

MLT Homework 13

Ana Borovac
Bas Haver

December 9, 2018

Question 1

Finite Θ .

Subquestion 1.1

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

Solution

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

Subquestion 1.2

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

Solution

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

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

Subquestion 1.3

Let us fix $N = |\Theta|$ and set $\Theta = \{1/(N+1), \dots, N/(N+1)\}$. 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

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

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