Prediction Learning and Games – Exercise Solutions

Ludovic Schwartz

Contents

Chapter 2

Prediction with Expert Advice

Exercise 2.2

Consider a weighted average forecaster based on a potential function

$$\Phi(u) = \psi\left(\sum_{i=1}^{N} \phi(u_i)\right).$$

Assume further that the quantity $C(r_t)$ appearing in the statement of Theorem 2.1 is bounded by a constant for all values of r_t and that the function $\psi(\phi(u))$ is strictly convex. Show that there exists a nonnegative sequnce $\epsilon_n \to 0$ such that the cumulative regret of the forecaster satisfies, for every n and for every outcome sequence y^n,

$$\frac{1}{n}\left(\max_{i=1,\ldots N} R_{i,n}\right) \leq \epsilon_n$$

Solution : We start by assuming that $C(r_t) \leq C$ for all values of r_t. We apply theorem 2.1 and we get that :

$$\Phi(R_n) \leq \Phi(0) + \frac{1}{2}\sum_{t=1}^{n} C(r_t) \leq \Phi(0) + \frac{Cn}{2}$$

Then, since we know that $\psi \circ \phi$ is non decreasing and strictly convex, it must be increasing and as a result, both $\psi \circ \phi$ and ϕ must be invertible and increasing and we have :

$$\psi\left(\phi\left(\max_{i=1,\ldots,N} R_{i,n}\right)\right) \leq \psi\left(\max_{i=1,\ldots,N} \phi(R_{i,n})\right) \leq \psi\left(\sum_{i=1}^{N} \phi(R_{i,n})\right) = \Phi(R_n)$$

Hence :

$$\max_{i=1,\dots,N} R_{i,n} \leq \phi^{-1}(\psi^{-1}(\Phi(R_n)))$$

$$\frac{1}{n}\left(\max_{i=1,\dots,N} R_{i,n}\right) \leq \underbrace{\frac{(\psi\circ\phi)^{-1}\left(\Phi(0)+\frac{Cn}{2}\right)}{n}}_{\epsilon_n}$$

Now we need to show that $\epsilon_n \to 0$. That is the same as saying that for a stricly convex increasing function F, we have $\lim_{x\to\infty}\frac{F^{-1}(x)}{x} = 0$ or equivalenty $\lim_{x\to\infty}\frac{F(x)}{x} = +\infty$. This doesn't seem to be true in general. Indeed, the function $F(x) = \sqrt{x^2+1}-1$ is stricly convex but $F(x) \underset{x\to\infty}{\sim} x$. The result would be true if we assume that $\psi\circ\phi$ is strongly convex as it would be lower bounded by a positive quadratic.

$\square$

Exercise 2.5 Nonuniform initial weights

By definition, the weighted average forecaster uses uniform initial weights $w_{i,0} = 1$ for all $i = 1,\dots,N$. However, there is nothing special abut this choice, and the analysis of the regret for this forecaster can be carried out using any set of nonnegative numbers for the inital weights.

Consider the exponentially weighted average forecaster run with arbitrary inital weights $w_{1,0},\dots,w_{N,0} > 0$, defined for all $t = 1,2,\dots,$ by

$$\hat{p}_t = \frac{\sum_{i=1}^N w_{i,t-1}f_{i,t}}{\sum_{j=1}^N w_{j,t-1}}, \quad w_{i,t} = w_{i,t-1}\exp -\eta\ell(f_{i,t},y_t).$$

Under the same conditions as in the statement of Theorem 2.2, show that for every n and for every outcome sequence y^n,

$$\widehat{L_n} \leq \min_{i=1,\dots,N}\left(L_{i,n} + \frac{1}{\eta}\log\frac{1}{w_{i,0}}\right) + \frac{\log W_0}{\eta} + \frac{\eta}{8}n$$

where $W_0 = w_{1,0} + \cdots + w_{N,0}$

Solution : As in the proof of theorem 2.2, we define :

$$\begin{aligned} W_t &= \sum_{i=1}^N w_{i,t} \\ &= \sum_{i=1}^N \exp\left(-\eta L_{i,t}\right) w_{i,0} \\ &= \sum_{i=1}^N \exp -\eta\left(L_{i,t} + \frac{1}{\eta}\log\frac{1}{w_{i,0}}\right) \end{aligned}$$

Now as in the proof of theorem 2.2, we analyse $\log \frac{W_t}{W_0}$ and $\log \frac{W_t}{W_{t-1}}$. We have :

$$\begin{aligned}
\log \frac{W_n}{W_0} &= \log \left(\sum_{i=1}^{N} \exp -\eta \left(L_{i,n} + \frac{1}{\eta} \log \frac{1}{w_{i,0}} \right) \right) - \log W_0 \\
&\geq \log \left(\max_{i=1,\ldots,N} \exp -\eta \left(L_{i,n} + \frac{1}{\eta} \log \frac{1}{w_{i,0}} \right) \right) \\
&= -\eta \left(\min_{i=1,\ldots,N} L_{i,n} + \frac{1}{\eta} \log \frac{1}{w_{i,0}} \right) - \log W_0
\end{aligned}$$

On the other hand, exactly as in the proof of Theorem 2.2, we get for each $t = 1, \ldots, n$,

$$\begin{aligned}
\log \frac{W_t}{W_{t-1}} &= \log \frac{\sum_{i=1}^{N} \exp\left(-\eta \ell(f_{i,t}, y_t)\right) \exp\left(-\eta \left(L_{i,t-1} + \frac{1}{\eta} \log \frac{1}{w_{i,0}}\right)\right)}{\sum_{i=1}^{N} \exp\left(-\eta \left(L_{i,t-1} + \frac{1}{\eta} \log \frac{1}{w_{i,0}}\right)\right)} \\
&= \log \frac{\sum_{i=1}^{N} w_{i,t-1} \exp\left(-\eta \ell(f_{i,t}, y_t)\right)}{\sum_{i=1}^{N} w_{i,t-1}} \\
&\leq -\eta \ell(\hat{p}_t, y_t) + \frac{\eta^2}{8}
\end{aligned}$$

Where the last inequality is obtained by using Hoeffding's Lemma and the convecity of the loss exactly like in the proof of Theorem 2.2. Summing over $t = 1, \ldots, n$, we get

$$\log \frac{W_n}{W_0} \leq -\eta \widehat{L_n} + \frac{\eta^2}{8} \eta.$$

Putting this together with the previous upper bound and solving for $\widehat{L_n}$ we get :

$$\widehat{L_n} \leq \min_{i=1,\ldots,N} \left(L_{i,n} + \frac{1}{\eta} \log \frac{1}{w_{i,0}} \right) + \frac{\log W_0}{\eta} + \frac{\eta}{8} n$$

□

Exercise 2.6 Many good experts

Sequences of outcomes on which many experts suffer a small loss are intuitively easier to predict. Adapt the proof of Theorem 2.2 to show that the exponentially weighted forecaster satisfies the following property : for every n, for every outcome sequence y^n and for all $L > 0$,

$$\widehat{L_n} \leq L + \frac{1}{\eta} \log \frac{N}{N_L} + \frac{\eta}{8} n,$$

where N_L is the cardinality of the set $\{1 \leq i \leq N : L_{i,n} \leq L\}$

Solution : We adapt the lower bound of $\log \frac{W_n}{W_0}$. We have :

$$\begin{aligned}
\log \frac{W_n}{W_0} &= \log \left(\sum_{i=1}^{N} e^{-\eta L_{i,n}} \right) - \log N \\
&\geq \log \left(\sum_{i=1}^{N} e^{-\eta L_{i,n}} \cdot \mathbb{I}_{L_{i,n} \leq L} \right) - \log N \\
&\geq \log \left(\sum_{i=1}^{N} e^{-\eta L} \cdot \mathbb{I}_{L_{i,n} \leq L} \right) - \log N \\
&= \log \left(N_L e^{-\eta L} \right) - \log N \\
&= -\eta L - \log \frac{N}{N_L}
\end{aligned}$$

We still have the upper bound of $\log \frac{W_n}{W_0}$ from the proof of Theorem 2.2 :

$$\log \frac{W_n}{W_0} \leq -\eta \widehat{L_n} + \frac{\eta^2}{8} n.$$

Putting everything together, we get :

$$\widehat{L_n} \leq L + \frac{1}{\eta} \log \frac{N}{N_L} + \frac{\eta}{8} n,$$

□

Exercise 2.8 THE DOUBLING TRICK

Consider the following forecasting strategy("Doubling Trick"): time is divided in periods $(2^m, \ldots, 2^{m+1} - 1)$, where $m = 0, 1, 2, \ldots$. In period $(2^m, \ldots, 2^{m+1} - 1)$ the strategy uses the exponentially weighted average forecaster initialized at time 2^m with parameter $\eta_m = \sqrt{8(\log N)/2^m}$. Thus, the weighted average forecaster is reset at each time instance that is an integer power of 2 and is restarted with a new value of η. Using Theorem 2.2 prove that, for any sequence $y_1, y_2, \ldots \in \mathcal{Y}$ of outcomes and for any $n \geq 1$, the regret of this forecaster is at most

$$\widehat{L_n} - \min_{i=1,\ldots,N} L_{i,n} \leq \frac{\sqrt{2}}{\sqrt{2} - 1} \sqrt{\frac{n}{2} \log N}$$

Solution : Let $n \geq 1$, and let M be the only integer such that $2^M \leq n \leq$

2^{M+1}. We have :

$$\begin{aligned}
&\widehat{L_n} - \min_{i=1,\dots,N} L_{i,n} \\
&= \sum_{t=1}^{n} \ell(\hat{p}_t, y_t) - \min_{i=1,\dots,N} \sum_{t=1}^{n} f_{i,t} \\
&= \left(\sum_{m=0}^{M} \sum_{t=2^m}^{2^{m+1}\wedge n} \ell(\hat{p}_t, y_t) \right) - \min_{i=1,\dots,N} \left(\sum_{m=0}^{M} \sum_{t=2^m}^{2^{m+1}\wedge n} f_{i,t} \right) \\
&\leq \sum_{m=0}^{M} \left(\sum_{t=2^m}^{2^{m+1}\wedge n} \ell(\hat{p}_t, y_t) \right) - \sum_{m=0}^{M} \left(\min_{i=1,\dots,N} \sum_{t=2^m}^{2^{m+1}\wedge n} f_{i,t} \right) \\
&= \sum_{m=0}^{M} \left(\underbrace{\sum_{t=2^m}^{2^{m+1}\wedge n} \ell(\hat{p}_t, y_t) - \min_{i=1,\dots,N} \sum_{t=2^m}^{2^{m+1}\wedge n} f_{i,t}}_{R_{m,n}} \right)
\end{aligned}$$

Where $R_{m,n}$ is the regret suffered by our forecaster during phase m(and up to time n) which has exactly $2^{m+1} \wedge n - 2^m \leq 2^m$ steps. In particular, using the bound of Theorem 2.2, and the choice of η_m we know that

$$\forall m \leq M, R_m \leq \sqrt{(2^m/2)\log N}$$

Finally :

$$\begin{aligned}
&\widehat{L_n} - \min_{i=1,\dots,N} L_{i,n} \\
&\leq \sum_{m=0}^{M} \sqrt{(2^m/2)\log N} \\
&\leq \sqrt{\frac{\log N}{2}} \cdot \frac{\sqrt{(2^{M+1})} - 1}{\sqrt{2} - 1} \\
&\leq \sqrt{\frac{\log N}{2}} \cdot \frac{\sqrt{(2^{M+1})}}{\sqrt{2} - 1} \\
&\leq \frac{\sqrt{2}}{\sqrt{2} - 1} \sqrt{\frac{n}{2} \log N}
\end{aligned}$$

Where we used $2^{M+1} \leq 2n$ on the last line.

$\square$