Exercise 1: Online Updates

Suppose $z_1, \ldots, z_t \in \mathbb{R}^d$ are the environmental data points seen until time $t \in \mathbb{N}$.

(a) Provide an update formula for the empirical mean of the data points for any time instance $s = 1, \ldots, t$ in form of a function $u : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{N} \to \mathbb{R}^d$, such that

$$\bar{z}_s = u(\bar{z}_{s-1}, z_s, s),$$

holds. Here, $\bar{z}_s = \frac{1}{s} \sum_{j=1}^{s} z_j$ denotes the empirical mean at time s and we have the convention that $\sum_{j=1}^{s} z_j = 0$, if $s = 0$.

(b) Provide an update formula for the empirical total variance of the data points for any time instance $s = 1, \ldots, t$ in form of a function $u : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{N} \to \mathbb{R}$, such that

$$v_s = u(\overline{z_{s-1}^2}, \bar{z}_{s-1}, z_s, s)$$

with $v_s = \frac{1}{s} \sum_{j=1}^{s} (z_j - \bar{z}_s)^\top (z_j - \bar{z}_s)$ holds. Here, $\overline{z_s^2} = \frac{1}{s} \sum_{j=1}^{s} z_j^\top z_j$ denotes the empirical mean of the inner products of the data points at time s.

(c) Explain the benefits of having such update formulas for a particular statistic in the online learning framework.

Exercise 2: Doubling Trick

Suppose `Algo` is an (online learning) algorithm for an online learning problem characterized by some action space $\mathcal{A}$, an environmental data space $\mathcal{Z}$ and a loss function L. Further, assume that `Algo` enjoys a (cumulative) regret bound of the form $R_T^{\texttt{Algo}} \leq C \cdot T^\alpha$, for some $\alpha \in (0, 1)$ and some constant $C > 0$, if its parameters are set as a function of T, so that for notational convenience `Algo` can be identified by $\texttt{Algo}_T$. Now, define a second online learning algorithm $\widetilde{\texttt{Algo}}$ for the same online learning problem as follows:

> Algorithm $\widetilde{\texttt{Algo}}$
> **for** epoch $m = 0, 1, 2, \ldots$ **do**
> Reset `Algo` with parameters chosen for $T = 2^m$, i.e., use $\texttt{Algo}_T$.
> **for** time steps $t = 2^m, \ldots, 2^{m+1} - 1$ **do**
> Run $\texttt{Algo}_T$
> **end for**
> **end for**

(a) Illustrate by means of a timeline what $\widetilde{\texttt{Algo}}$ is doing. In particular, what is the length of an epoch and in which time window is $\texttt{Algo}_T$ for the occurring T's used?

(b) Show that the regret of $\widetilde{\texttt{Algo}}$ up to time $\tilde{T}$ is bounded by $\sum_{m=0}^{\lfloor \log_2(\tilde{T}) \rfloor} R_{2^m}^{\texttt{Algo}}$ for any $\tilde{T}$, that is, show

$$R_{\tilde{T}}^{\widetilde{\texttt{Algo}}} \leq \sum\nolimits_{m=0}^{\lfloor \log_2(\tilde{T}) \rfloor} R_{2^m}^{\texttt{Algo}}.$$

Here, $\lfloor x \rfloor := \max\{z \in \mathbb{Z} \,|\, z \leq x\}$.

(c) Conclude that the regret of $\widetilde{\texttt{Algo}}$ up to time $\tilde{T}$ is of order $O(\tilde{T}^\alpha)$, i.e., $R_{\tilde{T}}^{\widetilde{\texttt{Algo}}} \leq \tilde{C} \cdot \tilde{T}^\alpha$ for some constant $\tilde{C} > 0$. Finally, explain why the strategy of algorithm $\widetilde{\texttt{Algo}}$ is called the *doubling trick* and what its advantages and disadvantages are.

Exercise 3: Practical Performance of FTL and FTRL

(a) Consider an online quadratic optimization problem with action space $\mathcal{A} = [-1, 1]^d$, environment data space $\mathcal{Z} = [-1, 1]^d$ and the loss function given by $L(a, z) = \frac{1}{2} ||a - z||_2^2$. Furthermore, let $T = 10000$ be the considered time horizon. Assume that the environmental data is generated uniformly at random in each time step $t \in \{1, \dots, T\}$. Compute the cumulative regret of FTL and of FTRL instantiated with the squared L2-norm regularization and the optimal choice for the regularization magnitude for any time step $t = 1, \dots, T$.

Repeat this procedure 100 times and compute the empirical average of the resulting cumulative regrets in each time step $t = 1, \dots, T$. Note that this results in a curve with support points $(t, \bar{R}_t^{\texttt{Algo}})_{t=1,\dots,T}$, where $\bar{R}_t^{\texttt{Algo}}$ is the average cumulative regret (over the 100 repetitions) of algorithm $\texttt{Algo} \in \{\text{FTL}, \text{FTRL}\}$ till time step t. Plot this mean cumulative regret curve together with the theoretical upper bound for the cumulative regret of the FTRL algorithm in this case into one chart. Include also the empirical standard error of the mean cumulative regret, i.e., include the points $(t, \bar{R}_t^{\texttt{Algo}} \pm \hat{\sigma}(R_t^{\texttt{Algo}}))_{t=1,\dots,T}$, where $\hat{\sigma}(R_t^{\texttt{Algo}})$ is the empirical standard deviation of the cumulative regret (over the 100 repetitions) of algorithm $\texttt{Algo} \in \{\text{FTL}, \text{FTRL}\}$ till time step t. How does the mean cumulative regret curve of FTRL vary with respect to the regularization magnitude? Illustrate this also by means of a chart.

Hint: For d you can, of course, consider different settings.

(b) Repeat (a), but this time consider an online linear optimization problem with the same action space, the same environment data space, but with the loss function given by $L(a, z) = a^\top z$. Comment on your findings.