E1 245 - Online Prediction and Learning, Aug-Dec 2018
Homework #2

1. *Hoeffding's inequality*
   Prove the following inequality for independent random variables $X_1, \ldots, X_T$ taking values in $[0,1]$.
$$\forall \varepsilon \geq 0 \quad \mathbb{P}\left[\frac{\sum_{t=1}^{T} X_t}{T} - \frac{\sum_{t=1}^{T} \mathbb{E}[X_t]}{T} \geq \varepsilon\right] \leq e^{-2T\varepsilon^2}.$$
   Hint: For any $\lambda > 0$ and a non-negative random variable Z, $\mathbb{P}[Z \geq \varepsilon] = \mathbb{P}\left[e^{\lambda Z} \geq e^{\lambda \varepsilon}\right]$; use Markov's inequality, Hoeffding's lemma (from HW 1), optimize over $\lambda > 0$.

2. *Exp-concavity and common loss functions*

   (a) Show that if for a $y \in \mathscr{Y}$ and $\eta > 0$ the function $F(z) := e^{-\eta l(z,y)}$ is concave, then $l(z,y)$ is a convex function of z.

   (b) Show that the relative entropy loss $l(x,y) := y \log \frac{y}{x} + (1-y) \log \frac{1-y}{1-x}$, $x, y \in [0,1]$, is 1-exp-concave for all valid values[1] of y.

   (c) Show that the squared loss $l(x,y) := (x-y)^2$, $x, y \in [0,1]$, is $\frac{1}{2}$-exp-concave for all valid values of y.

   (d) Show that the absolute value loss $l(x,y) := |x-y|$, $x, y \in [0,1]$, *cannot* be η-exp-concave for any $\eta > 0$.

3. *Improved regret with exp-concave loss functions*
   Show that if the Exponentially Weighted Forecaster (EXPWTS) is run in the prediction-with-expert-advice setting with a σ-exp-concave loss function $l : \mathscr{D} \times \mathscr{Y} \to [0,1]$ (over $\mathscr{D}$) and the learning rate $\eta = \sigma > 0$ over N experts, then the algorithm enjoys the regret bound
$$\sum_{t=1}^{T} l(p_t, y_t) - \min_{i \in [N]} \sum_{t=1}^{T} l(f_{i,t}, y_t) \leq \frac{\log N}{\sigma}.$$
   (note: regret does not grow with time T!)

4. Establish the following properties, that we used in class, to prove a regret bound for Cover's Universal Portfolio algorithm.

   (a) Let $b^* \in \Delta_m$ represent a Constantly Rebalancing Portfolio (CRP) on the (non-negative) unit simplex in $\mathbb{R}_+^m$. Let $\text{Ball}_\varepsilon(b^*) := \{(1-\varepsilon)b^* + \varepsilon b : b \in \Delta_m\}$ for $\varepsilon \in [0,1]$. If $\text{Vol}(A)$ denotes the $(m-1)$-dimensional volume[2] of a set $A \subseteq \Delta_m$, then show that $\text{Vol}(\text{Ball}_\varepsilon(b^*)) = \varepsilon^{m-1} \text{Vol}(\Delta_m)$.

   (b) Show that the CRP strategy $b \in \text{Ball}_\varepsilon(b^*)$ achieves wealth $S_T(b, x^T) \geq S_T(b^*, x^T)(1-\varepsilon)^T$ in T investment periods.

[1] By convention, we take $\frac{0}{0} := 0$ & $0 \cdot \log 0 := 0$.

[2] $\text{Vol}(A)$ can be taken to be the $(m-1)$-dimensional "surface area" of the surface defined by $x_m = f(x_1, \ldots, x_{m-1}) := 1 - \sum_{1=1}^{m-1} x_i$, for $x_1, \ldots, x_{m-1} \geq 0$, $\sum_{1=1}^{m-1} x_i \leq 1$. Alternatively, $\text{Vol}(A)$ can be defined to be the probability of a point lying in the set A when it is drawn from the uniform probability distribution over Δ_m.