^n)} \\
&= \log \sum_{z^n \in Z^n} \sup_{\theta \in \Theta} \theta_n(z^n) \\
&\Rightarrow \text{independant of } x^n \\
&\Rightarrow \text{regret is the same for } \forall x^n
\end{aligned}
$$

2. same as (i) with 'regret' replaced by 'cumulative loss'.

   Solution

   False.

$$
p_n^*(x^n) = \frac{\sup_{\theta \in \Theta} \theta_n(x^n)}{\sum_{z^n \in Z^n} \sup_{\theta \in \Theta} \theta_n(z^n)}
$$

$$
\begin{aligned}
\hat{L}(x^n) &= \log \frac{1}{p_n^*(x^n)} \\
&= \log \frac{\sum_{z^n \in Z^n} \sup_{\theta \in \Theta} \theta_n(z^n)}{\sup_{\theta \in \Theta} \theta_n(x^n)} \\
&\Rightarrow \text{idependant on } x^n \\
&\Rightarrow \text{cumulative loss is not the same for } \forall x^n
\end{aligned}
$$

3. the Bayesian predictor $\bar{p}$ achieves the same regret for every sequence $x^m \in \{0,1\}^m$.

   Solution

   False.

$$
\bar{p}_n(x^n) = \int w(\theta) p_\theta(x^n)\, d\theta = \int p_\theta(x^n)\, d\theta
$$

$$
\begin{aligned}
\hat{L}(x^n) - \inf_{\theta \in \Theta} L_\theta(x^n) &= \log \frac{\sup_{\theta \in \Theta} \theta_n(x^n)}{\bar{p}_n(x^n)} \\
&= \log \frac{\sup_{\theta \in \Theta} \theta_n(x^n)}{\int p_\theta(x^n)\, d\theta} \\
&\Rightarrow \text{dependant on } x^n \\
&\Rightarrow \text{regret is the not the same for } \forall x^n
\end{aligned}
$$

4. same as (iii) with 'regret' replaced by 'cumulative loss'.

Solution

???.

$$\bar{p}_n(x^n) = \int w(\theta) p_\theta(x^n)\, d\theta = \int p_\theta(x^n)\, d\theta$$

$$\begin{aligned}
\hat{L}(x^n) &= \log \frac{1}{\bar{p}_n(x^n)} \\
&= \log \frac{1}{\int p_\theta(x^n)\, d\theta} \\
&\Rightarrow \\
&\Rightarrow
\end{aligned}$$

5. for every fixed $x^m \in \{0,1\}^m$, the NML predictor p^* achieves the same regret on every y^m that is a permutation of x^m.

   Solution

   True.
   Since the regret is the same for every x^n, it is also the same for the permutation of x^n.

6. same as (vii) with 'regret' replaced by 'cumulative loss'.

   Solution

   True.

$$\begin{aligned}
\hat{L}(x^n) &= \log \frac{1}{p_n^*(x^n)} \\
&= \log \frac{\sum_{z^n \in Z^n} \sup_{\theta \in \Theta} \theta_n(z^n)}{\sup_{\theta \in \Theta} \theta_n(x^n)} \\
&\Rightarrow \sup_{\theta \in \Theta} \theta_n(x^n) \text{ is independant on the order of the elements of } x^n \\
&\Rightarrow \text{cumulative loss is the same for every permutation of } x^n
\end{aligned}$$

7. for every fixed $x^m \in \{0,1\}^m$, every $n < m$, the NML predictor p^* achieves the same loss on $y_1, \ldots, y_n$, for every $y_1, \ldots, y_n$ that are the initial segment of an $y_1, \ldots, y_m$ that is a permutation of x^m.

8. same as (ix) but with p^* replaced by $\bar{p}$.

Subquestion 2.2

[1 pt] Suppose we do not know in advance how many predictions we have to make, i.e. we do not know the *horizon* n. Explain why this is a problem for prediction with the NML p^* but not for prediction with $\bar{p}$.